

# Módulo 1

## Zeros de funções reais

### 1. Introdução

**Prezado estudante, seja bem vindo.**

Neste módulo você aprenderá o conceito de zero de uma função real,  $f: \mathbb{R} \rightarrow \mathbb{R}$ , onde  $\mathbb{R}$  é o conjunto dos números reais. Por definição,  $r$  é um zero de  $f$  se, e somente se,  $r$  é raiz da equação  $f(x) = 0$ ; ou seja,  $f(r) = 0$ . Um exemplo bastante conhecido é o de se obter as raízes da equação de segundo grau:  $f(x) = ax^2 + bx + c = 0$ . Os zeros da parábola (a função  $f$  dada anteriormente) são obtidos pela fórmula de Bhaskara, que é bastante conhecida. Se  $\Delta = b^2 - 4ac \geq 0$ , então os zeros reais da parábola ou, equivalentemente, as raízes reais da equação do segundo grau são  $r_1 = \frac{-b + \sqrt{\Delta}}{2a}$  e  $r_2 = \frac{-b - \sqrt{\Delta}}{2a}$ .

Em geral, equações do tipo  $f(x) = 0$  não possuem fórmulas, como a de Bhaskara, que forneçam as suas raízes através de um número finito de operações. Por exemplo, a equação  $f(x) = x - 1,5 \operatorname{sen}(x) = 0$  possui uma raiz  $r = 0$ , pois  $f(0) = 0 - 1,5 \operatorname{sen}(0) = 0$ . Porém, a raiz positiva somente será obtida através de um método numérico que vai gerar uma sequência infinita de números reais que se aproximará dessa raiz com qualquer precisão desejada. Uma tal sequência é dada, por exemplo, por  $r_n = 1,5 \operatorname{sen}(r_{n-1})$ , onde  $r_0 = 1,3$  e  $n \in \mathbb{N}$  (conjunto dos números naturais). Note que  $r_1 = 1,5 \operatorname{sen}(r_0) = 1,5 \operatorname{sen}(1,3) = 1,4453372278$ . Observe que o cálculo do seno está retornando valores em radianos. Esse será o padrão sempre que estivermos lidando com as funções trigonométricas.

Dessa forma, a aproximação do zero de  $f(x) = x - 1,5 \operatorname{sen}(x)$ , com 9 casas decimais, é dada por  $r = 1,495781568$ . Note que  $f(r)$  é aproximadamente  $-0,19713 \times 10^{-9}$ . Para esses valores serem obtidos, a sua calculadora deverá executar os cálculos em radianos (lembre-se de que a medida de um ângulo dada em graus não representa um número real).

## 2. Objetivos e conteúdo do Módulo 1

O objetivo deste módulo será o de estudar métodos numéricos que obtenham aproximações de zeros de uma função real. Para você ficar familiarizado com as técnicas numéricas de obtenção de raízes de equações não lineares do tipo  $f(x) = 0$ , serão utilizados vários recursos: resolução de listas de exercícios; leitura de materiais didáticos e elaboração de algoritmos computacionais.

Espero que possamos fazer uma ótima parceria nesta disciplina do curso de Licenciatura em Matemática do Plano Nacional de Formação de Professores da Educação Básica – PARFOR.

### Conteúdos básicos do Módulo 1

- Isolamento de raízes.
- Método da Bissecção.
- Método Iterativo Linear.
- Método de Newton - Raphson.

## 3. Isolamento de raiz

Em muitos problemas de Ciências e Engenharia há a necessidade de se determinar um número  $r$  para o qual uma função  $f(x)$  seja zero, isto é,  $f(r) = 0$ .

Esse número é chamado de zero de  $f$  ou raiz da equação  $f(x) = 0$ . Em nossos estudos  $r$  será real.

Graficamente, os zeros reais são representados pelos pontos de interseção da curva representada pelos pontos  $(x, f(x))$  com o eixo horizontal ( $Ox$ ;  $O = (0, 0)$ ), conforme a Figura 1 abaixo.

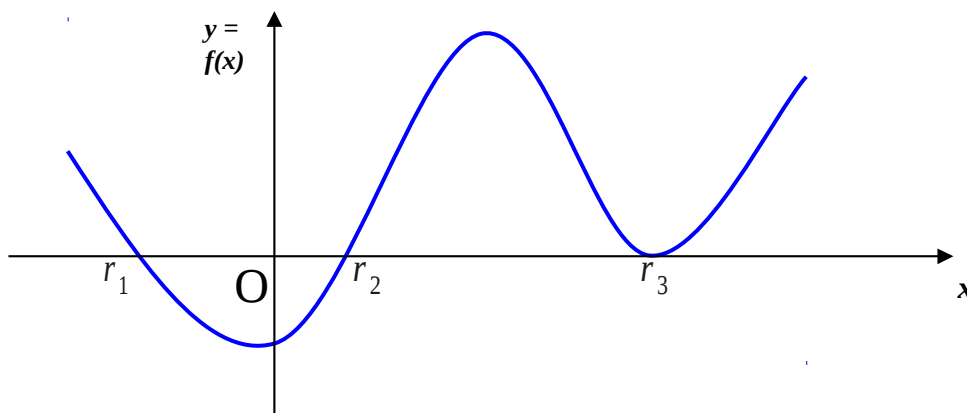


Figura 1: Zero de função

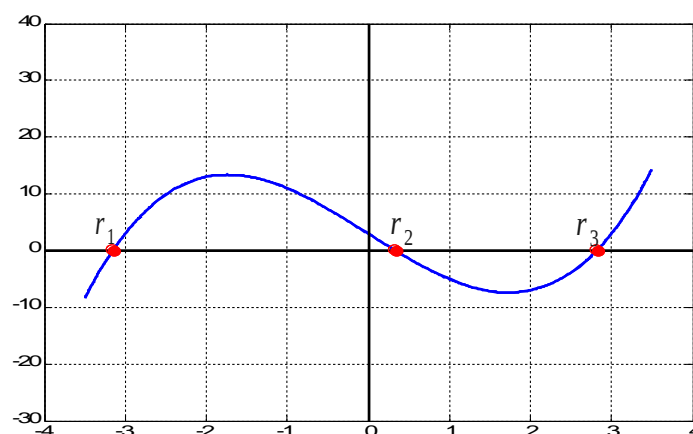
Duas fases são importantes para a determinação dos zeros de  $f$ : (I) isolamento das raízes da equação  $f(x) = 0$ ; (II) Geração de um processo iterativo que irá convergir para uma determinada raiz da equação  $f(x) = 0$ . Esse processo será chamado de refinamento da raiz ou refinamento do intervalo que contém uma raiz da equação  $f(x) = 0$ .

### Fase I - Localização ou isolamento das raízes

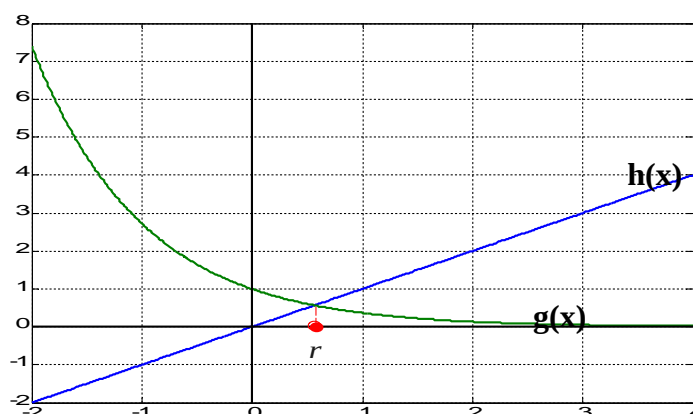
Consiste em obter um intervalo  $[a, b]$  que contém uma única raiz da equação  $f(x) = 0$ .

Nessa fase, é feita uma análise gráfica e teórica da função  $f$ . A precisão dessa análise é o pré-requisito para o sucesso da fase II. Os seguintes procedimentos são utilizados:

i) Esboçar o gráfico da função  $f(x)$  e localizar as abscissas dos pontos de interseção da curva com o eixo horizontal (Ox). Por exemplo, se  $f(x) = x^3 - 9x + 3$ , então podemos isolar cada uma das três raízes de  $f(x) = 0$  nos seguintes intervalos  $[-4, -3]$ ,  $[0, 1]$  e  $[2, 3]$ . Veja o gráfico apresentado a seguir.



ii) A partir da equação  $f(x) = 0$ , obtém-se uma equação equivalente  $g(x) = h(x)$ . Os pontos de interseção das curvas  $g$  e  $h$  são os zeros de  $f$ . Isto é,  $g(x) = h(x) \Leftrightarrow f(x) = 0$ . Por exemplo, a equação  $f(x) = e^{(-x)} - x = 0$  é equivalente à equação  $g(x) = e^{(-x)} = x = h(x)$ . Dessa forma, encontramos o intervalo  $[0, 1]$  que contém a raiz de  $f(x) = 0$ . Veja o gráfico exibido a seguir.



A análise teórica consiste em usar o Teorema do Valor Intermediário ou de Bolzano (veja, mais adiante, as referências [4, 5] – Elon Lages Lima), que será enunciado a seguir, sem demonstração (veja o **Exercício 2** no final da Seção 4).

**Teorema 1 (Teorema do Valor Intermediário).** Seja  $f$  uma função contínua no intervalo fechado  $[a, b]$ . Se  $f(a)f(b) < 0$ , então existe pelo menos um ponto  $x = r$  entre  $a$  e  $b$  tal que  $f(r) = 0$ .



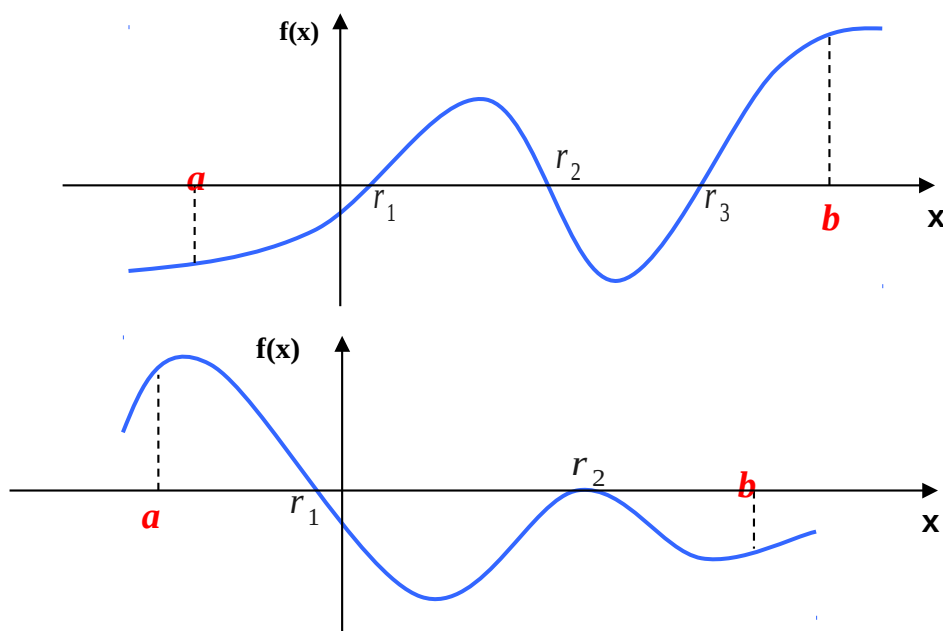
**Observação:** A demonstração do Teorema do Valor Intermediário (TVI) é vista em cursos de Análise Real. A **Figura 2** adiante exhibe interpretações geométricas desse teorema.



## Referências

- [1] ÁVILA, G. *Análise Matemática para a Licenciatura*, São Paulo, Edgard Blucher, 2006.
- [2] ÁVILA, G. *Introdução à análise matemática*, São Paulo, Edgard Blucher, 1992.
- [3] FIGUEIREDO, D. G. *Análise 1*, 2ª Edição, São Paulo, Livros Técnicos e Científicos, 1996.
- [4] LIMA, E. L. *Curso de análise*. Volume 1. Projeto Euclides, Rio de Janeiro, SBM, 2000.
- [5] LIMA, E. L. *Análise real*. Volume 1. Coleção Matemática Universitária, Rio de Janeiro, SBM, 2001.

### Interpretação Geométrica do TVI (Figura 2):



**Figura 2: Interpretação geométrica do TVI – 2 casos**

Incluindo mais uma hipótese no TVI, obtemos um outro teorema importante para o isolamento de raízes:

**Teorema 2.** Seja  $f$  uma função contínua no intervalo fechado  $[a, b]$  tal que  $f(a)f(b) < 0$ . Se  $f'(x)$  existir e preservar o sinal no intervalo aberto  $(a, b)$ , então existirá um único ponto  $x = r$  entre  $a$  e  $b$  tal que  $f(r) = 0$ .

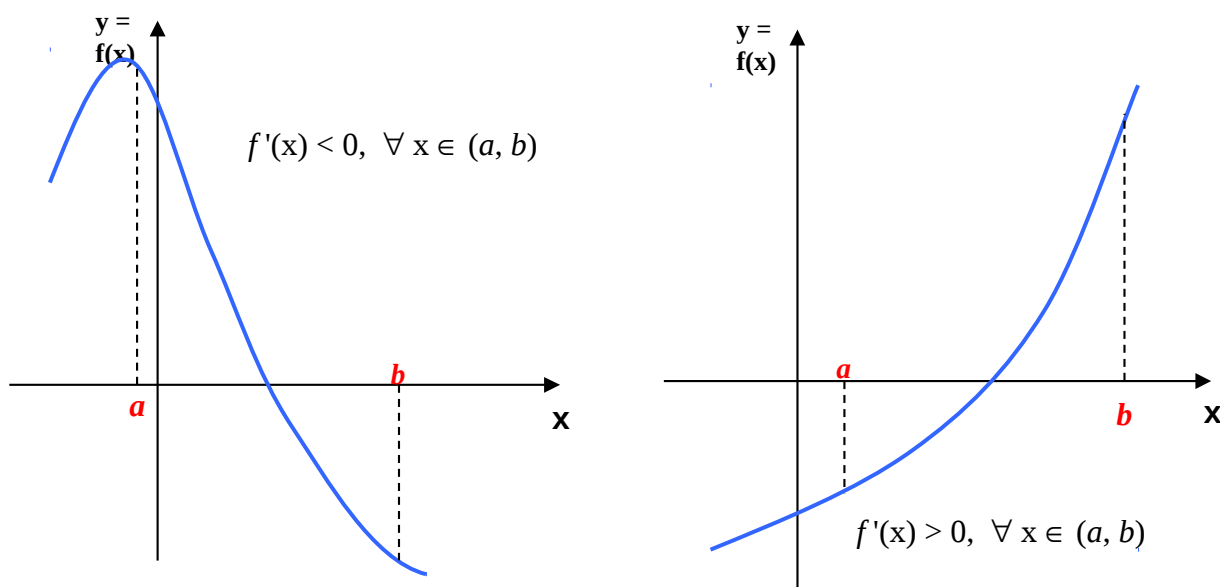
**Demonstração:** O TVI garante que existe  $r$  entre  $a$  e  $b$  tal que  $f(r) = 0$ . Suponha, por exemplo, que  $f'(x) > 0$ , para todo  $x$  em  $(a, b)$ . Nesse caso, a função  $f$  será estritamente crescente (\*) e poderá cortar o eixo horizontal (Ox) uma única vez (veja **Figura 3**).



**Observação (\*):** Para ajudar na compreensão da demonstração anterior, o Teorema do Valor Médio e o Teorema de Rolle podem ser consultados (referências [1 – 5] exibidas na página anterior).

O Teorema 2 também pode ser demonstrado supondo-se, por contradição, que existam  $r$  e  $s$  distintos tais que  $f(r) = f(s) = 0$ . Então, pelo Teorema de Rolle, existiria  $c$  entre  $r$  e  $s$  tal que  $f'(c) = 0$ . Mas isso seria um absurdo porque a derivada de  $f$  ou é positiva ou é negativa.

### Interpretação Geométrica do Teorema 2 (Figura 3):



**Figura 3: Interpretação geométrica do Teorema 2**

**Exemplo 1. (Aplicação do Teorema 2)** Considere  $f(x) = x^3 - 9x + 3$ . Como mostrado anteriormente, podemos isolar cada uma das três raízes de  $f(x) = 0$  nos seguintes intervalos  $[-4, -3]$ ,  $[0, 1]$  e  $[2, 3]$ . Note que  $f'(x) = 3x^2 - 9$  (parábola côncava para cima que possui raízes  $-\sqrt{3}$  e  $\sqrt{3} \approx 1,73$ ). Assim, se  $x < -\sqrt{3}$  ou  $x > \sqrt{3}$  então  $f'(x) > 0$ . Por outro lado, se  $-\sqrt{3} < x < \sqrt{3}$  então  $f'(x) < 0$ . Portanto, nos intervalos  $[-4, -3]$  e  $[2, 3]$  tem-se que  $f'(x) < 0$ ; no intervalo  $[0, 1]$  tem-se que  $f'(x) > 0$ . Dessa forma, pelo Teorema 2, em cada intervalo exibido anteriormente existe uma única raiz. ■

### Fase II – Refinamento de Intervalo

Partindo-se de  $I_0 = [a, b]$ , queremos obter intervalos  $I_n = [a_n, b_n]$ ,  $n \in \mathbb{N}$ , que contêm a raiz  $r$  da equação  $f(x) = 0$ , de modo que  $I_n \subset I_{n-1}$  e o comprimento de  $I_n$  tenda a zero. Isso pode ser feito até que se obtenha uma aproximação da raiz com precisão  $\varepsilon$  prefixada.

## Fase II – Refinamento da raiz (processo iterativo)

Partindo-se de uma aproximação  $r_0$  da raiz, queremos gerar um processo iterativo (uma sequência) caracterizado pelo termo geral  $r_n = \varphi(r_{n-1})$ ,  $n \in \mathbb{N}$ . A função  $\varphi$  é chamada de função de iteração. O processo iterativo deve ser convergente, isto é,  $\lim_{n \rightarrow \infty} r_n = r$ .

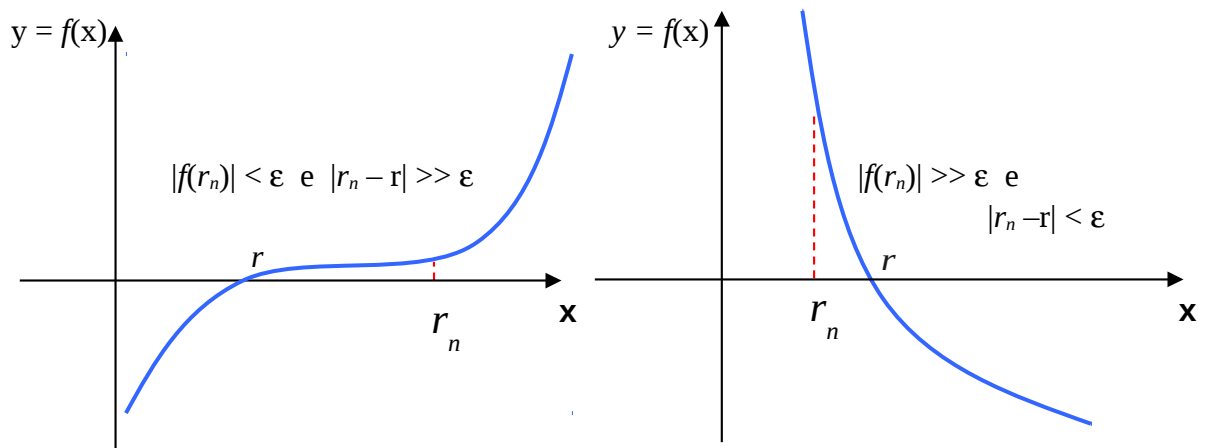
**Observação:** Todo processo iterativo necessita de um critério de parada. Os critérios mais utilizados são: (i) Erro Absoluto; (ii) Erro Relativo e (iii) Valor Absoluto da função. Assim,

(i) se  $|r_n - r_{n-1}| < \varepsilon = 0.5 \times 10^{-s}$ , então  $r_n$  será uma aproximação da raiz,  $r$ , da equação  $f(x) = 0$ , com precisão de  $s$  casas decimais;

(ii) se  $\frac{|r_n - r_{n-1}|}{|r_n|} < \varepsilon = 0.5 \times 10^{-s}$ , então  $r_n$  será uma aproximação da raiz,  $r$ , da equação  $f(x) = 0$ , com precisão de  $s$  casas decimais;

(iii) se  $|f(r_n)| < \varepsilon = 0.5 \times 10^{-s}$ , então  $r_n$  será uma aproximação da raiz,  $r$ , da equação  $f(x) = 0$ , com precisão de  $s$  casas decimais.

Nem sempre as condições anteriores são satisfeitas simultaneamente. Veja os exemplos gráficos a seguir (**Figura 4**).



**Figura 4: Critérios de parada**

O problema descrito a seguir foi apresentado pelo professor José Manoel Balthazar para alunos de graduação em matemática da Universidade Estadual Paulista (Unesp), Rio Claro – São Paulo. A disciplina era Cálculo Numérico e o ano não me recordo muito

bem, provavelmente 1987, mas os alunos iniciaram o curso de graduação em 1986. Eu era um deles.

**Problema 1.** Uma teoria sobre a formação de nuvens diz que o surgimento de gotas a partir da umidade do ar é facilitada pela presença de impurezas na atmosfera. Tais impurezas são chamadas “núcleos de condensação” e agem como concentradores de umidade e podem crescer até o tamanho de gotas de nuvens. As propriedades importantes que determinam o crescimento desses núcleos são a curvatura de sua superfície e a concentração molar da solução (isto é, a quantidade de sais dissolvidos na água). Quanto menor o tamanho da gota, maior será sua tendência para evaporar, enquanto a maior concentração da solução faz com que as gotas cresçam. A uma dada umidade relativa do ar, uma gota está em equilíbrio se a tendência ao crescimento é contrabalançada pela tendência à evaporação. O diâmetro,  $d$ , que deve ter uma tal gota em equilíbrio, foi deduzido por Kohler e deve obedecer à equação

$$100 - UR = A/d - B/d^3,$$

onde  $UR$  é a umidade relativa do ar em porcentagem ( $UR < 100\%$ ) e  $A$  e  $B$  são parâmetros que dependem da tensão superficial do fluido que se condensa, da concentração molar e da dissociação eletrolítica da solução. Se o diâmetro da gotícula for maior do que o diâmetro de equilíbrio, ela cresce rapidamente. Dessa forma, se o diâmetro de equilíbrio for pequeno, maior será a probabilidade de formação de nuvens. Portanto, para se produzir chuva, a uma certa umidade relativa e ajustando-se os parâmetros  $A$  e  $B$ , tenta-se mudar a salinidade do ar para se obter uma solução pequena da equação de Kohler.

*Caso hipotético:*  $UR = 70\%$ ;  $A = 1,0$  e  $B = 1,3 \times 10^{-4}$ . Qual é o valor do diâmetro de equilíbrio?

### **Solução do Problema 1:**

Reescrevendo a equação, obtemos:  $f(d) = 30d^3 - 1d^2 + 0.00013 = 0$ .

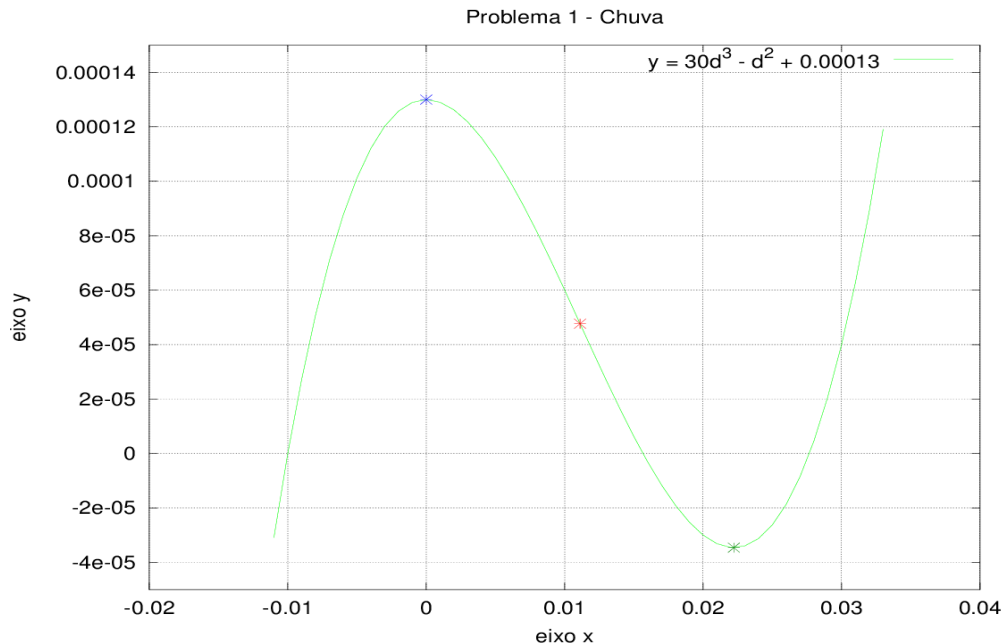
O esboço do gráfico da função  $f$  pode ser feito seguindo-se os procedimentos usuais vistos em disciplinas de Cálculo 1. Para isso, precisamos analisar tanto o sinal da derivada primeira quanto da derivada segunda da função  $f$ .

Observe que  $f'(d) = 90d^2 - 2d$  e  $f''(d) = 180d - 2$ . Então  $f'(d) = 0 \Leftrightarrow d = 0$  ou  $d = 1/45$ ;  $f''(d) = 0 \Leftrightarrow d = 1/90$ . Portanto, se  $d < 0$  ou  $d > 1/45$ , então  $f'(d) > 0$ ; se  $0 < d < 1/45$ , então  $f'(d) < 0$ . A derivada segunda satisfaz o seguinte:  $f''(d) < 0$ , se  $d < 1/90$  e  $f''(d) > 0$ , se  $d > 1/90$ .

Conclusão:  $d = 0$  é um ponto de máximo relativo da função  $f$ ;  $d = 1/45$  é um ponto de mínimo relativo da função  $f$ ;  $d = 1/90$  é um ponto de inflexão da função  $f$  (pontos indicados com “\*” no esboço do gráfico exibido a seguir). A função  $f$  é crescente nos intervalos  $(-\infty, 1/90)$  e  $(1/45, +\infty)$ , onde  $f'(d) > 0$ ;  $f$  é decrescente no intervalo  $(0, 1/45)$ , onde  $f'(d) < 0$ . O gráfico de  $f$  é côncavo para baixo no intervalo  $(-\infty, 1/90)$ ,



onde  $f''(d) < 0$ , e côncavo para cima no intervalo  $(1/90, +\infty)$ , onde  $f''(d) > 0$ . A seguir, veja o esboço do gráfico da função  $f$  (Figura 5).



**Figura 5. Gráfico do Problema da chuva**

Observando o gráfico, identificamos que o menor diâmetro de equilíbrio (a menor raiz positiva de  $f(d) = 0$ ) está no intervalo  $I_0 = [a, b] = [1/90, 1/45]$ . Para confirmar isso, basta utilizar o Teorema do Valor Intermediário. Note que  $f(1/90)f(1/45) < 0$  e  $f$  é uma função contínua.

Para encontrar uma aproximação da menor raiz positiva, vamos utilizar a técnica de refinamento de intervalo.

**OBJETIVO:** Partindo-se de  $I_0 = [a, b]$ , queremos obter  $I_n = [a_n, b_n]$ , de modo que (1)  $I_n \subset I_{n-1}$ ; (2)  $f(a_n)f(b_n) < 0$  e (3) o comprimento de  $I_n$  é a metade do comprimento de  $I_{n-1}$ . Isso pode ser feito até que se obtenha uma aproximação da raiz da equação  $f(d) = 0$  com precisão  $\varepsilon$  prefixada. Vamos utilizar  $\varepsilon = 0.5 \times 10^{-5}$  e o critério de parada dado pelo valor absoluto da função  $f$ .

A solução do exercício será obtida em quatro passos apresentados a seguir.

(i) Primeiro refinamento.

$f(d) = 30d^3 - d^2 + 0.00013$ ,  $d \in [1/90, 1/45]$ . Sabemos que  $f(1/90) > 0$  e  $f(1/45) < 0$ . No ponto médio,  $d_0 = (1/90 + 1/45)/2 = 1/60$ , tem-se que  $f(d_0) < 0$ . Então o primeiro intervalo refinado é  $I_1 = [1/90, 1/60]$ .

(ii) Segundo refinamento.

Sabemos que  $f(1/90) > 0$  e  $f(1/60) < 0$ . No ponto médio,  $d_1 = (1/90 + 1/60)/2 = 1/72$ , tem-se que  $f(d_1) > 0$ . Então o segundo intervalo refinado é  $I_2 = [1/72, 1/60]$ .

(iii) Avaliação de  $f$  no ponto médio do segundo intervalo refinado.

O ponto médio de  $I_2$  é dado por  $d_2 = (1/72 + 1/60)/2 = 11/720$ . Assim,  $f(d_2) \approx 0.357 \times 10^{-5}$ .

(iv) Precisão da aproximação.

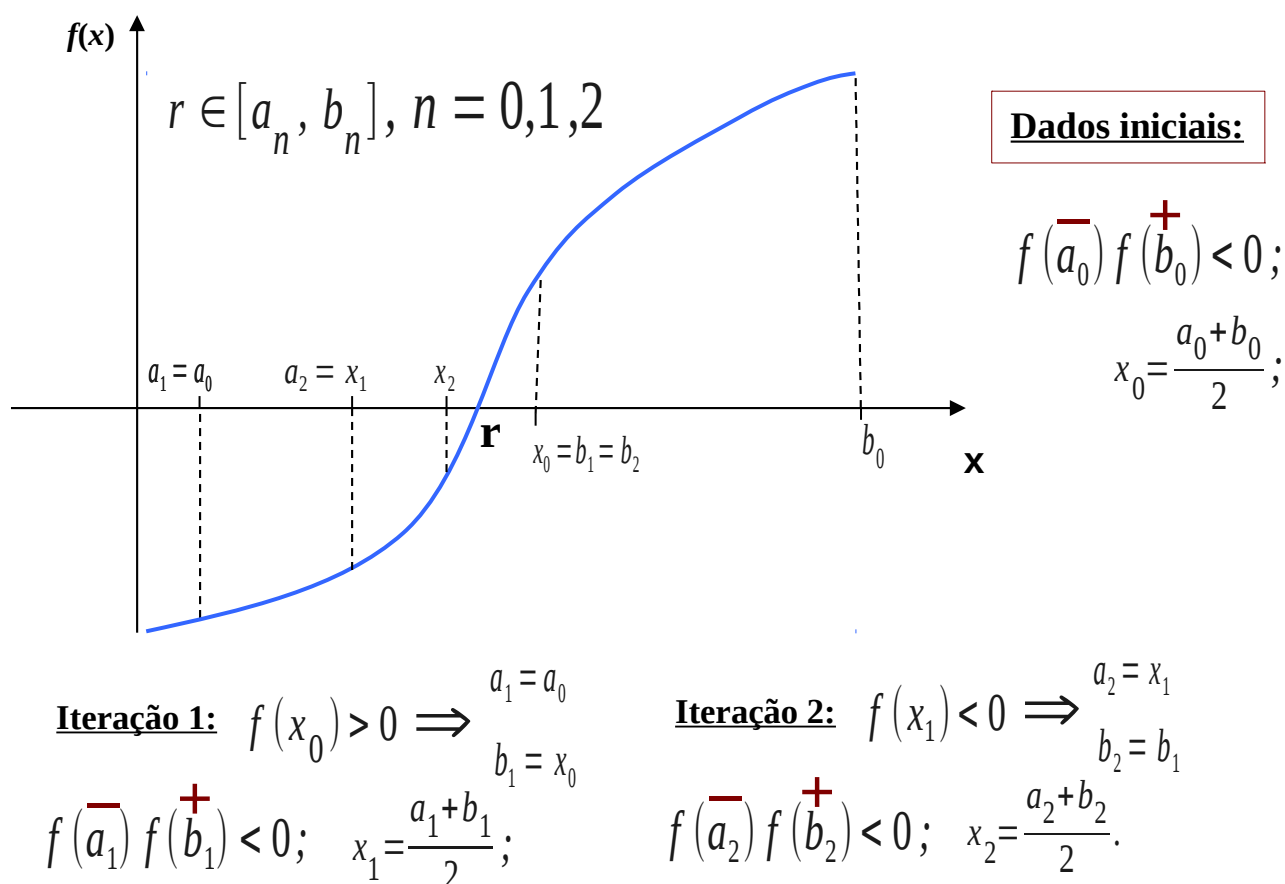
Note que  $|f(d_2)| \approx 0.357 \times 10^{-5} < 0.5 \times 10^{-5}$ . Portanto,  $d_2 = 0.0152778$  é uma aproximação da menor raiz com a precisão de 5 casas decimais. ■

## 4. Método da Bissecção

O método de refinamento de intervalo que utilizamos no **Problema 1** (final da seção anterior) é conhecido como Método da Bissecção. Tal método pode ser aplicado para se determinar uma raiz da equação  $f(x) = 0$ , sempre que a função contínua  $f$ , definida em um intervalo fechado  $[a, b]$ , satisfizer  $f(a)f(b) < 0$ .

**OBJETIVO DO MÉTODO DA BISSECÇÃO:** Partindo-se de  $I_0 = [a, b]$ , obter  $I_n = [a_n, b_n]$ , de modo que (1)  $I_n \subset I_{n-1}$ ; (2)  $f(a_n)f(b_n) < 0$  e (3) o comprimento de  $I_n$  é a metade do comprimento de  $I_{n-1}$ . Isso pode ser feito até que se obtenha uma aproximação da raiz da equação  $f(x) = 0$  com precisão  $\varepsilon$  prefixada.

Veja a seguir o esquema de aplicação desse método (**Figura 6**).



**Figura 6. Esquema do Método da Bissecção**

O algoritmo do Método da Bissecção é dado a seguir.

- 1) Dados iniciais: (i) intervalo inicial  $[a, b]$ ; (ii) precisão  $\varepsilon$ .
- 2) Se  $(b - a) < \varepsilon$  então escolha  $r$  qualquer em  $[a, b]$ . FIM.
- 3)  $k = 1$ .
- 4)  $x = (a + b)/2$ .
- 5) Se  $f(a)f(x) > 0$ , faça  $a = x$ . Vá para o passo 7.
- 6)  $b = x$ .

7) Se  $(b - a) < \varepsilon$ , escolha  $r$  qualquer em  $[a, b]$ . FIM.

8)  $k = k + 1$ . Volte ao passo 4.

No Método da Bissecção é possível calcular o número de iterações,  $n$ , que vai garantir que  $|x_n - r| < \varepsilon$ , onde  $x_n$  é o ponto médio do intervalo  $[a_n, b_n]$ , tal que  $f(a_n)f(b_n) < 0$ .

Para isso, observe que  $|x_n - r| < \frac{b_n - a_n}{2} = \frac{b_{n-1} - a_{n-1}}{2^2}$ , pois o comprimento do intervalo  $I_n$  é metade do comprimento do intervalo  $I_{n-1}$ . Por indução finita, podemos mostrar que  $|x_n - r| < \frac{b_0 - a_0}{2^{n+1}}$ . Dessa forma,  $\frac{b_0 - a_0}{2^{n+1}} < \varepsilon \rightarrow |x_n - r| < \varepsilon$ .

$$\text{Note que } \frac{b_0 - a_0}{2^{n+1}} < \varepsilon \rightarrow 2^{(n+1)} > \frac{b_0 - a_0}{\varepsilon} \rightarrow (n+1) \ln(2) > \ln\left(\frac{b_0 - a_0}{\varepsilon}\right) \rightarrow$$
$$n > \frac{\ln\left(\frac{b_0 - a_0}{\varepsilon}\right)}{\ln(2)} - 1.$$

**Exercício 1. (Número de iterações do Método da Bissecção)** Para  $x \in [a_0, b_0] = [1, 2]$ , determine o número mínimo de iterações para se obter uma aproximação, com precisão de pelo menos seis casas decimais ( $\varepsilon = 10^{-6}$ ), da raiz da equação:

$$f(x) = \sqrt{e^{-x} + \sin(x)} - x^2 \ln(x) = 0.$$

**Solução do Exercício 1:**

Note que a função  $f$  é contínua e satisfaz  $f(1)f(2) < 0$ . Então, utilizando a fórmula anterior,

$$n > \frac{\ln\left(\frac{b_0 - a_0}{\varepsilon}\right)}{\ln(2)} - 1 = \frac{\ln\left(\frac{2-1}{10^{-6}}\right)}{\ln(2)} - 1 = -1 + 6 \frac{\ln(10)}{\ln(2)} \simeq 18,93.$$

Portanto, a aproximação desejada é obtida com 19 iterações. ■

**Exercício 2. (Convergência do Método da Bissecção)** Mostre que o Método da Bissecção é convergente.

**Solução do Exercício 2** (Esta é uma demonstração construtiva do Teorema do Valor Intermediário).

Seja  $I_0 = [a_0, b_0]$ , tal que  $f(a_0)f(b_0) < 0$ , onde  $f$  é uma função contínua. De acordo com o Método da Bissecção, três sequências são construídas:  $(a_n)$ ,  $(b_n)$  e  $(x_n)$ , formadas,

respectivamente, pelos extremos inferiores, extremos superiores e pontos médios dos intervalos  $I_n = [a_n, b_n]$ , nos quais  $f$  troca de sinal, ou seja,  $f(a_n)f(b_n) < 0$ .

Por construção,  $I_n \subset I_{n-1}$  e  $b_n - a_n = \frac{b_{n-1} - a_{n-1}}{2}$ . Portanto,  $a_n \geq a_{n-1}$ ;  $b_n \leq b_{n-1}$  e  $x_n = \frac{a_n + b_n}{2}$ , com  $a_n \leq x_n \leq b_n$ ,  $\forall n \in \mathbb{N}$ . Dessa forma,

(i)  $(a_n)$  é não decrescente e limitada superiormente por  $b_0$ , ou seja,  $a_n \leq b_0$ ,  $\forall n \in \mathbb{N}$ ;

(ii)  $(b_n)$  é não crescente e limitada inferiormente por  $a_0$ , ou seja,  $b_n \geq a_0$ ,  $\forall n \in \mathbb{N}$ .

Por um teorema de convergência de sequências de números reais (veja a observação (\*) abaixo),  $(a_n)$  e  $(b_n)$  são convergentes. Além disso, como  $b_n - a_n = \frac{b_0 - a_0}{2^n}$  e

$\lim_{n \rightarrow \infty} \left(\frac{1}{2}\right)^n = 0$  então  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = r$ . Pelo Teorema do Confronto (veja a observação (\*\*) abaixo),  $\lim_{n \rightarrow \infty} x_n = r$ .

Agora fica faltando mostrar que  $f(r) = 0$ . Com efeito, note que  $f$  é contínua e, portanto,

$$\begin{aligned} f(a_n)f(b_n) < 0 &\rightarrow \lim_{n \rightarrow \infty} f(a_n)f(b_n) \leq 0 \rightarrow f\left(\lim_{n \rightarrow \infty} a_n\right)f\left(\lim_{n \rightarrow \infty} b_n\right) \leq 0 \rightarrow \\ &\rightarrow f(r)f(r) \leq 0. \end{aligned}$$

Dessa forma,  $0 \leq [f(r)]^2 \leq 0$ ; de onde segue que  $f(r) = 0$ . □



#### Saiba Mais

**Observação (\*):** Um importante teorema de convergência de sequências diz que toda sequência  $(x_n)$  monótona e limitada é convergente. Mais especificamente, tem-se que

(i)  $x_{n-1} \leq x_n$  e  $x_n \leq L$ ,  $\forall n \in \mathbb{N}$  (sequência não decrescente e limitada superiormente)  $\rightarrow$   
 $\lim_{n \rightarrow \infty} x_n = s = \sup\{A\} = \sup\{x_n; n \in \mathbb{N}\}$ , onde  $\sup\{A\}$  é o supremo do conjunto  $A$ , que é a menor cota superior de  $A$  (observe que  $L$  é uma cota superior de  $A$ );

(ii)  $x_n \leq x_{n-1}$  e  $x_n \geq M$ ,  $\forall n \in \mathbb{N}$  (sequência não crescente e limitada inferiormente)

$\rightarrow \lim_{n \rightarrow \infty} x_n = t = \inf\{A\} = \inf\{x_n; n \in \mathbb{N}\}$ , onde  $\inf\{A\}$  é o ínfimo do conjunto  $A$ ,

que é a maior cota inferior de  $A$  (observe que  $M$  é uma cota inferior de  $A$ ).

Esse teorema deverá ser apresentado para você nas disciplinas “Cálculo” e “Introdução à Análise”. As referências apresentadas na Seção 3 podem ser consultadas, também.



Saiba Mais

**Observação (\*\*):** O Teorema do Confronto é demonstrado em cursos de Análise na reta. O resultado importante desse Teorema é apresentado a seguir.

$$a_n \leq x_n \leq b_n, \forall n \in \mathbb{N} \text{ e } \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = r \Rightarrow \lim_{n \rightarrow \infty} x_n = r.$$

## 5. Método do Ponto Fixo

O Método do Ponto Fixo (MPF) ou Método Iterativo Linear (MIL) consiste em transformar a equação  $f(x) = 0$  em uma equação equivalente  $x = \varphi(x)$ , onde  $\varphi(x)$  é chamada de função de iteração.

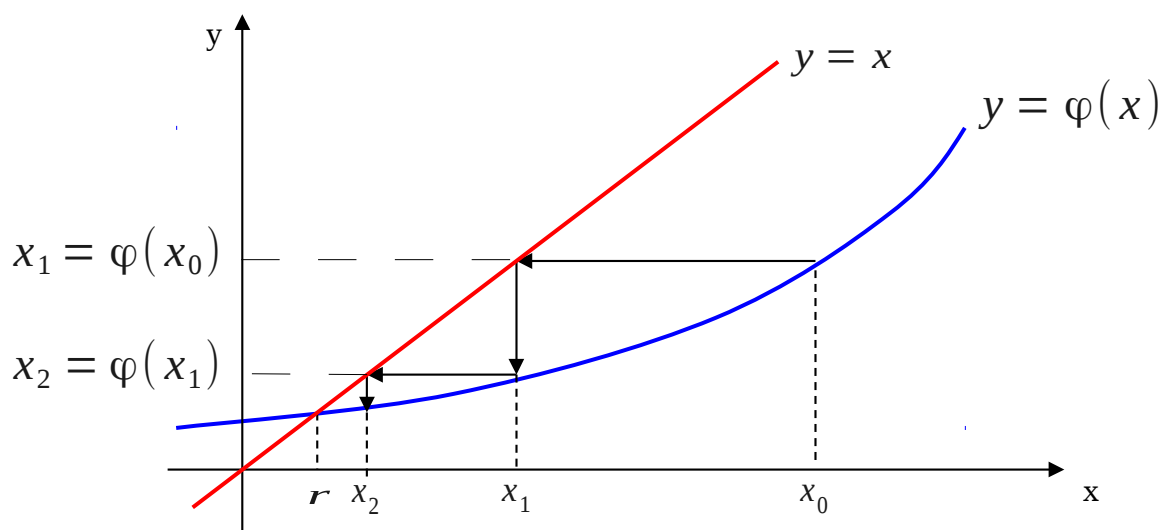
Conforme vimos no processo de refinamento de raiz ([Fase II – processo iterativo](#)), na Seção 3, dado  $x_0$  pertencente a um intervalo contendo a raiz,  $r$ , da equação  $f(x) = 0 \Leftrightarrow x = \varphi(x)$ , podemos considerar o seguinte processo iterativo:  $x_n = \varphi(x_{n-1})$ ,  $n \in \mathbb{N}$ . O processo iterativo deve ser convergente, isto é,

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \varphi(x_{n-1}) = r.$$

Observe que se a função de iteração for contínua então  $\lim_{n \rightarrow \infty} \varphi(x_{n-1}) = \varphi(\lim_{n \rightarrow \infty} x_{n-1})$ .

Dessa forma,  $r$  é um ponto fixo de  $\varphi$  (Veja a **Fig. 7**), ou seja,  $r = \varphi(r) \Leftrightarrow f(r) = 0$ .

### Interpretação Geométrica do Método do Ponto Fixo:



**Figura 7. Interpretação geométrica do Método do Ponto Fixo**

**Exemplo 2. (funções de iteração)** Dada a equação  $x^3 + x - 6 = 0$ , podemos encontrar as seguintes funções de iteração:

(i)  $x^3 + x - 6 = 0 \rightarrow x = 6 - x^3 \rightarrow \varphi(x) = 6 - x^3$ ;

(ii)  $x^3 + x - 6 = 0 \rightarrow x^3 = 6 - x \rightarrow x = \sqrt[3]{6 - x} \rightarrow \varphi(x) = \sqrt[3]{6 - x}$  ;

(iii)  $x^3 + x - 6 = 0 \rightarrow x(x^2 + 1) = 6 \rightarrow x = \frac{6}{x^2 + 1} \rightarrow \varphi(x) = \frac{6}{x^2 + 1}$  ;

(iv)  $x^3 + x - 6 = 0 \rightarrow x^3 = 6 - x \rightarrow x = \frac{6 - x}{x^2} = \frac{6}{x^2} - \frac{1}{x}$ .

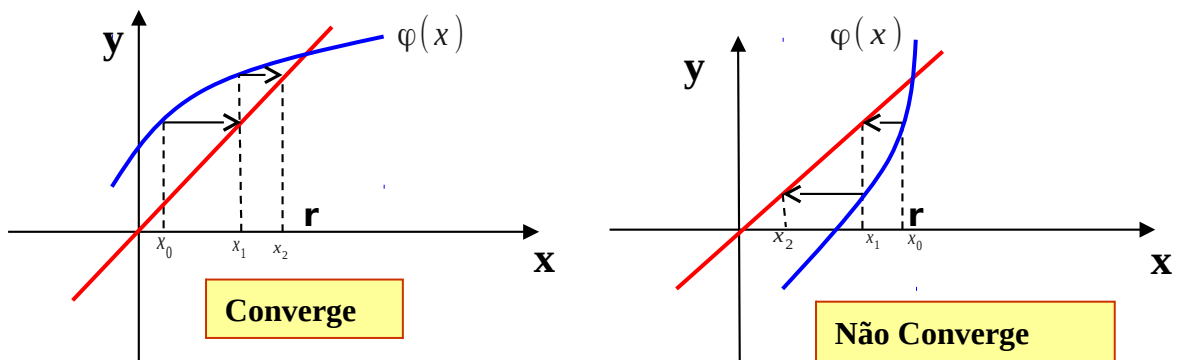
Nem todas as funções de iteração irão gerar processos iterativos convergentes. Por exemplo, considerando-se  $x_0 = 1.5$ , a função dada no primeiro item (i) será divergente e a função dada no item (ii) será convergente. De fato,

$$\begin{aligned}\varphi_1(x) &= 6 - x^3; \\ x_1 &= 6 - 1.5^3 = 2.625 \\ x_2 &= 6 - 2.625^3 \approx -12.0879 \\ x_3 &= 6 - (-12.0879)^3 \approx 1772.2516 \\ &\vdots \\ &\text{n\~ao converge}\end{aligned}$$

$$\begin{aligned}\varphi_2(x) &= \sqrt[3]{6-x}; \\ x_1 &= \sqrt[3]{6-1.5} \approx 1.65096 \\ x_2 &= \sqrt[3]{6-1.65096} \approx 1.63229 \\ x_3 &= \sqrt[3]{6-1.63229} \approx 1.63462 \\ &\vdots \\ &\text{converge}\end{aligned}$$

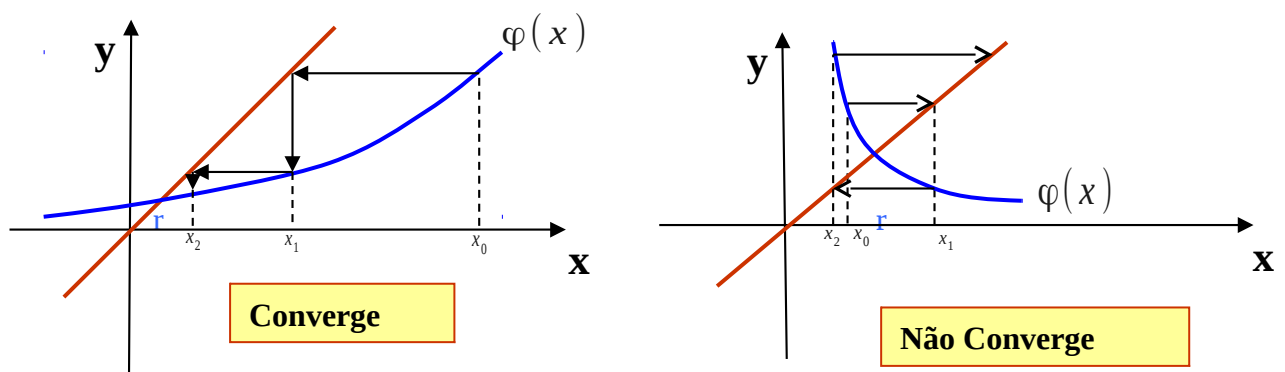
A seguir s\~ao exibidas alguns gr\'aficos mostrando os comportamentos das sequ\~encias geradas pelo M\'etodo do Ponto Fixo (**Figuras 8 e 9**).

Interpreta\~ao Geom\'etrica da Converg\~encia e N\~ao converg\~encia do M\'etodo do Ponto Fixo:



**Figura 8. Comportamentos do MPF: converge (quadro 1) e diverge (quadro 2)**

Interpreta\~ao Geom\'etrica da Converg\~encia e N\~ao converg\~encia do M\'etodo do Ponto Fixo:



**Figura 9. Comportamentos do MPF: converge (quadro 1) e diverge (quadro 2)**



**Teorema 3 (Condição Suficiente de Convergência do MPF).** Seja  $r$  a única raiz da equação  $f(x) = 0$ , pertencente ao intervalo  $I = [r - d, r + d]$ ,  $d > 0$ . Seja  $\varphi(x)$  uma função de iteração para a equação  $f(x) = 0$ . Suponha que

- (i)  $\varphi$  e  $\varphi'$  forem contínuas em  $I$ ;
- (ii)  $|\varphi'(x)| \leq M < 1$ ;  $\forall x \in I$ ;
- (iii)  $x_0 \in I$ .

Então  $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \varphi(x_{n-1}) = r$ .

**Demonstração:** Como a função  $\varphi$  é contínua no intervalo fechado  $I$  e derivável no intervalo aberto  $(r - d, r + d)$ , o Teorema do Valor Médio (mencionado na Seção 3) pode ser utilizado. Assim, dados  $x$  e  $t$  em  $I$ ,  $x \neq t$ , tem-se que  $\frac{\varphi(x) - \varphi(t)}{x - t} = \varphi'(c)$  para algum  $c$  entre  $x$  e  $t$ . Portanto,  $|\varphi(x) - \varphi(t)| = |\varphi'(c)| |x - t| \leq M |x - t|$ , porque, por hipótese (ii),  $|\varphi'(x)| \leq M < 1$ ,  $\forall x \in I$ . Dessa forma, podemos mostrar que se  $x_0 \in I$  então  $x_1 = \varphi(x_0) \in I$ .

Com efeito, considerando o argumento do parágrafo anterior e observando-se que  $x \in I$  se, e somente se,  $|x - r| \leq d$ , tem-se que  $|x_1 - r| = |\varphi(x_0) - \varphi(r)| \leq M |x_0 - r| < |x_0 - r| \leq d$ . Portanto,  $x_1 \in I$ . Utilizando um argumento de Indução Finita, é fácil provar que  $x_n = \varphi(x_{n-1}) \in I$ ,  $\forall n \in \mathbb{N}$ .

Utilizando o mesmo argumento do primeiro parágrafo (Teorema do Valor Médio e hipótese (ii)), obtém-se que  $0 \leq |x_n - r| = |\varphi(x_{n-1}) - \varphi(r)| \leq M |x_{n-1} - r|$ . Repetindo o raciocínio, obtém-se que  $|x_{n-1} - r| = |\varphi(x_{n-2}) - \varphi(r)| \leq M |x_{n-2} - r|$ , de onde segue que  $|x_n - r| \leq M^2 |x_{n-2} - r|$ . Novamente, utilizando indução finita, podemos concluir que  $0 \leq |x_n - r| \leq M^n |x_0 - r|$ . Sabendo que  $\lim_{n \rightarrow \infty} M^n = 0$ , pois  $M < 1$ , segue-se, do Teorema do Confronto (veja o final da Seção 4), que  $\lim_{n \rightarrow \infty} |x_n - r| = 0$ , ou seja,  $\lim_{n \rightarrow \infty} x_n = r$ . □

O algoritmo do Método do Ponto Fixo é dado a seguir:

- 1) Dados iniciais: (i)  $x_0$  : aproximação inicial, (ii) precisões  $\varepsilon_1$  e  $\varepsilon_2$ .
- 2) Se  $|f(x_0)| < \varepsilon_1$ , então faça  $r = x_0$ . FIM.

3)  $k = 1$ .

4)  $x_1 = \varphi(x_0)$ .

5) Se  $|f(x_1)| < \varepsilon_1$  ou se  $|x_1 - x_0| < \varepsilon_2$ , faça  $r = x_1$ . FIM (critério do erro absoluto)

6)  $x_0 = x_1$ .

7)  $k = k + 1$ . Volte ao passo 4.

### Ordem de Convergência de métodos iterativos

**Definição (ordem de convergência):** Seja  $(x_n)$  uma sequência que converge para um número  $r$  e seja  $e_n = x_n - r$  o erro na iteração  $n$ . Se existir um número  $p \geq 1$  e uma constante  $C > 0$ , tais que  $\lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|^p} = C$ , então  $p$  é a ordem de convergência da sequência e  $C$  é a constante assintótica do erro.

**Observação:** Da definição anterior, podemos considerar a seguinte estimativa:  $|e_{n+1}| \approx C|e_n|^p$ , ou seja, o valor absoluto do erro na iteração  $n + 1$  é (aproximadamente) proporcional ao valor absoluto do erro na iteração  $n$  elevado à potência  $p$ .


**Exercício 3. (Ordem Linear do MPF)** Mostre que o Método do Fixo possui convergência Linear.

**Solução do Exercício 3:** Pela definição do método iterativo e utilizando o Teorema do Valor Médio, segue-se que  $x_{n+1} - r = \varphi(x_n) - \varphi(r) = \varphi'(c_n)(x_n - r)$ , onde  $c_n$  está entre  $x_n$

e  $r$ . Assim,  $|e_{n+1}| = |e_n| |\varphi'(c_n)| \rightarrow \lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|} = \lim_{n \rightarrow \infty} |\varphi'(c_n)| \rightarrow$

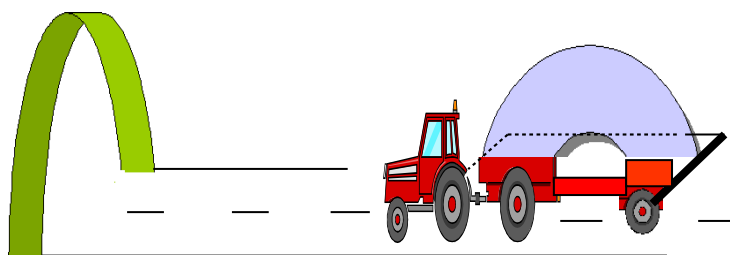
$\rightarrow \lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|} = |\varphi'(r)|$ . Na última passagem, utilizamos o Teorema do Confronto e o

fato de  $\varphi'$  ser uma função contínua.

Note que  $c_n$  está entre  $x_n$  e  $r$  e, além disso,  $x_n$  converge para  $r$ . Portanto,  $0 < C = |\varphi'(r)| < 1$ , o que garante a convergência linear do Método do Ponto Fixo. Daí vem o outro nome desse método: Método Iterativo Linear. 

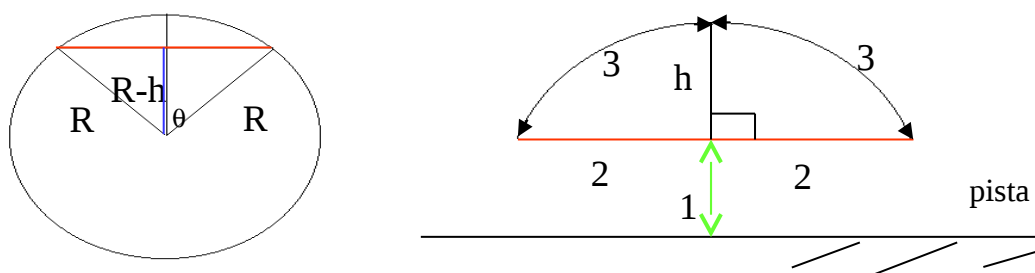
Vamos finalizar essa seção com mais um problema retirado da lista do professor Balthazar (mesmo estilo do **Problema 1**, apresentado no final da Seção 3), com pequenas adaptações (Veja a **Figura 10**).

**Problema 2.** Um caminhão transporta uma peça com 4,0 m de largura em formato de um arco de circunferência medindo 6,0 m de comprimento. À sua frente encontra-se um túnel de 3,0 metros de altura. Sabendo que a base do caminhão está a 1,0 m de distância da pista, responda: o caminhão passará pelo túnel?



**Figura 10. Problema adaptado de um exercício proposto, em 1987, pelo professor Balthazar – Unesp – Rio Claro**

**Solução do Problema 2:** Vamos utilizar conceitos da Geometria Euclidiana Plana para resolver esse problema. Para isso, observe as figuras abaixo (**Figura 11**).



**Figura 11. Modelo Geométrico**

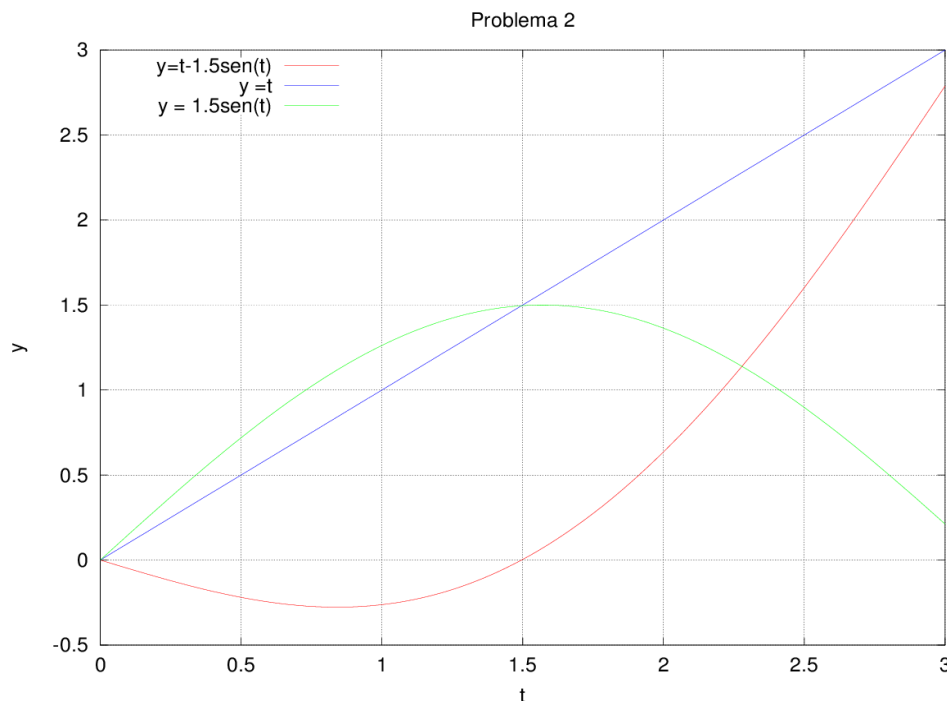
Com os dados fornecidos pelo problema, podemos construir uma circunferência de raio  $R$  (a ser determinado) e considerar o arco de 6 metros que dará origem a uma corda de 4 metros (linha de cor vermelha na **Figura 11**). O objetivo é encontrar a altura  $h$  (ponto mais alto da peça transportada) para saber se o caminhão irá passar ou não pelo túnel.

É fácil mostrar que a reta que passa pelo centro da corda e pelo centro da circunferência é perpendicular à corda (basta usar congruência de triângulos; caso Lado-Lado-Lado). Além disso, utilizando o teorema que garante que toda reta tangente a um ponto da circunferência é perpendicular ao raio que tem uma das extremidades no ponto de tangência, podemos concluir que o ponto mais alto da peça transportada é a medida do segmento de reta que passa perpendicularmente pelo centro da corda e divide o arco de circunferência em dois arcos de mesmo comprimento (veja a **Figura 11**, imagem do lado direito).

Note que o arco de 6 m corresponde a um ângulo central, digamos  $2\theta$ , que tem medida, em radianos, igual a  $6/R$ , ou seja,  $2\theta = 6/R$ ; logo  $\theta = 3/R$ .

De acordo com as relações trigonométricas em um triângulo retângulo, tem-se que  $\sin(\theta) = 2/R$  (note que a medida da hipotenusa é igual a  $R$ ). Como  $R = 3/\theta$ , segue-se que  $\sin(\theta) = 2\theta/3$ , ou seja,  $\theta = 1,5 \sin(\theta)$ . Note que  $\theta = 0$  é uma solução da equação anterior, mas não é a solução que precisamos para o problema.

A **Figura 12** exibe os gráficos associados às equações:  $f(t) = t - 1,5 \sin(t)$  e  $t = 1,5 \sin(t)$ . O gráfico da função  $f$  está exibido na cor verde. Estão exibidos nas cores azul e vermelho os gráficos de  $y = t$  e  $y = 1,5 \sin(t)$ , respectivamente.



**Figura 12. Gráficos associados ao Problema 2**

De acordo com o gráfico (**Fig. 12**), o ponto fixo da função  $\phi(t) = 1,5 \sin(t)$ , ou o zero da função  $f$ , encontra-se no intervalo  $[1 \ 1,5]$ . Para confirmar esse resultado, basta utilizar o

Teorema do Valor Intermediário. É fácil verificar que  $f$  é contínua e satisfaz  $f(1)f(1.5) < 0$ .

Observe que  $\phi'(t) = 1.5 \cos(t)$ , com  $t$  pertencente ao intervalo  $[1 \ 1.5]$ . Nesse intervalo, a função cosseno é decrescente e positiva. Portanto  $|\phi'(t)| = 1.5 \cos(t) \leq 1.5 \cos(1) < 1$ . De acordo com o **Teorema 3**, o Método do Ponto Fixo é convergente.

Vamos utilizar o Método do Ponto Fixo para determinar o valor do ângulo  $\theta$ , tal que  $\theta = 1.5 \sin(\theta)$ . Para isso, considere  $\theta_0 = (1 + 1.5)/2 = 1.25$  e calcule as demais aproximações do ponto fixo através da sequência:  $\theta_n = \phi(\theta_{n-1})$ . Isso dará origem aos seguintes valores (não se esqueça de calcular os valores da função seno em radianos):

$$\begin{aligned}\theta_1 &= 1.5 \sin(\theta_0) = 1.423476929; \\ \theta_2 &= 1.5 \sin(\theta_1) = 1.483752164; \\ \theta_3 &= 1.5 \sin(\theta_2) = 1.494321072; \\ \theta_4 &= 1.5 \sin(\theta_3) = 1.495615789; \\ \theta_5 &= 1.5 \sin(\theta_4) = 1.495762911; \\ \theta_6 &= 1.5 \sin(\theta_5) = 1.495779471; \\ \theta_7 &= 1.5 \sin(\theta_6) = 1.495781332; \\ \theta_8 &= 1.5 \sin(\theta_7) = 1.495781542.\end{aligned}$$

Calculando o erro relativo, na iteração 8, obtemos:  $ER = \frac{|\theta_8 - \theta_7|}{|\theta_8|} \approx 1.402031 \times 10^{-7}$ .

Portanto,  $\theta_8$  é uma aproximação do ponto fixo da função  $\phi$  com pelo menos 6 algarismos significativos ( $ER < 0.5 \times 10^{-6}$ ).

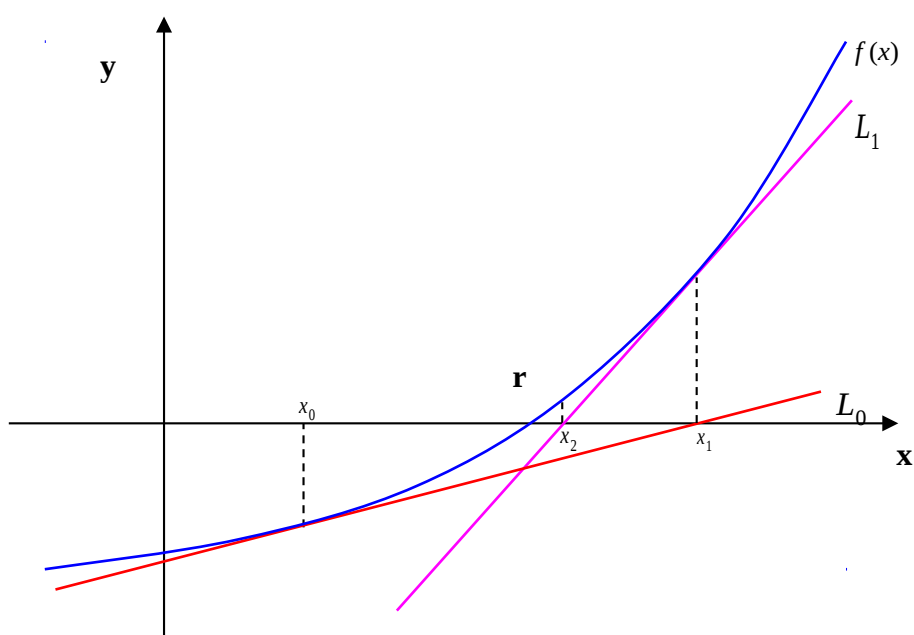
Utilizando a aproximação anterior, obtemos que  $R = 3/\theta \approx 2.005640473$ . Assim, pelo Teorema de Pitágoras (veja a **Figura 11**, desenho do lado esquerdo),

$$\begin{aligned}2^2 + (R - h)^2 &= R^2 \rightarrow (R - h)^2 = R^2 - 4 \rightarrow R - h = (R^2 - 4)^{1/2} \rightarrow \\ \rightarrow h &= R - (R^2 - 4)^{1/2} \rightarrow h = 1.855328436.\end{aligned}$$

Lembre-se de que a distância da base do caminhão até a pista é igual a 1m. Portanto, a altura máxima ( $h$ ) da peça que está sendo transportada somada a 1 dará aproximadamente 2,86. Conclusão: o caminhão passa pelo túnel, mas a folga é algo em torno de 14 cm. ■

## 6. Método de Newton – Raphson

O método de Newton – Raphson é obtido a partir da seguinte construção: tome  $(x_{n-1}, f(x_{n-1}))$  sobre o gráfico de  $f$  e considere o coeficiente angular  $m = f'(x_{n-1})$ ; o próximo valor da sequência gerada pelo método de Newton-Raphson é  $x_n = x_{n-1} - [f(x_{n-1})/f'(x_{n-1})]$ , obtido da interseção da reta tangente  $y = f(x_{n-1}) + m(x - x_{n-1})$  com o eixo  $x$ . Essa construção está exibida na **Figura 13**. (Observação: As retas  $L_0$  e  $L_1$  são as retas tangentes ao gráfico de  $f$ , respectivamente, nos pontos  $(x_0, f(x_0))$  e  $(x_1, f(x_1))$  e  $r$  é tal que  $f(r) = 0$ .)



**Figura 13. Interpretação Geométrica do Método de Newton – Raphson**

**Teorema 4 (Convergência do Método de Newton – Raphson).** Sejam  $f(x)$ ,  $f'(x)$ ,  $f''(x)$  contínuas num intervalo  $I$  que contém a raiz  $r$  da equação  $f(x) = 0$ . Suponha que  $f'(r) \neq 0$ . Dessa forma, existirá um intervalo  $I_2 \subset I$ , que contém a raiz  $r$ , tal que se  $x_0 \in I_2$

$$\text{então } \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} = r.$$

**Demonstração:** Vamos mostrar que as hipóteses do **Teorema 3** valem para a função de iteração do método de Newton – Raphson,  $\varphi(x) = x - \frac{f(x)}{f'(x)}$ .

Primeiramente, observe que  $r = \varphi(r) \Leftrightarrow f(r) = 0$ . Agora, note que  $f'(x)$  é contínua em  $I$  e  $f'(r) \neq 0$ ; logo, existe um intervalo  $I_1 \subset I$  tal que  $f'(x) \neq 0$ , para todo  $x \in I_1$ .

Esse resultado é conhecido, nos livros de Cálculo, como Teorema da permanência do sinal. Em nosso caso, basta utilizar a definição de continuidade, no ponto  $r$ , da função  $|f'(x)|$ . Isso, juntamente com o fato de  $f'(x)$  e  $f''(x)$  serem contínuas, garante que as funções

$$\varphi(x) = x - \frac{f(x)}{f'(x)} \quad \text{e} \quad \varphi'(x) = 1 - \frac{[f'(x)]^2 - f(x)f''(x)}{[f'(x)]^2} = \frac{f(x)f''(x)}{[f'(x)]^2}$$

sejam contínuas no intervalo  $I_1$ .

Agora, note que  $\varphi'(r) = 0$ , pois  $f(r) = 0$ . Dessa forma, utilizando a continuidade de  $\varphi'$  no ponto  $r$ , pode-se mostrar a existência de um intervalo  $I_2 \subset I_1$ , contendo  $r$ , tal que  $|\varphi'(x)| \leq M < 1$ ,  $\forall x \in I_2$ . Assim, as hipóteses do **Teorema 3** são satisfeitas e, portanto, o método de Newton – Raphson converge para  $r$ . ■



Se  $\varphi'$  é contínua em  $r$ , então dado  $\varepsilon = M = 1/2$  existe  $\delta > 0$  tal que  $|x - r| < \delta \rightarrow |\varphi'(x) - \varphi'(r)| < M$ . Como  $\varphi'(r) = 0$  então  $|\varphi'(x)| < M$ , para todo  $x \in (r - \delta, r + \delta)$ . Pode-se escolher  $\delta_2 \leq \delta$ , tal que  $I_2 = [r - \delta_2, r + \delta_2] \subset I_1$  (intervalo mencionado na demonstração do Teorema 4). Desse modo,  $|\varphi'(x)| \leq M < 1$ ,  $\forall x \in I_2$ .

O algoritmo do Método de Newton – Raphson é dado a seguir:

- 1) Dados iniciais: (i)  $x_0$  : aproximação inicial; (ii) precisões  $\varepsilon_1$  e  $\varepsilon_2$ .
- 2) Se  $|f(x_0)| < \varepsilon_1$  então faça  $r = x_0$ . FIM.
- 3)  $k = 1$ .

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

5) Se  $|f(x_1)| < \varepsilon_1$  ou se  $|x_1 - x_0| < \varepsilon_2$ , faça  $r = x_1$ . FIM (critério do erro absoluto)

6)  $x_0 = x_1$ .

7)  $k = k + 1$ . Volte ao passo 4.

**Exercício 4. (Ordem quadrática do Método de Newton - Raphson)** Mostre que o Método de Newton – Raphson possui convergência quadrática.

**Solução do Exercício 4:** Vamos usar a mesma notação utilizada na definição de ordem de convergência, dada na Seção 5. Dessa forma, seja  $e_n = x_n - r$  e considere o desenvolvimento de Taylor das funções  $f$  e  $f'$ :

$$f(r + e_{n-1}) = f(r) + f'(r)e_{n-1} + \frac{f''(c_{n-1})e_{n-1}^2}{2!};$$

$$f'(r + e_{n-1}) = f'(r) + f''(d_{n-1})e_{n-1},$$

onde  $c_{n-1}$  e  $d_{n-1}$  estão entre  $r$  e  $x_{n-1}$ .

$$\text{Note que } x_n - r = x_{n-1} - r - \frac{f(x_{n-1})}{f'(x_{n-1})} \rightarrow e_n = e_{n-1} - \frac{f(e_{n-1} + r)}{f'(e_{n-1} + r)}.$$

$$\text{Portanto, } e_n = \frac{e_{n-1}[f'(r) + f''(d_{n-1})e_{n-1}] - [f(r) + f'(r)e_{n-1} + \frac{f''(c_{n-1})e_{n-1}^2}{2!}]}{f'(e_{n-1} + r)}.$$

Logo,

$$e_n = \frac{[f''(d_{n-1}) - \frac{f''(c_{n-1})}{2}]e_{n-1}^2}{f'(x_{n-1})} \rightarrow \frac{e_n}{e_{n-1}^2} = \frac{f''(d_{n-1}) - \frac{f''(c_{n-1})}{2}}{f'(x_{n-1})}. \text{ Assim}$$

$$\lim_{n \rightarrow \infty} \frac{|e_n|}{|e_{n-1}|^2} = \lim_{n \rightarrow \infty} \left| \frac{f''(d_{n-1}) - \frac{f''(c_{n-1})}{2}}{f'(x_{n-1})} \right| = \frac{\left| \frac{f''(r)}{2} \right|}{|f'(r)|} = \left| \frac{f''(r)}{2f'(r)} \right|.$$

Na última passagem, utilizamos que  $c_{n-1}$  e  $d_{n-1}$  estão entre  $r$  e  $x_{n-1}$  e que a função  $f''$  é contínua. Depois, aplicamos o Teorema do Confronto nas sequências com termos gerais  $c_{n-1}$  e  $d_{n-1}$  para concluir que ambas convergem para  $r$ , já que  $x_{n-1}$  converge para  $r$ . Isso garantiu a convergência quadrática do Método de Newton – Raphson. ■



**Exemplo 3.** Para verificar que o Método de Newton - Raphson converge mais rápido que o Método do Ponto Fixo, considere a mesma equação dada no **Problema 2** (apresentado no final da Seção 5),  $x = 1,5 \sin(x) \Leftrightarrow f(x) = x - 1,5 \sin(x) = 0$ , e o mesmo valor inicial  $x_0 = 1.25$ . As iterações serão realizadas com a função  $\varphi(x) = x - \frac{f(x)}{f'(x)}$ . Note que  $f'(x) = 1 - \cos(x)$ . Assim,  $x_n = \varphi(x_{n-1})$  dará origem aos seguintes valores (não se esqueça de calcular os valores das funções seno e cosseno em radianos):

$$x_1 = \varphi(x_0) = 1.579167955;$$

$$x_2 = \varphi(x_1) = 1.500929896;$$

$$x_3 = \varphi(x_2) = 1.495803715;$$

$$x_4 = \varphi(x_3) = 1.495781569;$$

$$x_5 = \varphi(x_4) = 1.495781568.$$

Observe que na quinta iteração o erro relativo é dado por  $ER = \frac{|x_5 - x_4|}{|x_5|} \approx 0.066855 \times 10^{-8}$ . Além disso,  $|f(x_5)| = |x_5 - 1,5 \sin(x_5)| \approx 0.197 \times 10^{-9}$  (resultado muito menor do que o obtido com o MPF – Confira!).

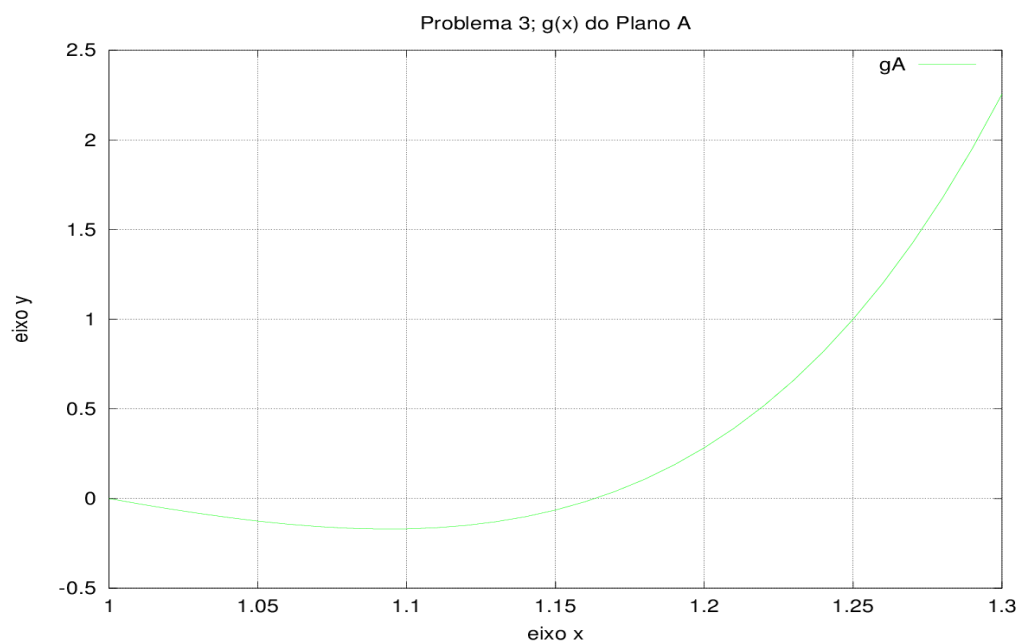
Vamos finalizar essa seção com a apresentação de um problema que foi adaptado da Monografia de Especialização “Modelagem de Problemas de Matemática Financeira e suas Resoluções Utilizando Técnicas Matemáticas e Computacionais”, de Leone Alves Leite, então aluna da Faculdade de Matemática da Universidade Federal de Uberlândia. A monografia foi apresentada, em agosto de 2005, para a Banca Examinadora composta pelos professores Edson Agustini, Marcos Antônio da Câmara e César Guilherme de Almeida (orientador).

**Problema 3.** Uma loja de eletrodomésticos oferece dois planos de financiamento para um produto cujo preço à vista é R\$ 1.200,00. Plano A: entrada de R\$ 200,00 mais 7 prestações mensais de R\$ 250,00. Plano B: entrada de R\$ 200,00 mais 10 prestações mensais de R\$ 200,00. Fica a dúvida: qual plano apresenta a menor taxa de juros?

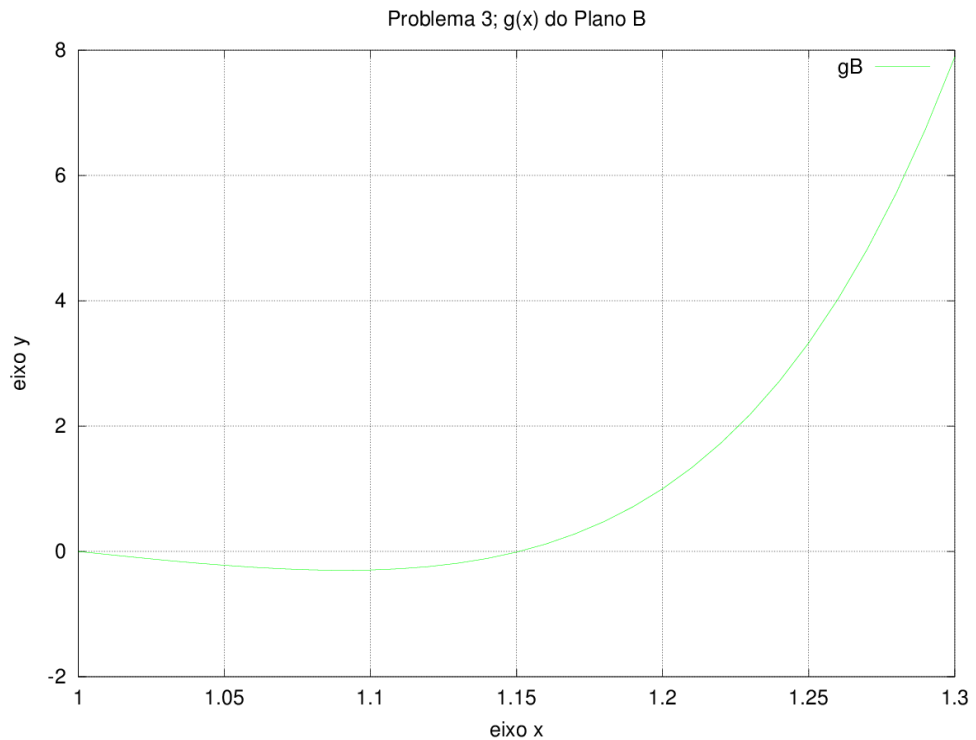
**Solução:** Toda vez que alguém for financiar um produto, ou um bem qualquer, deve ficar atento à equação:  $F \cdot j = P \cdot [1 - (1 + j)^{-n}]$ , onde **F** é o **Valor Financiado**; **j** é a **Taxa Mensal de Juros**; **P** é o **Valor Mensal da Prestação** e **n** é o **Número de Prestações**. O mais importante é saber calcular a taxa de juros.

OBSERVAÇÕES: (i) Se  $x = 1 + j$ , então a equação apresentada no início dessa solução pode ser reescrita como:  $F \cdot (x - 1) = P \cdot [1 - x^{-n}]$ , onde  $F = \text{R\$}1.000,00$  (que é o preço à vista menos a entrada) em ambos os planos. (ii) Seja  $K = F/P$ . Assim, a seguinte equação é obtida  $g(x) = Kx^{n+1} - (K + 1)x^n + 1 = 0$ . (iii) No plano A tem-se que  $n = 7$  e  $K = 1000/250 = 4$ . No Plano B,  $n = 10$  e  $K = 1000/200 = 5$ . A equação associada ao Plano A será denotada por  $g(x) = 4x^8 - 5x^7 + 1 = 0$  e a do Plano B será denotada por  $G(x) = 5x^{11} - 6x^{10} + 1$ .

Analisando os gráficos dessas funções (veja as **Figuras 13 e 14**) percebemos que as raízes das equações  $g(x) = 0$  e  $G(x) = 0$  pertencem ao intervalo  $[1.1 \ 1.2]$ .



**Figura 13: Gráfico de  $g(x) = 4x^8 - 5x^7 + 1$**



**Figura 14: Gráfico de  $G(x) = 5x^{11} - 6x^{10} + 1$**

Como  $g$  e  $G$  são funções contínuas e  $g(1.1) \cdot g(1.2) < 0$  e  $G(1.1) \cdot G(1.2) < 0$  então, pelo Teorema do Valor Intermediário, existem  $c$  e  $d$  pertencentes ao intervalo aberto  $(1.1, 1.2)$  tais que  $g(c) = 0$  e  $G(d) = 0$ . Observações:  $g(1.1) < 0$ ;  $g(1.2) > 0$ ;  $G(1.1) < 0$  e  $G(1.2) > 0$ .

Vamos utilizar o Método da Bissecção para fazer dois refinamentos do intervalo Inicial  $I_0 = [1.1, 1.2]$ .

(i) Primeiro refinamento.

Note que no ponto médio,  $x_0 = (1.1 + 1.2)/2 = 1.15$ , tem-se que  $g(x_0) < 0$  e  $G(x_0) < 0$ . Então o primeiro intervalo refinado é  $I_1 = [1.15, 1.2]$ .

(ii) Segundo refinamento.

Sabemos que  $g(1.15) < 0$  e  $g(1.2) > 0$ ;  $G(1.15) < 0$  e  $G(1.2) > 0$ . No ponto médio,  $x_1 = (1.15 + 1.2)/2 = 1.175$ , tem-se que  $g(x_1) > 0$  e  $G(x_1) > 0$ . Então o segundo intervalo refinado é  $I_2 = [1.15, 1.175]$ .

Vamos utilizar o Método de Newton – Raphson para encontrar as raízes das equações  $g(x) = 0$  e  $G(x) = 0$ . As iterações serão obtidas a partir da seguinte fórmula:  $x_m = x_{m-1} - [f(x_{m-1})/f'(x_{m-1})]$ , onde  $x_0 = 1.1625$  (ponto médio do intervalo  $I_2$ , dado anteriormente).

Primeiramente utilizaremos  $f(x) = g(x) = 4x^8 - 5x^7 + 1$  e  $f'(x) = g'(x) = 32x^7 - 35x^6$ . Depois utilizaremos  $f(x) = G(x) = 5x^{11} - 6x^{10} + 1$  e  $f'(x) = G'(x) = 55x^{10} - 60x^9$ .

(iii) Aproximação da raiz de  $g(x) = 0$ .

$$x_1 = 1.163272901; \quad x_2 = 1.163267091; \quad x_3 = 1.16326709.$$

Note que o erro relativo  $ER_2 = |x_2 - x_1| / |x_2| \approx 0.0000049946 < 0.5 \times 10^{-5}$ . Além disso,  $|g(x_2)| \approx 0.422 \times 10^{-8} < 0.5 \times 10^{-8}$ . A partir da quarta iteração, os valores de  $x_3$  serão repetidos, caso a calculadora exiba no visor apenas 9 dígitos.

Considerando  $x_2$  (ou  $x_3$ ) como a aproximação da raiz desejada então a precisão será de pelo menos oito casas decimais.

(iv) Aproximação da raiz de  $G(x) = 0$ .

$$x_1 = 1.15235692; \quad x_2 = 1.151006706; \quad x_3 = 1.150984151; \quad x_4 = 1.150984145.$$

Note que o erro relativo  $ER_4 = |x_4 - x_3| / |x_4| \approx 0.052 \times 10^{-7} < 0.5 \times 10^{-7}$ . Além disso,  $|G(x_4)| \approx 0.27 \times 10^{-8} < 0.5 \times 10^{-8}$ . A partir da quinta iteração, os valores de  $x_4$  serão repetidos, caso a calculadora exiba no visor apenas 10 dígitos. Considerando  $x_4$  como a aproximação da raiz desejada, então a precisão será de pelo menos oito casas decimais.

(v) Taxa de juros do Plano A.

$$j = x - 1; \text{ assim, } j \approx 1.16326709 - 1 = 0.16326709; \text{ ou seja, } j \approx 16,33 \, \% .$$

(vi) Taxa de juros do Plano B.

$$j = x - 1; \text{ assim, } j \approx 1.150984145 - 1 = 0.150984145; \text{ ou seja, } j \approx 15,10 \, \% .$$

Conclusão: O plano que apresenta a menor taxa de juros é o Plano B. □



A monografia de Leone Alves Leite foi publicada, em setembro de 2005, na revista eletrônica “FAMAT em Revista”. Esse artigo pode ser encontrado no seguinte endereço:

[www.portal.famat.ufu.br/sites/famat.ufu.br/files/Anexos/Bookpage/Famat\\_Revista\\_05.pdf](http://www.portal.famat.ufu.br/sites/famat.ufu.br/files/Anexos/Bookpage/Famat_Revista_05.pdf)

## 7. Atividades do Módulo 1

Não deixe de consultar outras referências bibliográficas, além desse material didático, para auxiliar você na resolução das atividades propostas a seguir. Veja, por exemplo:



[6] FRANCO, Neide Bertoldi, *Cálculo Numérico*, São Paulo, Pearson Prentice Hall, 2006.

### Atividade 1 – Lista de Exercícios

#### Enunciado da atividade

Em praticamente todos os exercícios desta lista você precisará fazer algum tipo de gráfico; portanto, é fundamental que você recorde a teoria e as técnicas, estudadas nas disciplinas de Cálculo Diferencial, referentes a esboço de gráfico de funções.

Essa lista contém exercícios referentes ao conteúdo do Módulo 1: Isolamento de raízes; Método da Bissecção; Método Iterativo Linear e Método de Newton – Raphson. Você tem que se esforçar para tentar fazer todos os exercícios.

#### Primeira Lista de Cálculo Numérico

1) Seja  $f(R) = -(10/R^2) + 4\pi R$ . Use algum tipo de gráfico para isolar o zero de  $f$ . Justifique a escolha do intervalo. Faça dois refinamentos do intervalo  $[0.5, 1]$ ; use o método da bissecção.

2) Um professor da área de exatas com talento para adivinhações lançou o seguinte desafio para os seus alunos: “eu sou capaz de adivinhar o número memorizado por qualquer um de vocês, em no máximo sete palpites, desde que o número pensado seja inteiro e esteja entre 1 e 100.” O professor, de fato, não era adivinhão; ele era conhecedor do método da bissecção!

i) No método da bissecção a sequência dos pontos médios dos intervalos  $[a_n, b_n]$ , que contém a raiz  $r$  da equação  $f(x) = 0$ , é dada por  $x_n = (a_n + b_n)/2$ . Logo,  $|x_n - r| \leq (b_n - a_n)/2 \leq (b - a)/2^{(n+1)} < \varepsilon$ . Considere  $\varepsilon = 1/2$  e obtenha  $n = 7$  (os palpites).

ii) Justifique a seguinte afirmação: “Se fosse considerado  $\epsilon = 1$  (ou seja  $|x_n - r| < 1$ ) não seria garantido o acerto em 7 palpites”. Observe que  $r$  é um número natural e a sua aproximação,  $x_n$ , é obtida por arredondamento simétrico.

3) Seja  $f(t) = t - \exp(t^2 - 1)$ , onde  $\exp(x) = e^x$ . Observe que  $f(1) = 0$ .

i) Faça os gráficos de  $y = t$  e  $y = \exp(t^2 - 1)$ ; obtenha um intervalo que contenha a outra raiz positiva de  $f(t) = 0$ .

ii) Se  $Q(t) = \exp(t^2 - 1)$ , então  $Q'(t) = 2t \exp(t^2 - 1)$ . Verifique que  $|Q'(t)| < 1$ , se  $0 < t < 1/2$  (Sugestão:  $0 < t < 1/2 \Rightarrow t^2 < 1/4 \Rightarrow t^2 - 1 < -3/4$ ). O M.I.L será convergente ?

4) Coloque V ou F e justifique.

i.( ) Seja  $f(x) = 4\sin(x) - e^x = 0$ , com  $x \in I = [-2\pi, 1/2]$ . Uma função de iteração pode ser dada por:  $Q(x) = x - 2\sin(x) + 0.5e^x$ .

ii.( ) Se  $x_0 = 1$ , então a sequência  $x_n = x_{n-1}(2 - a x_{n-1})$  convergirá para  $1/a$ , em que  $0 < a < 1$ , porque tal sequência será crescente e limitada superiormente por  $1/a$ .

iii.( ) Seja  $Q'(x) = 1 - 2\cos(x) + 0.5e^x$ . Como  $|Q'(-2\pi)| < 1$  e  $|Q'(0.5)| < 1$ , então  $|Q'(x)| < 1$ , para  $\forall x \in I = [-2\pi, 1/2]$ .

iv.( ) O método iterativo gerado pela função de iteração  $\phi(x) = 2x - ax^2$  convergirá para a raiz da equação  $f(x) = (1/x) - a = 0$ , onde  $1 > a > 0$ , se  $x_0 > 2/a$ .

5) Seja  $r$  a única raiz da equação  $f(x) = 0$ ,  $x \in I = [r - \delta, r + \delta]$  (intervalo centrado na raiz). Escrevendo a equação anterior na forma equivalente – problema de ponto fixo –:  $x = \phi(x)$ , uma das condições (a principal) de convergência do processo iterativo dado por  $x_n = \phi(x_{n-1})$ , com valor inicial  $x_0 \in I$ , é a seguinte:  $|\phi'(x)| \leq M < 1$ . Essa é uma condição suficiente. Para entender isso, considere as equações: (A)  $x = 2/x$ ,  $x > 0$ ; (B)  $x = \sin(x)$ ,  $x \in [-\pi/2, \pi/2]$ .

i) Verifique que  $|\phi'(r)| = 1$ , onde  $r = 2^{1/2}$ , em A, e  $r = 0$ , em B.

ii) Mostre que o processo iterativo, proveniente de A, vai gerar uma sequência divergente que ficará oscilando entre dois valores, qualquer que seja  $x_0 \neq r$ . Exiba esses valores.

iii) Mostre através de análise gráfica que o processo iterativo proveniente de B vai gerar uma sequência convergente. Considere  $x_0 > r$ .

iv) Considere  $x_0 = -0.05$  e faça algumas iterações. Observe que a sequência será crescente e a convergência será muito lenta. O método de Newton-Raphson, que tem convergência quadrática, apresentaria resultados muito melhores nesse caso? Justifique.  
Conclusão: Se  $\phi(x)$  não satisfaz as condições de convergência, o método iterativo pode convergir ou não; quando converge, a convergência pode ser muito lenta e, daí, o método será ineficiente. O melhor é trabalhar com funções que satisfazem as condições de convergência.

6) O dono de uma pequena obra exigiu que fosse gasto o mínimo possível de material na construção de uma caixa d'água cilíndrica, com capacidade para cinco mil litros ( $5 m^3$ ). Esse é um problema simples de otimização, onde se quer minimizar a área lateral de um cilindro, levando-se em conta as suas bases inferior e superior. Deixando-se a área lateral em função do raio da base, obtém-se a seguinte função:  $AL(R) = (10/R) + 2\pi R^2$ .

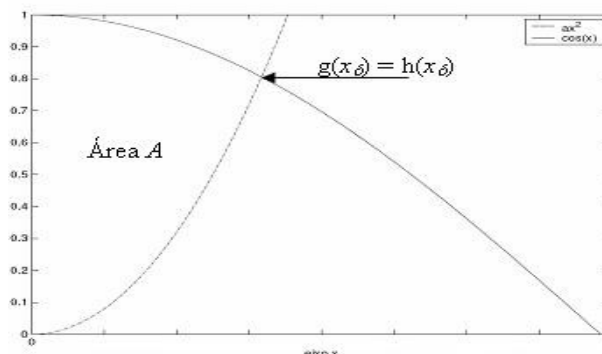
i) Seja  $f(R)$  a derivada primeira de  $AL(R)$ . Use um procedimento gráfico para isolar o zero de  $f$ , justifique. Faça dois refinamentos do intervalo  $[0.5 \ 1]$ ; use o método da bissecção.

ii) Utilize o valor inicial  $R_0 = 0.93$  e aplique o método de Newton-Raphson para obter o zero de  $f(R)$ , com erro relativo menor do que  $0.5 \times 10^{-6}$ .

iii) Considere as funções de iterações provenientes da equação  $f(R) = 0$ : (a)  $Q(R) = 10/(4\pi R^2)$  e (b)  $Q(R) = (10/4\pi)^{1/2} (R)^{-1/2}$ . Verifique as condições de convergência para ambas as funções. Qual delas vai gerar, com certeza, um processo iterativo convergente?

7) Sejam  $g(x) = ax^2$  e  $h(x) = \cos(x)$ . Se  $x_\delta$  (no primeiro quadrante) é ponto de interseção entre  $g$  e  $h$ , isto é  $g(x_\delta) = h(x_\delta)$ , obtenha o valor de  $a > 0$ , com precisão de pelo menos seis casas decimais, de modo que “ $A = 1/4$ ”, veja a figura abaixo. Note que a área é calculada pela seguinte integral:

$A = \int_0^{x_\delta} [h(x) - g(x)] dx$ . Use o método de Newton-Raphson.



8) O sistema linear  $Ax = b$  é equivalente ao sistema de equações  $x = Cx + g$ ;  $C$  é uma matriz de ordem  $n$ ;  $x$ ,  $b$  e  $g$  são vetores do  $R^n$ . Seja  $\lambda_i$  um autovalor de  $C$ ;  $|\lambda_i| < 1$ ,  $1 \leq i \leq n$ , se, e somente se, a sequência de vetores  $x^{(k+1)} = Cx^{(k)} + g$  converge para a solução do sistema linear, independentemente da escolha do vetor inicial  $x^{(0)}$ . Os autovalores de  $C$  são as raízes do polinômio (característico) de grau  $n$  dado por  $p(\lambda) = \det(C - \lambda I)$ ;  $I$  é a

matriz identidade. O polinômio característico de  $C = \begin{pmatrix} 0 & -1/5 & -1/5 \\ 3/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$  é  $p(\lambda) = -\lambda^3 - (\lambda/8) - 1/10$ .

**i)** Use algum tipo de gráfico, juntamente com argumentos teóricos, para exibir um intervalo contendo a única raiz real de  $p$

**ii)** Faça dois refinamentos de  $I = [-1/2 \ 0]$  (intervalo que contém a raiz de  $p$ ) com o método da bissecção. Use o ponto médio do último intervalo refinado como valor inicial para o método de Newton. Obtenha uma aproximação com erro absoluto menor do que  $0.5 \times 10^{-6}$ .

**iii)** Exiba uma função de iteração do M.I.L. associada à equação  $p(\lambda) = 0$ ; faça a análise de  $\phi(\lambda) = -(0.1 + 0.125\lambda)^{1/3}$ .

9) Calculadoras científicas possuem uma tecla que fornece rapidamente o valor do inverso de qualquer número real  $a \neq 0$ . Esse valor é calculado pelo método de Newton-Raphson, que usa a equação  $f(x) = (1/x) - a = 0$  (note que  $f(1/a) = 0$ ). Como é escolhido o valor inicial  $x_0$  do processo iterativo usado no programa implementado na calculadora? Pode-se mostrar que  $x_0 = 1$  é um valor inicial adequado para se obter a aproximação de  $1/a$ , se  $0 < a < 1$ .



i) Seja  $\phi(x) = 2x - ax^2$  a função de iteração do método de Newton-Raphson. Calcule  $\phi'(x)$  e determine uma condição para que  $|\phi'(x)| < 1$ .

Para o caso  $1 < a$ , a escolha de um valor inicial para o método de Newton pode ser feita considerando-se a seguinte condição:  $a x_0 < 2$  (essa condição é um pouco menos restritiva do que a condição obtida no item (i) anterior). O seguinte procedimento pode ser adotado para a escolha do valor inicial. Dado  $a > 1$  construa a sequência  $x_{0,0} = 0.005$  (ou qualquer outro número próximo de zero) e  $x_{0,i} = x_{0,i-1} \div 10$ . Escolha  $x_0 = x_{0,i}$  assim que  $ax_{0,i} < 2$ .

ii) Faça 4 iterações com o método de Newton para obter o inverso de  $a = 500$ . Use o processo anterior para a escolha de  $x_0$ . Calcule o erro absoluto.

10) Dados o ponto  $(x_{n-1}, f(x_{n-1}))$  e o coeficiente angular  $m = f'(x_{n-1})$ , o próximo valor da sequência gerada pelo método de Newton-Raphson é  $x_n = x_{n-1} - [f(x_{n-1})/f'(x_{n-1})]$ , obtido da interseção da reta tangente  $y = f(x_{n-1}) + m(x - x_{n-1})$  com o eixo  $x$ . Nos casos em que  $f'(x)$  é uma função complicada (exigiria muito esforço computacional nos cálculos de  $f'(x_{n-1})$ ), usa-se o método de Newton Modificado:  $x_n = x_{n-1} - [f(x_{n-1})/f'(x_0)]$ . Note que, nesse método, as retas tangentes são substituídas por retas com um mesmo coeficiente angular  $m = f'(x_0)$ .

i) Faça a interpretação geométrica desse método.

ii) Seja  $f(x) = x^5 - (10/9)x^3 + (5/21)x = 0$ ; use  $x_0 = -0.8984375$  e faça duas iterações com o método de Newton modificado. Qual a precisão da aproximação?

11.1) Obtenha um intervalo que contenha a interseção entre a parábola  $h(p) = p^2 - p + 1$  e a função  $g(p) = p(1 + p)^{1/21}$ , considerando  $0 \leq p \leq 1$ . Para isso, faça o gráfico de  $h(p)$ , exibindo o ponto de mínimo da parábola; faça o gráfico de  $g(p)$ , mostrando que  $g'(p) > 0$  e  $g''(p) > 0$ . Use o Teorema do Valor Intermediário, com a função  $f(p) = h(p) - g(p)$ , para garantir que o intervalo que você encontrou possui a raiz de  $f(p) = 0$ .

11.2) Considere a função  $f(p)$ , anterior (11.1), definida no intervalo  $I_0 = [0.8125, 0.84375]$ . Faça dois refinamentos com o método da Bissecção. Quantos refinamentos serão necessários para se obter uma aproximação da raiz de  $f(p) = 0$  com precisão de 3 casas decimais ( $\varepsilon = 0.5 \times 10^{-3}$ ) ?

11.3) No artigo de J. B. Keller, "Probability of a shutout in racquetball", publicado no SIAM Review 26, n. 2 (1984), 267-268, QA1.S2, a seguinte fórmula é apresentada:

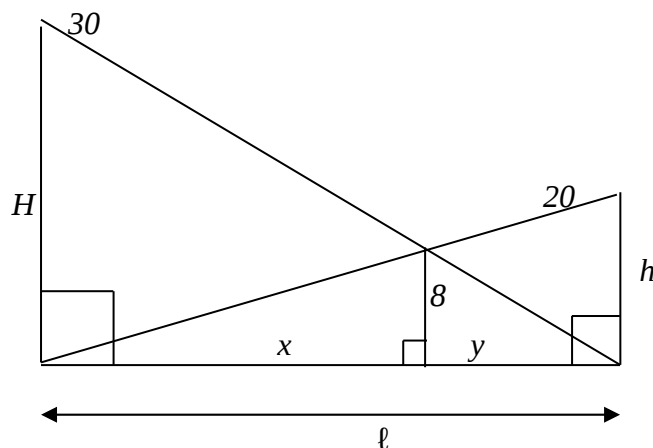
$$P = \frac{1+p}{2} \left( \frac{p}{p^2 - p + 1} \right)^{21}.$$

Nessa fórmula,  $P$  é a probabilidade de certo jogador A eliminar um jogador B, vencendo-o por um placar de 21 a 0, em uma partida de raquetebol;  $p$  é a probabilidade de o jogador A rebater um saque com precisão, qualquer que seja o sacador. Qual o valor de  $p$  dado que  $P = 1/2$  ?

Sugestão: Use o método de Newton-Raphson. Tome como valor inicial para o método de N-R o ponto médio do intervalo  $I_2$  (veja o item 11.2). Faça duas iterações e analise o erro relativo.

11.4) Obtenha uma função de iteração,  $\phi(p)$ , para o método iterativo linear (método do ponto fixo) associado ao item anterior (exercício 11.3). Faça algumas iterações com essa função e observe se o método irá convergir. Tente mostrar que  $|\phi'(p)| < 1$ .

12) Considere a figura dada a seguir. Para obter o valor de  $\ell$  siga as sugestões abaixo.



Sugestões:

i) Utilize um teorema conhecido para encontrar as seguintes relações:

$$H^2 + \ell^2 = 900 \quad \text{e} \quad \text{ii) } h^2 + \ell^2 = 400;$$

ii) Use semelhança de triângulos para obter a próxima relação:

$$(\ell/h) + (\ell/H) = \ell/8.$$

iii) Utilize os itens anteriores e obtenha a seguinte expressão

$$-H^2 + (64.H^2)/(H-8)^2 = -500 \Leftrightarrow H(H-16) = 500.(H-8)^2/H^2.$$

iv) Faça os gráficos de  $f(H) = H(H-16)$  e de  $g(H) = 500.(H-8)^2/H^2$ , com  $H > 0$ .

v) Verifique que a equação polinomial  $p(H) = H^4 - 16 H^3 - 500 H^2 + 8000 H - 32000 = 0$  pode ser escrita como  $f(H) = g(H)$ .

vi) Use o Teorema do Valor Intermediário para garantir que a raiz positiva,  $r$ , de  $p$  está no intervalo  $[24, 26]$ .

vii) Faça dois refinamentos do intervalo anterior por bissecção e considere o ponto médio do último intervalo refinado como valor inicial para o método de Newton-Raphson. Obtenha uma aproximação de  $r$  com pelo menos 5 casas decimais.

13) Seja  $a \in \mathbb{R}$ ,  $a > 0$ . Aplicando-se o método de Newton-Raphson na equação  $x^p = a$ ,  $p \in \mathbb{N}$ ,  $p \geq 2$ , a seguinte função de iteração é obtida:  $Q(x) = [(p-1)/p]x + (a/p)x^{(1-p)}$ . Considerando  $x \in \mathbb{R}$ ,  $x > 0$ , a condição  $|Q'(x)| < 1$  será satisfeita para  $x > [a]^{1/p} [(p-1)/(2p-1)]^{1/p}$  (\*).

i) Justifique por que a condição anterior (\*) é muito restritiva para a escolha de um valor inicial para o cálculo da raiz  $p$ -ésima de  $a$  ( $[a]^{1/p}$ ), através do método de Newton-Raphson.

Sugestão: Verifique que a condição anterior vai gerar um intervalo contendo a raiz (centrado na raiz). Atribua valores para  $p$  (variando de 2 até o infinito) e observe o que ocorre com tal intervalo.

ii) A condição anterior é apenas suficiente para a convergência do método de N-R. De fato, a sequência gerada pela função de iteração,  $Q(x)$ , convergirá para  $r = [a]^{1/p}$ , independentemente do valor inicial escolhido. Para entender o comportamento dessa sequência faça, em primeiro lugar, o gráfico de  $f(x) = x^p - a$ ,  $a > 0$ , considerando  $p$  par ou  $p$  ímpar. A seguir, lembrando-se de que o método de N-R é baseado em retas tangentes, faça a interpretação geométrica da convergência, indicando se a sequência obtida será crescente ou decrescente.

Nota: Matematicamente, o resultado que está por trás da convergência anterior é o seguinte: “Se existir um intervalo  $I = [c, r)$  ou  $I = (r, d]$ , tal que  $f(x).f''(x) > 0$ ,  $\forall x \in I$ , e  $f(r) = 0$ , então o método de Newton-Raphson convergirá para  $r$ , dado qualquer valor inicial em  $I$ ”.

iii) Dado um valor qualquer  $x_0$ , com  $0 < x_0 < r$ , mostre que  $x_1 = Q(x_0) > r$ . Mostre, também, que a condição expressa no resultado da nota anterior é válida no intervalo  $I = (r, x_1]$ .

Sugestões: 1) Lembre-se de que  $Q(r) = r$ ; 2) Se  $Q'(x) < 0$ , então  $Q(x)$  é decrescente; 3)  $Q'(x) = [f(x)f''(x)]/(f'(x))^2$ ; 4) Verifique o sinal de  $f(x)$  para  $0 < x < r$  e para  $x > r$ .

Observação: O teorema enunciado na Nota anterior é demonstrado seguindo-se os seguintes passos:

1º) Mostra-se que  $Q(x) = x - (f(x)/f'(x))$  é crescente em  $I$ .

2º) Mostra-se que: i)  $f(x) < 0 \Rightarrow f''(x) < 0 \Rightarrow f'(x)$  é decrescente; ii)  $f(x) > 0 \Rightarrow f''(x) > 0 \Rightarrow f'(x)$  é crescente. O fato de  $f'$  ser crescente ou decrescente determinará se  $f'$  será positiva ou negativa no intervalo  $I$ ; tal análise dependerá do sinal de  $f'(r) = \lim_{x \rightarrow r} [f(x) - f(r)]/[x - r]$ .

3º) Mostra-se que: i)  $f(x)/f'(x) < 0, \forall x \in I = [c, r)$ ; ii)  $f(x)/f'(x) > 0, \forall x \in I = (r, d]$ . Conclui-se, portanto, que  $x_n = x_{n-1} - [f(x_{n-1})/f'(x_{n-1})]$  ou é uma sequência crescente e limitada (caso i), ou é uma sequência decrescente e limitada (caso ii).

4º) Dos passos anteriores, segue que  $x_{n-1} \in I \Rightarrow x_n \in I$ . Logo, a sequência convergirá para  $r$ .

## Atividade 2 – Esboço de gráfico de função e revisão de Cálculo Diferencial

### Prezado(a) aluno(a),

Será fundamental, para o bom acompanhamento deste tópico, que você esteja familiarizado(a) com as técnicas para fazer esboço de gráfico de funções. Essas técnicas foram estudadas na disciplina de Cálculo Diferencial. Para recordar essas técnicas, recomendo que você leia novamente o material didático que foi disponibilizado em sua disciplina de Cálculo.

Para fazer o esboço do gráfico de uma função real, você deverá recordar os conceitos de função monótona (crescente ou decrescente) e concavidade de gráfico (concavidade para cima e concavidade para baixo) e relacionar esses conceitos com as definições de derivada primeira e derivada segunda de uma função  $f$ . Além disso, você tem que recordar os conceitos de valores máximos e mínimos de função e relacioná-los com os testes da derivada primeira ou da derivada segunda; recordar a definição de assíntota vertical e horizontal e de ponto de inflexão.

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre o método da Bissecção e Isolamento de raízes: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 1), ambos localizados em [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Como exercício, proponho que você faça os esboços dos seguintes gráficos:

1)  $f(H) = H(H-16)$ , com  $H > 0$ ;

2)  $g(H) = 500.(H-8)^2/H^2$ , com  $H > 0$ .

Faça o esboço dos dois gráficos em um mesmo plano cartesiano. Obtenha um intervalo fechado  $I = [a, b]$  que contenha a intersecção de  $f$  e  $g$ . Nesse caso, o ponto de intersecção,  $h$ , é tal que  $f(h) = g(h)$ , ou seja,  $F(h) = f(h) - g(h) = 0$ . Portanto, o zero da função  $F$  coincide com o ponto de intersecção das funções  $f$  e  $g$ .

### Informações sobre a Atividade 2

1)  $f(H) = H(H - 16)$ , com  $H > 0$ .

2)  $g(H) = 500 (H - 8)^2/H^2$ , com  $H > 0$ .

**Observações:** i)  $g'(H) = 8.000 (H - 8)/H^3$ ;      ii)  $g''(H) = 16.000 (12 - H)/H^4$ .

3) Pontos críticos de cada uma das funções - Ponto crítico de  $f$ :  $H = 8$ ; ponto crítico de  $g$ :  $H = 8$ .

4) Obtenção do ponto de inflexão de  $g$ :  $H = 12$ .

5) Estudo do sinal:  $g' > 0$  se  $H > 8$  e  $g' < 0$  se  $H < 8$ ;  $g'' > 0$  se  $H > 12$  e  $g'' < 0$  se  $H < 12$ .

6) Um intervalo  $I$  que contém a raiz de  $F(H) = 0$  é, por exemplo,  $I = [25, 26]$ .

7) O Teorema do Valor Intermediário garante que o intervalo está correto, pois  $F(a).F(b) < 0$ .

Gráficos das funções da atividade 2.

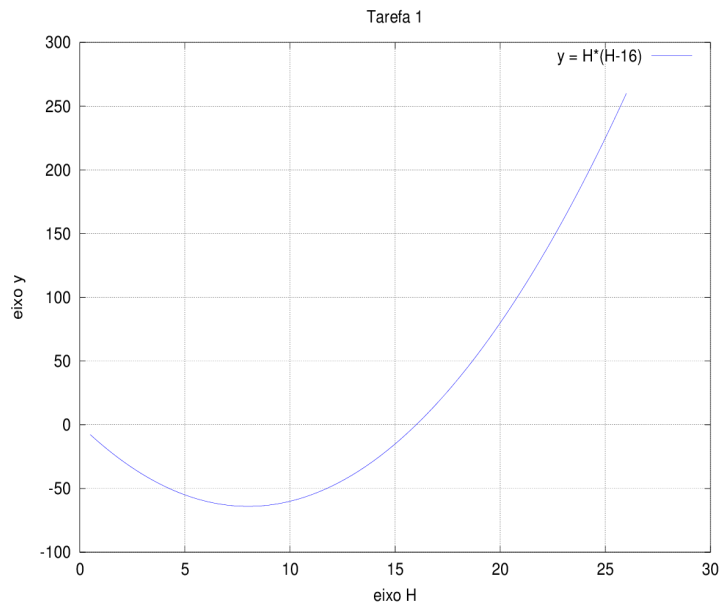


Gráfico de  $f(H) = H(H - 16)$ , com  $H > 0$

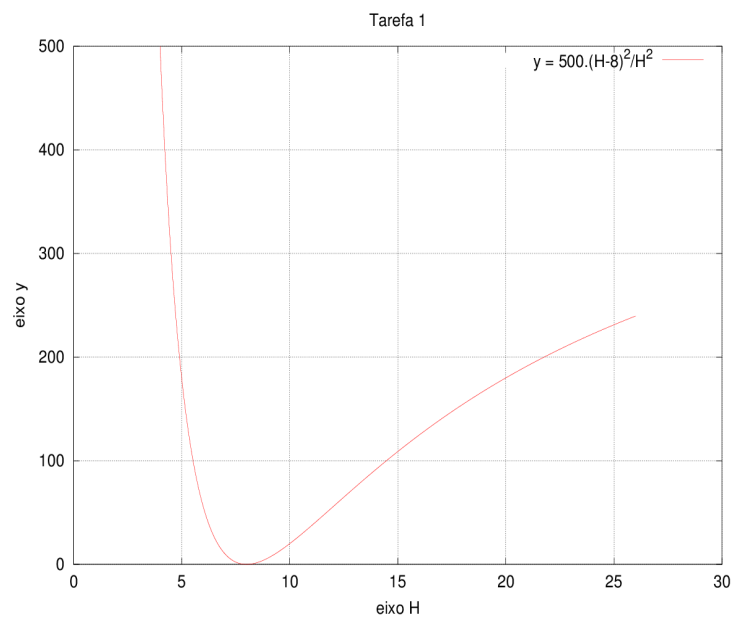


Gráfico de  $g(H) = 500 (H - 8)^2 / H^2$ ; com  $H > 0$

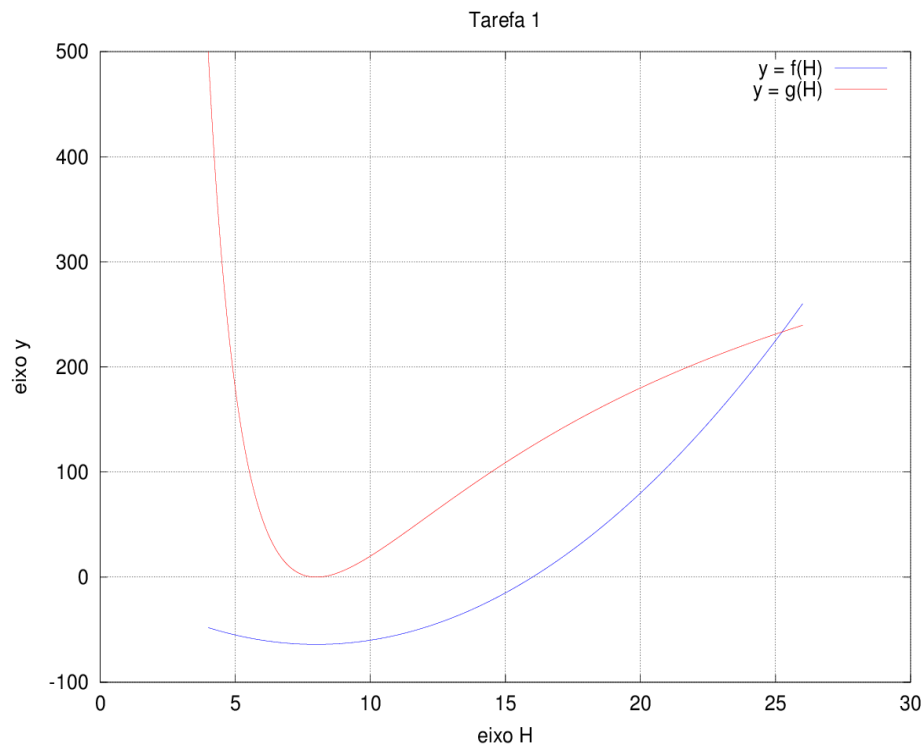


Gráfico de  $g(H) = 500 (H - 8)^2/H^2$  e  $f(H) = H(H - 16)$ ; com  $H > 0$

### Atividade 3 – Isolamento de raiz e o método da Bissecção

#### Prezado(a) aluno(a),

Você já deve ter visto na disciplina de Cálculo Diferencial o enunciado do Teorema do Valor Intermediário. A demonstração desse teorema é feita em um curso de Análise na reta. Porém, uma demonstração construtiva desse importante resultado de Análise pode ser elaborada com os conhecimentos que você adquiriu nas disciplinas de Cálculo Diferencial, tais como: funções contínuas e sequências de números reais, juntamente com os teoremas lá demonstrados. Essa construção construtiva é conhecida como o Método da Bissecção.

#### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre o método da bissecção: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 1), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Se  $F(H) = f(H) - g(H)$ , onde  $g(H) = 500.(H - 8)^2/H^2$  e  $f(H) = H(H - 16)$ , então  $F(H) = 0 \Leftrightarrow p(H) = H^4 - 16 H^3 - 500 H^2 + 8000 H - 32000 = 0$ .

Obtenha uma aproximação com duas casas decimais para o zero de  $F$ . Considere a sequência formada pelos pontos médios de cada intervalo refinado, partindo do intervalo inicial  $I_0 = [a_0, b_0] = [25.21875, 25.25]$ . Note que as aproximações, nesse caso, devem ter pelo menos três casas decimais. Mostre que  $F(a_0) \cdot F(b_0) < 0$ . Como teste de parada, utilize o seguinte critério:  $|F(x_n)| < 0.5 \times 10^{-2}$ , onde  $x_n$  é o ponto médio do  $n$ -ésimo intervalo refinado. Quantos refinamentos do intervalo  $I_0$  são necessários para se obter uma aproximação com erro menor do que  $10^{-9}$ ?

### Informações sobre a Atividade 3.

**Observação:**  $I_0 = [a, b] = [25.21875, 25.25]$ . Na primeira parte da questão a tolerância é  $\varepsilon = 0.5 \times 10^{-2}$ . Utilizando o Método da Bissecção, na equação:  $p(H) = H^4 - 16H^3 - 500H^2 + 8000H - 32000 = 0$ , obtêm-se:

- 1)  $I_1 = [25.234375, 25.25]$ .
- 2)  $I_2 = [25.234375, 25.2421875]$ .
- 3)  $I_3 = [25.23828125, 25.2421875]$ .
- 4) Sequência dos pontos médios:  $x_1 = 25.234375$ ;  $x_2 = 25.23828125$ ;  $x_3 = 25.24023438$ .
- 5) Os valores de  $p(x_n)$  não estão próximos de zero, mesmo que  $x_n$  esteja próximo da raiz.

**Observação:** na segunda parte da questão a tolerância é  $\varepsilon = 0.5 \times 10^{-9}$ .

- 6) Utilizando a fórmula  $n > -1 + [\ln(b - a) - \ln(\varepsilon)]/\ln(2)$ , obtém-se  $n \geq 25$ .

### **Atividade 4 – O Método do Ponto Fixo**

#### **Prezado(a) aluno(a),**

Dada uma equação do tipo  $f(x) = 0$  sempre é possível reescrevê-la da forma equivalente  $x = \varphi(x)$ , com  $\varphi(x) = x - f(x)$ , por exemplo. Nesse caso, o problema de se encontrar os zeros de uma função  $f$  (encontrar  $x$  tal que  $f(x) = 0$ ) é equivalente ao problema de se encontrar os pontos fixos da função  $\varphi$  (encontrar  $x$  tal que  $\varphi(x) = x$ ).

O Método do Ponto Fixo consiste na construção do seguinte processo iterativo: a partir de um dado valor inicial  $x_0$  obtém-se os valores de uma sequência com termo geral  $x_n = \varphi(x_{n-1})$ . Sob certas condições a respeito da função  $\varphi$ , pode-se mostrar que a sequência



anterior converge para  $r$ , a raiz da equação  $f(x) = 0$  ou, equivalentemente, para o ponto fixo de  $\varphi$ . As condições suficientes de convergência são as seguintes:  $\varphi$  e  $\varphi'$  são funções contínuas e  $|\varphi'(x)| \leq M < 1$ , para todo  $x$  pertencente a uma vizinhança de  $r$ .

O Método do Ponto Fixo também é conhecido como Método Iterativo Linear, devido à sua convergência linear, ou seja,  $\lim_{n \rightarrow \infty} \frac{|e_n|}{|e_{n-1}|} = C$ , onde  $|e_n| = |x_n - r|$  e  $C$  é uma constante tal que  $0 < C < 1$ .

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre o método iterativo linear: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 1), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Seja  $f(x) = x - 1,5 \sin(x)$ .

(a) Faça os gráficos das funções  $h(x) = x$  e  $g(x) = 1,5 \sin(x)$ . Obtenha um intervalo que contenha a raiz positiva de  $f(x) = 0$ . Use o Teorema do Valor Intermediário para garantir que realmente existe uma raiz no intervalo dado.

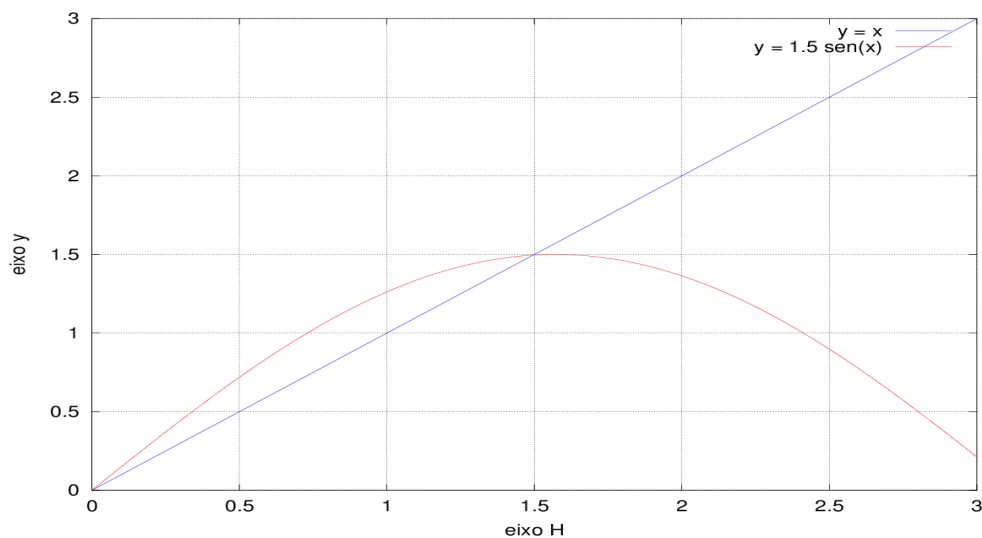
(b) Obtenha uma aproximação da raiz positiva de  $f(x) = 0$  com precisão de 9 casas decimais. Use o Método do Ponto Fixo e como teste de parada use o erro relativo:  $ER_k = \frac{|x_k - x_{k-1}|}{|x_k|} < 0,5 \times 10^{-9}$ .

(c) Ajuste o intervalo encontrado, se for necessário, de modo a garantir a validade das condições suficientes de convergência do Método do Ponto Fixo:  $\varphi$  e  $\varphi'$  são funções contínuas e  $|\varphi'(x)| \leq M < 1$ , para todo  $x$  pertencente a uma vizinhança de  $r$ .

### Informações sobre a atividade 4.

**Observação:**  $f(x) = x - 1,5 \sin(x) = 0$ ; a tolerância é  $\varepsilon = 0,5 \times 10^{-9}$ .

(a.1) Construção dos gráficos.



(a.1) Intervalo que contém a raiz positiva de  $f$ :  $[1, 1.5]$ .

(a.2) Aplicação do Teorema do Valor Intermediário:  $f(1)f(1.5) < 0$ .

(b.1) A sequência gerada pelo MIL tem termo geral  $x_n = 1.5 \sin(x_{n-1})$ ; com  $x_0 = 1.25$ , por exemplo.

(b.2) Use o erro relativo como critério de parada.

**Observação:**  $ER_k = \frac{|x_k - x_{k-1}|}{|x_k|} < 0.5 \times 10^{-9}$ .

(c.1) Note que  $\phi(x) = 1.5 \sin(x)$  e  $\phi'(x) = 1.5 \cos(x)$  são contínuas.

(c.2) Não se esqueça de mostrar que  $|\phi'(x)| < 1$ , no intervalo dado. Note que  $\cos(x)$  é decrescente e positiva no intervalo  $[1, 1.5]$ .

### Atividade 5 – O Método de Newton - Raphson

**Prezado(a) aluno(a),**

Dada uma equação do tipo  $f(x) = 0$ , o Método de Newton – Raphson (MNR) tem função de iteração  $\phi(x) = x - \frac{f(x)}{f'(x)}$ . É fácil mostrar que  $\phi'(x) = \frac{f(x)f''(x)}{[f'(x)]^2}$ . Se  $r$

representa o zero de  $f$ , então  $\phi'(r) = 0$ . Supondo que as funções  $f$ ,  $\phi$  e  $\phi'$  são contínuas, pode-se mostrar que  $|\phi'(x)| \leq M < 1$ , para todo  $x$  pertencente a uma vizinhança de  $r$ ; além disso,  $\phi$  e  $\phi'$  serão contínuas, garantindo a validade das condições suficientes de convergência para a função de iteração do MNR. Essa convergência é quadrática, ou seja,  $\lim_{n \rightarrow \infty} \frac{|e_n|}{|e_{n-1}|^2} = C$ , onde  $|e_n| = |x_n - r|$  e  $C = \frac{1}{2} \frac{|f''(r)|}{|f'(r)|}$ . Grosso modo, podemos pensar que o valor absoluto do erro cometido na  $n$ -ésima iteração é aproximadamente o quadrado do erro da iteração anterior multiplicado pela constante  $C$ . Por essa razão, o MNR é mais rápido que o Método Iterativo Linear.

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre o método de Newton - Raphson: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 1), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Considere as seguintes funções, associadas às atividades 3 e 4 dadas anteriormente,  $p(H) = H^4 - 16H^3 - 500H^2 + 8000H - 32000$  e  $f(x) = x - 1,5 \sin(x)$ . Obtenha o zero de cada uma das funções utilizando o método de Newton-Raphson. Considere o erro relativo  $ER_k = \frac{|x_k - x_{k-1}|}{|x_k|} < 0,5 \times 10^{-9}$ .

### Informações sobre a atividade 5.

**Observação:**  $f(x) = x - 1,5 \sin(x) = 0$ ,  $p(H) = H^4 - 16H^3 - 500H^2 + 8000H - 32000$ ; a tolerância do erro relativo é  $\varepsilon = 0,5 \times 10^{-7}$ .

(1) O zero de  $f$  é dado por 1.495781568 (com precisão de pelo menos 8 casas decimais).

(2) O zero de  $p$  é dado por 25.24216659 (com precisão de 7 casas decimais).

## 8. Atividades suplementares



### Atividade 1

Os códigos exibidos a seguir foram desenvolvidos na linguagem de programação do OCTAVE. A sua tarefa será testar todos os códigos utilizando o mesmo *software* (OCTAVE) ou desenvolver as suas próprias rotinas utilizando a linguagem de programação que lhe seja mais conveniente.

(i) Rotina Computacional para fazer gráfico de função - Referência: gráfico do **Problema 2** (Seção 5).

(Início da Rotina.)

%analise grafica de  $f(x) = 0$ , onde  $f(x) = x - 1.5 \sin(x)$

t=0:0.01:3; %vetor iniciando em 0 e terminando em 3, com espaçamentos iguais a 0.01.

y=t-1.5.\*sin(t); %vetor que armazena os valores da função em cada ponto do vetor t anterior.

y1= 1.5.\*sin(t); %vetor que armazena o valor da função  $1.5\sin(t)$  em cada ponto do vetor t.

plot(t,y,'-r', t,t,'-b', t,y1,'-g');

%exibe o gráfico de 3 funções:  $y(t)$ ,  $y = t$  e  $y1(t)$ ; -r: linha contínua na cor %vermelha; -b: linha contínua na cor azul e -g: linha contínua na cor verde.

legend('y=t-1.5sen(t)', 'y =t', 'y = 1.5sen(t)', 0) %exibe a legenda de cada um dos 3 gráficos

grid; %produz um quadriculado.

xlabel('eixo t') %exibe um nome para o eixo horizontal

ylabel('eixo y') %exibe um nome para o eixo vertical.

(Fim da Rotina.)

(ii) Rotina Computacional: Método da Bissecção acoplado ao Método de Newton-Raphson - Referência: **Problema 3** – Plano A (Seção 6).

(Início da Rotina.)

```
%Método de Newton_Raphson  
%Cálculo da raiz da equação  $g(x) = 0$ 
```

```
clear
```

```
a = 1.1; %extremo inferior do intervalo onde f muda de sinal  
b = 1.2; %extremo superior do intervalo onde f muda de sinal
```

```
pm = (a+b)/2; %ponto médio do intervalo [a,b]  
comp = (b-a)/2; %metade do comprimento do intervalo [a,b]
```

```
cont = 0; %variável que conta o número de iterações
```

```
eps = 0.5*10-2; %tolerância usada no teste de parada
```

```
fa= funcao(a); %cálculo de f(a)  
fb = funcao(b);  
fpm = funcao(pm); %cálculo de f(pm)
```

```
VAf = abs(fpm); % valor absoluto de fpm
```

```
while((comp > eps) && (VAf > eps))  
    if(fa*fpm<0)  
        b=pm;  
    else  
        a=pm;  
        fa=fpm;  
    end  
  
    pm=(a+b)/2;  
    fpm=funcao(pm);  
    VAf = abs(fpm);  
    cont = cont+1;  
    comp=(b-a)/2;  
end
```

```
fprintf('\n');
```

```
fprintf('A raiz aproximada por bissecao eh dada por pm =%12.10f\n',pm);  
fprintf('O numero de refinamentos do intervalo inicial foi cont = %d\n',cont);  
fprintf('A tolerancia usada foi eps = %12.10f\n',eps);  
fprintf('O valor de f(pm) eh dado por fpm =%12.12f\n',fpm);
```

```
x0 = pm;
```

```

f0 = fpm;

df0 = dfuncao(x0); %derivada da função no ponto x0

ER = 1; %armazena o valor do Erro Relativo

%Método de Newton_Raphson

s = 8;
cont = 0;
tol = 0.5*10^(-s); %tolerância usada no Erro Relativo

while((ER > tol) && (VAf > tol))
    cont = cont+1;

    x = x0 - (f0/df0);
    ER = abs((x-x0)/(x));
    x0 = x;
    f0 = funcao(x0);
    df0 = dfuncao(x0);
    VAf = abs(f0);
end

fprintf('O n. de iteracoes do N-R eh cont = %i \n',cont);
fprintf('O erro relativo da aproximacao da raiz eh ER = %12.12f\n',ER);
fprintf('O valor da raiz procurada eh x = %12.12f\n',x);
fprintf('f(x) = %12.12f\n',f0);
(Fim da Rotina.)

```

**Observações:** As funções utilizadas na rotina anterior foram definidas em dois arquivos com os conteúdos descritos a seguir.

(1) arquivo funcao.m:

```

function g = funcao(x)
    g = 4*x^8 -5*x^7 +1;

```

(2) arquivo dfuncao.m:

```

function g = dfuncao(x)
    g = 32*x -35;
    g = g*x^6;

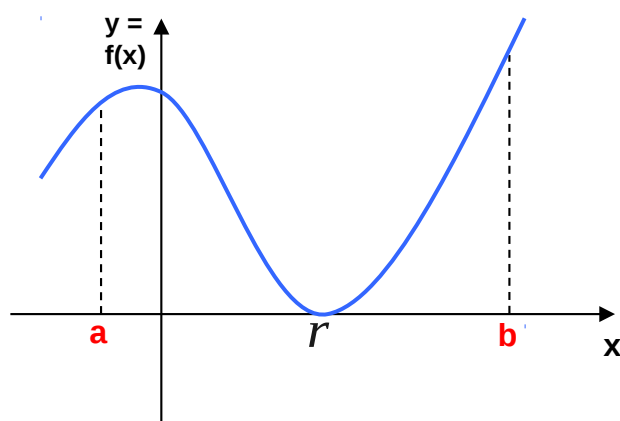
```

**Observação:** Em Octave, o nome de um arquivo que representa uma função que é chamada em uma determinada rotina (ou código computacional) deve coincidir com o nome que foi utilizado para a função, na rotina. Por exemplo, a função  $\text{funcao}(x)$  foi chamada na rotina anterior (Método de Newton-Raphson) e o nome do arquivo correspondente teve que ser  $\text{funcao.m}$ . O mesmo procedimento foi adotado para a função  $\text{dfuncao}(x)$ . O arquivo correspondente recebeu o nome de  $\text{dfuncao.m}$

**Desafio 1:** Desenvolva rotinas computacionais para calcular as iterações do Método do Ponto Fixo relacionado ao Problema 2 (Seção 5).

### Atividade 2

A figura abaixo mostra um caso para o qual o Método da Bissecção não pode ser utilizado (note que  $f(a)f(b) > 0$ ). Já o Método de Newton-Raphson será convergente; porém a convergência da sequência gerada por esse método não é garantida pelo **Teorema 4**, que foi demonstrado na Seção 6. Isso ocorre porque  $f'(r) = 0$ .



**Figura 15: Caso especial de convergência do Método de Newton – Raphson**

Considere  $f(x) = x^2$ . Mostre que a sequência gerada pelo Método de Newton-Raphson é convergente para  $r = 0$ . Considere  $x_0 > 0$  e demonstre que a sequência será decrescente e limitada inferiormente por zero ( $x_n > 0, \forall n \in \mathbb{N}$ ).

## Módulo 2

# Sistemas Lineares

### 1. Introdução

**Prezado estudante, seja bem vindo.**

Neste módulo você aprenderá a resolver sistemas de equações lineares representados matricialmente por  $Ax = b$ , onde  $A$  é uma matriz de ordem  $n$  com entradas reais  $a_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq n$ ;  $b$  é o vetor de termos independentes com coordenadas reais  $b_i$ ,  $1 \leq i \leq n$ ; e  $x$  é o vetor das incógnitas com coordenadas reais  $x_i$ ,  $1 \leq i \leq n$ .

Primeiramente, você estudará os métodos iterativos. Nesse caso, o sistema de equações lineares  $Ax = b$  será reescrito como um sistema de equações equivalentes (ambos os sistemas possuem o mesmo vetor solução) dado por  $x = Cx + g$ , onde  $C$  é uma matriz de ordem  $n$  com entradas reais  $c_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq n$  e  $g$  é o vetor de termos independentes com coordenadas reais  $g_i$ ,  $1 \leq i \leq n$ . Sob certas condições, por exemplo  $\|C\| < 1$  (a norma da matriz  $C$  é menor do que 1), mostraremos que o método iterativo (sequência de vetores) dado por  $x^{(k)} = Cx^{(k-1)} + g$ , com  $x^{(0)}$  vetor inicial dado, convergirá para a solução do sistema original  $Ax = b$ .

**Observação:** A norma de matriz de que faremos uso com mais frequência nesse curso é chamada de norma linha. Considerando uma matriz  $C$  de ordem  $n$ , a norma linha de  $C$  é dada por:

$\|C\| = \text{máximo}\{L_1, L_2, \dots, L_n\}$ , onde  $L_i = \sum_{j=1}^n |c_{ij}|$  (que corresponde à soma dos valores absolutos dos elementos da linha  $i$  da matriz  $C$ ),  $1 \leq i \leq n$ .

Depois dos métodos iterativos, você estudará os métodos baseados em escalonamento de matriz: Método de Eliminação de Gauss, Método de Eliminação de Gauss com Pivoteamento Parcial e Decomposição LU. Os métodos de escalonamento consistem em realizar operações elementares nas linhas da matriz  $A$  do sistema linear de modo a transformá-la em uma matriz triangular superior  $U$ .  $U$  é dita triangular superior quando  $u_{ij} = 0$ , nos casos em que  $i > j$ . O sistema linear  $Ax = b$  possui a mesma solução do sistema linear escalonado  $Ux = B$ , onde o vetor  $B$  foi obtido do vetor  $b$  por meio das mesmas operações elementares (aplicadas nas linhas da matriz do sistema) que transformaram  $A$  em  $U$ .



Para recordar, as operações elementares são 3:

(e1)  $L_i \leftrightarrow L_j$ : a linha  $i$  é trocada com a linha  $j$ ;

(e2)  $L_i \rightarrow L_i + t.L_j$ ,  $t \neq 0$ : a linha  $i$  é trocada pela soma da linha  $i$  com a linha  $j$  multiplicada por um número real  $t$  não nulo (se  $t$  for nulo então a matriz não sofrerá nenhuma alteração).

**Observação.** A operação anterior (e2) pode ser invertida, ou seja, para recuperar a linha  $i$  que foi alterada basta executar a seguinte operação:  $L_i \rightarrow L_i - t.L_j$ . Além disso, este tipo de operação elementar não altera o valor do determinante da matriz.

(e3)  $L_i \rightarrow t.L_i$ ,  $t \neq 0$ : a linha  $i$  é trocada pela linha  $i$  multiplicada por um número real  $t$  não nulo.

**Observação.** A operação anterior (e3) também pode ser invertida, ou seja, para recuperar a linha  $i$  que foi alterada basta executar a seguinte operação:  $L_i \rightarrow (t^{-1}).L_i$ .

**Observação.** Do processo de escalonamento, utilizando as operações elementares anteriores e o método da Eliminação de Gauss, é possível determinar duas matrizes, uma triangular superior  $U$  e uma triangular inferior  $L$  ( $l_{ij} = 0$ , se  $i < j$ ), tais que: ou  $A = LU$ , caso não haja nenhuma troca de linhas no processo de escalonamento, ou  $PA = LU$ , caso haja trocas de linhas e, nesse caso,  $PA$  denotará a matriz  $A$  permutada. Esse tipo de fatoração de  $A$  ou de  $PA$  é conhecida como decomposição  $LU$ . O sistema  $Ax = b$  é resolvido em duas etapas:  $Ly = b$  e  $Ux = y$ , caso não sejam efetuadas trocas de linhas; ou, se não,  $(PA)x = Pb$  é equivalente a  $(LU)x = Pb$  que é resolvido em duas etapas:  $Ly = Pb$  e  $Ux = y$ , onde  $Pb$  é o vetor  $b$  permutado de acordo com as trocas de linhas que foram efetuadas no Pivoteamento Parcial.

**Observação.** Não deixe de rever os tópicos sobre matriz inversa (Leia, por exemplo, o material visto na disciplina de Álgebra Linear). Lembre-se de que uma matriz quadrada  $A$  de ordem  $n$  tem inversa se existir uma matriz quadrada de ordem  $n$  indicada por  $A^{(-1)}$  tal que  $A \cdot A^{(-1)} = A^{(-1)} \cdot A = I$ , onde  $I$  é a matriz identidade, que é formada por elementos iguais a 1 na diagonal e elementos nulos fora da diagonal. Para obter a inversa de  $A$ , basta resolvermos  $n$  sistemas lineares do tipo  $Ax^{(i)} = b^{(i)}$ , onde  $x^{(i)}$  será a  $i$ -ésima coluna da matriz inversa e  $b^{(i)}$  será a  $i$ -ésima coluna da identidade,  $1 \leq i \leq n$ .

## 2. Objetivos e conteúdo do Módulo 2

O objetivo deste módulo será o de estudar métodos numéricos aplicados a sistemas lineares: métodos baseado em escalonamento de matriz e métodos iterativos. Para você ficar familiarizado com os métodos apresentados nesse módulo, serão utilizados vários recursos: resolução de listas de exercícios; leitura de materiais didáticos e elaboração de algoritmos computacionais.

## Conteúdos básicos do Módulo 2

- Método Iterativo de Jacobi e de Gauss - Seidel.
- Método da Eliminação de Gauss e da Decomposição LU.

### 3. Métodos Iterativos: Jacobi e Gauss - Seidel

Inicialmente, será apresentado um sistema linear, com quatro equações e quatro incógnitas, que servirá de motivação para a dedução dos métodos de Jacobi e de Gauss – Seidel.

O sistema linear  $Ax = b$ , com  $A = \begin{pmatrix} -1.0 & 0.5 & -0.1 & 0.1 \\ 0.2 & -0.6 & -0.2 & -0.1 \\ -0.1 & -0.2 & -1.5 & 0.2 \\ -0.1 & -0.3 & -0.2 & -1.0 \end{pmatrix}$ ,  $b = \begin{pmatrix} 0.2 \\ -2.6 \\ 1.0 \\ -2.5 \end{pmatrix}$  e

$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$  pode ser reescrito, em termos de equações lineares, como segue:

$$-1.0x_1 + 0.5x_2 - 0.1x_3 + 0.1x_4 = 0.2;$$

$$0.2x_1 - 0.6x_2 - 0.2x_3 - 0.1x_4 = -2.6;$$

$$-0.1x_1 - 0.2x_2 - 1.5x_3 + 0.2x_4 = 1.0;$$

$$-0.1x_1 - 0.3x_2 - 0.2x_3 - 1.0x_4 = -2.5 .$$

Como os elementos da diagonal da matriz  $A$  são não nulos, o seguinte sistema equivalente de equações lineares é obtido:

$$x_1 = \{ 0.2 - 0.5x_2 + 0.1x_3 - 0.1x_4 \} \div \{-1.0\};$$

$$x_2 = \{-2.6 - 0.2x_1 + 0.2x_3 + 0.1x_4\} \div \{-0.6\};$$

$$x_3 = \{ 1.0 + 0.1x_1 + 0.2x_2 - 0.2x_4 \} \div \{-1.5\};$$

$$x_4 = \{-2.5 + 0.1x_1 + 0.3x_2 + 0.2x_3\} \div \{-1.0\}.$$

Tanto o método de Jacobi quanto o de Gauss – Seidel partem da mesma estrutura anterior, porém os esquemas numéricos diferem na forma em que as iterações são efetuadas.

Partindo-se de um vetor inicial  $x^{(0)}$ , as iterações com o método de Jacobi são dadas por:

$$(x_1)^{(k)} = \{ 0.2 - 0.5(x_2)^{(k-1)} + 0.1(x_3)^{(k-1)} - 0.1(x_4)^{(k-1)} \} \div \{-1.0\};$$

$$(x_2)^{(k)} = \{-2.6 - 0.2(x_1)^{(k-1)} + 0.2(x_3)^{(k-1)} + 0.1(x_4)^{(k-1)}\} \div \{-0.6\};$$

$$(x_3)^{(k)} = \{ 1.0 + 0.1(x_1)^{(k-1)} + 0.2(x_2)^{(k-1)} - 0.2(x_4)^{(k-1)} \} \div \{-1.5\};$$

$$(x_4)^{(k)} = \{-2.5 + 0.1(x_1)^{(k-1)} + 0.3(x_2)^{(k-1)} + 0.2(x_3)^{(k-1)}\} \div \{-1.0\}.$$

As iterações com o método de Gauss – Seidel são dadas por:

$$(x_1)^{(k)} = \{ 0.2 - 0.5(x_2)^{(k-1)} + 0.1(x_3)^{(k-1)} - 0.1(x_4)^{(k-1)} \} \div \{-1.0\};$$

$$(x_2)^{(k)} = \{-2.6 - 0.2(x_1)^{(k)} + 0.2(x_3)^{(k-1)} + 0.1(x_4)^{(k-1)}\} \div \{-0.6\};$$

$$(x_3)^{(k)} = \{ 1.0 + 0.1(x_1)^{(k)} + 0.2(x_2)^{(k)} - 0.2(x_4)^{(k-1)} \} \div \{-1.5\};$$

$$(x_4)^{(k)} = \{-2.5 + 0.1(x_1)^{(k)} + 0.3(x_2)^{(k)} + 0.2(x_3)^{(k)}\} \div \{-1.0\}.$$

Nas próximas seções, apresentaremos com mais detalhes as deduções desses dois métodos.

### 3.1 Método de Jacobi

Dado um sistema de  $n$  equações lineares,  $Ax = b$ , podemos reescrevê-lo de forma equivalente como  $(L + D + U)x = b$ , onde  $L$  é a parte inferior da matriz  $A$ , ou seja,  $\ell_{ij} = 0$ , se  $i \leq j$ , e  $\ell_{ij} = a_{ij}$ , se  $i > j$ ;  $D$  é a diagonal da matriz  $A$ , isto é,  $d_{ij} = 0$ , se  $i \neq j$  e  $d_{ii} = a_{ii}$ ; e  $U$  é a parte superior da matriz  $A$ , ou seja,  $u_{ij} = 0$ , se  $i \geq j$ , e  $u_{ij} = a_{ij}$ , se  $i < j$ . Se  $D$  tem entradas não nulas (os elementos da diagonal da matriz  $A$  são não nulos) então  $Ax = b$  é equivalente ao sistema  $Dx = b - (L + U)x$  que, por sua vez, é equivalente ao sistema  $x = D^{-1}b - D^{-1}(L + U)x$ , onde  $D^{-1}$  é a matriz diagonal inversa da matriz  $D$  e possui elementos dados por  $1/a_{ii}$ ,  $1 \leq i \leq n$ .

Sejam a matriz  $C_J$  e o vetor  $g_J$  definidos como segue,  $C_J = -D^{-1}(L + U)$  e  $g_J = D^{-1}b$ . Então o sistema  $x = C_J x + g_J$  é equivalente ao sistema original  $Ax = b$ . Como  $L + U = A - D$ , é fácil notar que os elementos da matriz  $C_J$  são dados por  $c_{ij} = \frac{-a_{ij}}{a_{ii}}$ , se  $i \neq j$ , e

$c_{ii} = 0$ . Já o vetor  $g_J$  possui elementos dados por  $g_i = \frac{b_i}{a_{ii}}$ .

**Exemplo 1. (Sistema equivalente)** Considere o sistema linear  $Ax = b$ , onde

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}; \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ -2 \end{pmatrix}. \quad \text{Então, de acordo com a notação}$$

$$\text{anterior, obtemos: } L = \begin{pmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}; \quad D = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -2 \end{pmatrix}; \quad U = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix};$$

$$D^{-1} = \begin{pmatrix} 1/5 & 0 & 0 \\ 0 & -1/4 & 0 \\ 0 & 0 & -1/2 \end{pmatrix}; \quad C_J = \begin{pmatrix} 0 & -1/5 & -1/5 \\ 3/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix} \quad \text{e} \quad g_J = \begin{pmatrix} 5/5 \\ 8/-4 \\ -2/-2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

**Definição 1 (Método Iterativo de Jacobi).** Partindo-se de um vetor inicial,  $x^{(0)}$ , o método de Jacobi vai produzir uma sequência de vetores com termo geral dado por  $x^{(k)} = C_J x^{(k-1)} + g_J$ , onde  $C_J = -D^{-1}(L + U)$  é a matriz de iteração e  $g_J = D^{-1}b$  é o vetor de termos independentes.

**Exemplo 2. (Método de Jacobi)** O sistema linear  $Ax = b$ , onde  $A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}$ ,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ -2 \end{pmatrix} \quad \text{tem solução } x = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}. \quad \text{Considerando } x^{(0)} = \begin{pmatrix} 0.95 \\ -0.95 \\ 0.95 \end{pmatrix},$$

após duas iterações com o método de Jacobi, obtemos:

$$\begin{aligned} (x_1)^{(1)} &= \{ 5 - 1(x_2)^{(0)} - 1(x_3)^{(0)} \} \div \{5\} = 1; \\ (x_2)^{(1)} &= \{ 8 - 3(x_1)^{(0)} - 1(x_3)^{(0)} \} \div \{-4\} = -1.05; \\ (x_3)^{(1)} &= \{-2 - 1(x_1)^{(0)} - 1(x_2)^{(0)}\} \div \{-2\} = 1. \end{aligned}$$

$$\begin{aligned} (x_1)^{(2)} &= \{ 5 - 1(x_2)^{(1)} - 1(x_3)^{(1)} \} \div \{5\} = 1.01; \\ (x_2)^{(2)} &= \{ 8 - 3(x_1)^{(1)} - 1(x_3)^{(1)} \} \div \{-4\} = -1; \\ (x_3)^{(2)} &= \{-2 - 1(x_1)^{(1)} - 1(x_2)^{(1)}\} \div \{-2\} = 0.975. \end{aligned}$$

**Definição 2 (Critério de Parada do Erro Relativo).** Considere um sistema linear  $Ax = b$ , com solução única dada pelo vetor  $x$  e uma tolerância  $\varepsilon = 0.5 \times 10^{-s}$ ,  $s \in \mathbb{N}$ . O processo iterativo  $x^{(k)} = C_J x^{(k-1)} + g_J$ , com vetor inicial  $x^{(0)}$  conhecido, é interrompido se o seguinte critério de parada do erro relativo for satisfeito:  $ER^{(k)} = \frac{\|x^{(k)} - x^{(k-1)}\|_{max}}{\|x^{(k)}\|_{max}} < \varepsilon$ .

Nesse caso, o vetor  $x^{(k)}$  será uma aproximação do vetor solução  $x$  com a precisão estabelecida pela tolerância  $\epsilon$ .

**Observação:** A norma do máximo do vetor  $x = (x_1, x_2, \dots, x_n)$  é dada por:

$$\|x\|_{\max} = \max\{|x_1|, |x_2|, \dots, |x_n|\}.$$

**Exemplo 3. (Critério de Parada)** Vamos calcular o erro relativo cometido tanto na primeira quanto na segunda iteração apresentada no **Exemplo 2**. Antes, porém, note que

o vetor diferença  $x^{(1)} - x^{(0)}$  tem coordenadas  $\begin{pmatrix} 1 - 0.95 \\ -1.05 + 0.95 \\ 1 - 0.95 \end{pmatrix} = \begin{pmatrix} 0.05 \\ -0.1 \\ 0.05 \end{pmatrix}$  e o vetor

$x^{(2)} - x^{(1)}$  tem coordenadas  $\begin{pmatrix} 1.01 - 1 \\ -1 + 1.05 \\ 0.975 - 1 \end{pmatrix} = \begin{pmatrix} 0.01 \\ 0.05 \\ -0.025 \end{pmatrix}$ . Dessa forma,

$$ER^{(1)} = \frac{\|x^{(1)} - x^{(0)}\|_{\max}}{\|x^{(1)}\|_{\max}} = 0.1/1.05 \approx 0.09524 < 0.5 \times 10^{-1};$$

$$ER^{(2)} = \frac{\|x^{(2)} - x^{(1)}\|_{\max}}{\|x^{(2)}\|_{\max}} = 0.05/1.01 \approx 0.049505 < 0.5 \times 10^{-1}. \quad \blacksquare$$

**Observação:** Em vez de considerar a forma matricial do método de Jacobi, conforme a **Definição 1**, podemos definir o processo iterativo em termos das coordenadas do vetor

$x^{(k)} = C_J x^{(k-1)} + g_J$ :  $(x_i)^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(k-1)} \right)$ ,  $1 \leq i \leq n$ . Retorne ao **Exemplo 2** e perceba que as duas iterações foram efetuadas de acordo com a expressão anterior.

### 3.2 Método de Gauss - Seidel

Na tentativa de acelerar a convergência do processo iterativo que aproxima a solução do sistema linear  $Ax = b$ , considere a seguinte modificação no método de Jacobi:

$$(x_1)^{(k)} = \frac{1}{a_{11}} \left( b_1 - \sum_{j=2}^n a_{1j} x_j^{(k-1)} \right),$$

$$(x_i)^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k-1)} \right), \quad 2 \leq i \leq n-1;$$

$$(x_n)^{(k)} = \frac{1}{a_{nn}} \left( b_n - \sum_{j=1}^{n-1} a_{nj} x_j^{(k)} \right).$$

O esquema anterior é conhecido como método iterativo de Gauss – Seidel.

Observe que a primeira equação desse novo método é idêntica à primeira equação do método de Jacobi e, nesse caso, somente as coordenadas do vetor  $x^{(k-1)}$  são utilizadas. Nas demais equações, os índices,  $j$ , das coordenadas do vetor  $x^{(k-1)}$  correspondem a índices de colunas que estão depois da posição diagonal (índices associados à matriz  $U$  – veja o início da Seção 3.1), ou seja,  $i + 1 \leq j \leq n$ . A partir da segunda equação, os valores das coordenadas do vetor atualizado  $x^{(k)}$  também são utilizadas no processo iterativo. Note que os índices,  $j$ , dessas coordenadas correspondem a índices de colunas que estão antes da posição diagonal (índices associados à matriz  $L$  – veja o início da Seção 3.1), ou seja,  $1 \leq j \leq i - 1$ . Tal fato irá garantir a aceleração da convergência do processo iterativo.

Considerando as equações anteriores e as notações utilizadas no início da Seção 3.1, vamos reagrupar as coordenadas do vetor  $x^{(k)}$  de modo que  $(D + L)x^{(k)} = b - Ux^{(k-1)}$ . Portanto, em termos matriciais, podemos reescrever as equações anteriores de acordo com a

**Definição 3 (Método Iterativo de Gauss - Seidel).** Partindo-se de um vetor inicial,  $x^{(0)}$ , o método de Gauss – Seidel vai produzir uma sequência de vetores com termo geral dado por  $x^{(k)} = C_S x^{(k-1)} + g_S$ , onde  $C_S = -(D+L)^{-1}U$  é a matriz de iteração e  $g_S = (D+L)^{-1}b$  é o vetor de termos independentes.

**Exemplo 4. (Método de Gauss – Seidel)** Vamos considerar o mesmo sistema linear proposto no **Exemplo 2**, porém iremos fazer as iterações com o método de Gauss – Seidel, utilizando não a forma matricial, que é mais útil para a demonstração de teoremas de convergência, mas, sim, as iterações das coordenadas do vetor  $x^{(k)}$ :

$$(x_i)^{(k)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k-1)} \right), \quad 1 \leq i \leq n \text{ (quando } i = 1, \text{ o primeiro somatório}$$

é eliminado; quando  $i = n$ , o segundo somatório é eliminado). Dessa forma, sabendo que

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix} \text{ e } b = \begin{pmatrix} 5 \\ 8 \\ -2 \end{pmatrix} \text{ e } x^{(0)} = \begin{pmatrix} 0.95 \\ -0.95 \\ 0.95 \end{pmatrix}, \text{ após duas iterações com o método}$$

de Gauss - Seidel, obtemos:

$$(x_1)^{(1)} = \{ 5 - 1(x_2)^{(0)} - 1(x_3)^{(0)} \} \div \{5\} = 1;$$

$$(x_2)^{(1)} = \{ 8 - 3(x_1)^{(1)} - 1(x_3)^{(0)} \} \div \{-4\} = -1.0125;$$

$$(x_3)^{(1)} = \{-2 - 1(x_1)^{(1)} - 1(x_2)^{(1)}\} \div \{-2\} = 0.99375.$$

$$(x_1)^{(2)} = \{ 5 - 1(x_2)^{(1)} - 1(x_3)^{(1)} \} \div \{5\} = 1.00375;$$

$$(x_2)^{(2)} = \{ 8 - 3(x_1)^{(2)} - 1(x_3)^{(1)} \} \div \{-4\} = -0.99875;$$

$$(x_3)^{(2)} = \{-2 - 1(x_1)^{(2)} - 1(x_2)^{(2)}\} \div \{-2\} = 1.0025.$$

Agora, vamos calcular o erro relativo cometido na segunda iteração do método de Gauss – Seidel e compará-lo com o erro do método de Jacobi. Antes, porém, note que o vetor

diferença  $x^{(2)} - x^{(1)}$  tem coordenadas  $\begin{pmatrix} 1.00375 - 1 \\ -0.99875 + 1.0125 \\ 1.0025 - 0.99375 \end{pmatrix} = \begin{pmatrix} 0.00375 \\ 0.01375 \\ 0.00875 \end{pmatrix}$ . Dessa forma,

$$ER^{(2)} = \frac{\|x^{(2)} - x^{(1)}\|_{\max}}{\|x^{(2)}\|_{\max}} = 0.01375/1.00375 \approx 0.0137 < 0.049505, \text{ que é o valor do erro}$$

relativo associado à segunda iteração do método de Jacobi. ■

### 3.3 Convergência dos Método Iterativos de Jacobi e de Gauss – Seidel

Para estudar a convergência dos métodos iterativos, vamos precisar da noção de norma de uma matriz quadrada de ordem  $n$ . Aqui, nesse curso, utilizaremos a norma linha de matriz.

**Definição 4 (Norma Linha de matriz).** Seja  $C$  uma matriz de ordem  $n$  com elementos  $c_{ij}$ ,  $1 \leq i \leq n$  e  $1 \leq j \leq n$ . A norma linha da matriz  $C$  é definida como:

$\|C\|_L = \text{máximo} \{L_1, L_2, \dots, L_n\}$ , onde  $L_i = \sum_{j=1}^n |c_{ij}|$  (que é a soma dos valores absolutos dos elementos da linha  $i$  da matriz  $C$ ),  $1 \leq i \leq n$ .

**Exemplo 5. (Norma Linha)** Vamos calcular a norma linha da matriz de iteração do método de Jacobi associada ao sistema linear que foi dado no **Exemplo 1**. A matriz de

iteração é dada por  $C_J = \begin{pmatrix} 0 & -1/5 & -1/5 \\ 3/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$ . Utilizando a **Definição 4**, tem-se que

$L_1 = |0| + |-1/5| + |-1/5| = 2/5$ ;  $L_2 = |3/4| + |0| + |1/4| = 1$  e  $L_3 = |1/2| + |1/2| + |0| = 1$ . Portanto,  $\|C\|_L = \text{máximo} \{L_1, L_2, L_3\} = \text{máximo} \{2/5, 1, 1\} = 1$ .

**Observação:** No exemplo anterior, a norma linha da matriz de iteração do método de Jacobi resultou igual a 1. Nesse caso, fica um pouco mais difícil demonstrar que o método numérico é convergente. Porém, se  $\|C\| < 1$ , então o método iterativo  $x^{(k)} = C x^{(k-1)} + g$  sempre será convergente, independente da norma utilizada e independente do vetor inicial  $x^{(0)}$ . Antes de enunciarmos esse resultado, considere a definição mais geral de uma norma de matriz.

**Definição 5 (Norma de matriz).** Uma norma de matriz deve satisfazer as seguintes condições:

(N1)  $0 \leq \|M\|$ , qualquer que seja  $M$ , matriz quadrada de ordem  $n$ . Além disso,  $\|M\| = 0$  se, e somente se,  $M = 0$  (matriz nula).

(N2)  $\|M + N\| \leq \|M\| + \|N\|$ , quaisquer que sejam  $M$  e  $N$ , matrizes quadradas de ordem  $n$ .

(N3)  $\|tM\| = |t| \|M\|$ , para todo número real  $t$  e para qualquer  $M$ , matriz quadrada de ordem  $n$ .

**Observação:** Dizemos que uma norma de matriz é consistente com uma norma de vetor se  $\|Cx\| \leq \|C\| \|x\|$ , onde  $C$  é uma matriz de ordem  $n$  e  $x$  é um vetor coluna com  $n$  coordenadas. A norma linha de matriz, por exemplo, é consistente com a norma do máximo de vetor, ou seja,  $\|Cx\|_{\max} \leq \|C\|_L \|x\|_{\max}$ .

**Teorema 1 (Condição Suficiente de Convergência de métodos iterativos aplicados a sistemas lineares)** Considere um sistema linear  $Ax = b$ , com solução única dada pelo vetor  $x$  e que seja equivalente ao sistema linear  $x = Cx + g$ . Se  $\|C\| < 1$  e a norma de matriz for consistente com a norma de vetor, então o processo iterativo dado por  $x^{(k)} = Cx^{(k-1)} + g$  será converge independentemente da escolha do vetor inicial  $x^{(0)}$ .

**Demonstração:**  $0 \leq \|x^{(k)} - x\| = \|Cx^{(k-1)} + g - Cx - g\| = \|C(x^{(k-1)} - x)\| \leq \|C\| \|x^{(k-1)} - x\|$ . Repetindo-se o mesmo argumento anterior e utilizando indução finita, obtém-se que

$$0 \leq \|x^{(k)} - x\| \leq \|C\|^k \|x^{(0)} - x\|.$$

Como  $\|C\| < 1$ , então  $\lim_{k \rightarrow \infty} \|C\|^k = 0$ . Assim, pelo Teorema do Confronto (veja a

**Observação (\*\*)** posterior ao **Exercício 2** da Seção 4 do **Módulo 1**), segue-se que  $\lim_{k \rightarrow \infty} \|x^{(k)} - x\| = 0$  ou, equivalentemente,  $\lim_{k \rightarrow \infty} x^{(k)} = x$ . ■



## Critério das Linhas

Existe uma maneira prática para verificar se o método de Jacobi será convergente. Esse critério de convergência é conhecido como Critério das Linhas. Basta lembrar que os elementos da matriz  $C_J$  estão relacionados com os elementos da matriz  $A$  do sistema

linear de modo que:  $c_{ij} = \frac{-a_{ij}}{a_{ii}}$ , se  $i \neq j$ , e  $c_{ii} = 0$ . Além disso,  $\|C_J\|_L < 1$  se, e somente

se,  $L_i = \sum_{j=1}^n |c_{ij}| < 1$ . Portanto,  $\|C_J\|_L < 1 \leftrightarrow \sum_{j=1}^n |a_{ij}| < |a_{ii}|$ ,  $1 \leq i \leq n$ . Quando essa última condição é satisfeita dizemos que a matriz  $A$  é diagonal dominante.

**Exemplo 6. (Critério das Linhas)** Considere o sistema linear  $Ax = b$ , com

$$A = \begin{pmatrix} -1.0 & 0.5 & -0.1 & 0.1 \\ 0.2 & -0.6 & -0.2 & -0.1 \\ -0.1 & -0.2 & -1.5 & 0.2 \\ -0.1 & -0.3 & -0.2 & -1.0 \end{pmatrix}, \quad b = \begin{pmatrix} 0.2 \\ -2.6 \\ 1.0 \\ -2.5 \end{pmatrix} \quad \text{e} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Note que a matriz  $A$  é diagonal dominante, ou seja,

$$1.0 = |a_{11}| > |a_{12}| + |a_{13}| + |a_{14}| = 0.7; \quad 0.6 = |a_{22}| > |a_{21}| + |a_{23}| + |a_{24}| = 0.5;$$

$$1.5 = |a_{33}| > |a_{31}| + |a_{32}| + |a_{34}| = 0.5; \quad 1.0 = |a_{44}| > |a_{41}| + |a_{42}| + |a_{43}| = 0.6.$$

Portanto, o Critério das Linhas é satisfeito; logo o método de Jacobi será convergente. ■

## Critério de Sassenfeld

A matriz de iteração  $C_S = -(D+L)^{-1}U$  do método de Gauss – Seidel não é tão simples de ser calculada. Por essa razão, vamos obter um critério de convergência para o método de Gauss – Seidel que seja mais prático de ser avaliado. Essa condição suficiente de convergência é conhecida como Critério de Sassenfeld.

O Critério de Sassenfeld é uma variação do critério das linhas e é descrito a seguir.

$$\beta_1 = L_1 = \frac{1}{|a_{11}|} \sum_{j=2}^n |a_{1j}|,$$

$$\beta_i = \frac{1}{|a_{ii}|} \sum_{j=1}^{i-1} |a_{ij}| \beta_j + \frac{1}{|a_{ii}|} \sum_{j=i+1}^n |a_{ij}|, \quad 2 \leq i < n;$$

$$\beta_n = \frac{1}{|a_{nn}|} \sum_{j=1}^{n-1} |a_{nj}| \beta_j .$$

Se  $\beta = \text{máximo } \{\beta_1, \beta_2, \dots, \beta_n\} < 1$ , então o método de Gauss – Seidel é convergente.

**Exercício 1. (Critério de Sassenfeld)** Mostre que se o Critério de Sassenfeld for satisfeito então o método de Gauss – Seidel será convergente.

**Solução do Exercício 1:**

Lembre-se de que  $x^{(k)} = C_S x^{(k-1)} + g_S$ , onde  $C_S = -(D+L)^{-1}U$  e  $g_S = (D+L)^{-1}b$ . Além disso, o sistema  $Ax = b$  é equivalente ao sistema  $x = C_S x + g_S$ . O erro na iteração  $k$  do método de Gauss – Seidel será representado por  $e^{(k)} = x - x^{(k)} = C_S x - C_S x^{(k-1)}$ . Portanto,  $e^{(k)} = C_S e^{(k-1)} \Leftrightarrow De^{(k)} = -(L e^{(k)} + U e^{(k-1)}) \Leftrightarrow -e^{(k)} = D^{-1}(L e^{(k)} + U e^{(k-1)})$ . Portanto, utilizando a notação do início da Seção 3.2 e as expressões de  $\beta_i$  dadas anteriormente, tem-se que

$$(x_1)^{(k)} - x_1 = \frac{1}{a_{11}} \sum_{j=2}^n a_{1j} (x_j - x_j^{(k-1)}) \rightarrow |(x_1)^{(k)} - x_1| \leq \frac{1}{|a_{11}|} \sum_{j=2}^n |a_{1j}| |x_j - x_j^{(k-1)}| \rightarrow$$

$$|(x_1)^{(k)} - x_1| \leq \frac{1}{|a_{11}|} \sum_{j=2}^n |a_{1j}| \|e^{(k-1)}\|_{\max} = \|e^{(k-1)}\|_{\max} \beta_1. \text{ Portanto,}$$

$$|(e_1)^{(k)}| \leq \beta_1 \|e^{(k-1)}\|_{\max}.$$

Por indução finita, suponhamos que  $|(e_j)^{(k)}| \leq \beta_j \|e^{(k-1)}\|_{\max}$ , para todo  $j$  tal que  $1 \leq j \leq i-1$ .

Dessa forma, tem-se que

$$|(x_i)^{(k)} - x_i| \leq \frac{1}{|a_{ii}|} \sum_{j=1}^{i-1} |a_{ij}| |x_j - x_j^{(k)}| + \frac{1}{|a_{ii}|} \sum_{j=i+1}^n |a_{ij}| |x_j - x_j^{(k-1)}|, \quad 2 \leq i \leq n-1. \text{ Portanto,}$$

$$|(x_i)^{(k)} - x_i| \leq \frac{1}{|a_{ii}|} \sum_{j=1}^{i-1} |a_{ij}| \|e_j^{(k-1)}\|_{\max} \beta_j + \frac{1}{|a_{ii}|} \sum_{j=i+1}^n |a_{ij}| \|e_j^{(k-1)}\|_{\max} \leq \beta_i \|e^{(k-1)}\|_{\max},$$

$2 \leq i \leq n-1$ . Além disso,

$$|(x_n)^{(k)} - x_n| \leq \frac{1}{|a_{nn}|} \sum_{j=1}^{n-1} |a_{nj}| |x_j - x_j^{(k)}| \leq \frac{1}{|a_{nn}|} \sum_{j=1}^{n-1} |a_{nj}| \|e_j^{(k-1)}\|_{\max} \beta_j \leq \beta_n \|e^{(k-1)}\|_{\max}.$$

Conclusão:  $\|e^{(k)}\|_{\max} \leq \beta \|e^{(k-1)}\|_{\max}$ . (Lembre-se de que  $\beta = \text{máximo } \{\beta_1, \beta_2, \dots, \beta_n\} < 1$ ). Portanto, por indução finita,  $0 \leq \|e^{(k)}\|_{\max} \leq \beta^k \|e^{(0)}\|_{\max}$ . Novamente, o Teorema do Confronto garante que  $\lim_{k \rightarrow \infty} \|e^{(k)}\|_{\max} = 0 \Leftrightarrow \lim_{k \rightarrow \infty} x^{(k)} = x$ . ■

**Observação:** Se o critério das linhas for satisfeito, então o critério de Sassenfeld também será satisfeito. Verifique este resultado! Portanto, antes de efetuar os cálculos mais custosos dos  $\beta_i$ , avalie, de forma mais rápida, se a matriz do sistema linear é diagonal dominante.

**Exemplo 7. (Critério de Sassenfeld)** Fizemos duas iterações com o método de Gauss – Seidel no **Exemplo 4**. Vamos mostrar que o Critério de Sassenfeld é satisfeito para o

$$\text{sistema linear } Ax = b, \text{ onde } A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ e } b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ -2 \end{pmatrix}.$$

De fato,  $\beta_1 = |1/5| + |1/5| = 2/5 < 1$ ;  $\beta_2 = \beta_1 |-3/4| + |1/4| = 11/20 < 1$  e  $\beta_3 = \beta_1 |-1/2| + \beta_2 |-1/2| = 1/5 + 11/40 = 19/40 < 1$ . ■

## Critério Geral de Convergência

Um Teorema importante (condição necessária e suficiente) sobre convergência de métodos iterativos aplicados a sistemas lineares é enunciado a seguir, sem demonstração.

**Teorema 2 (Condição Necessária e Suficiente de Convergência de métodos iterativos aplicados a sistemas lineares)** O processo iterativo  $x^{(k)} = C x^{(k-1)} + g$ , com qualquer vetor inicial  $x^{(0)}$ , converge para a solução do sistema linear  $Ax = b$ , que é equivalente ao sistema  $x = Cx + g$ , se, e somente se,  $|\lambda| < 1$ , para todo auto-valor  $\lambda$  da matriz de iteração  $C$ .

**Observação:** Os auto-valores,  $\lambda$ , de uma matriz  $C$  de ordem  $n$  são tais que  $\det(C - \lambda I) = 0$ , em que  $I$  é a matriz identidade e  $\det(C - \lambda I)$  é o determinante da matriz  $C - \lambda I$ . Esta teoria faz parte do conteúdo da disciplina Álgebra Linear.

**Observação:** A demonstração do **Teorema 2** pode ser encontrada em livros de Análise Numérica. Algumas referências são apresentadas a seguir.



## Referências

[1] BURDEN, R.L. e FAIRES, J.D. *Análise Numérica*. Tradução da 8ª edição norte-americana, São Paulo: Cengage Learning, 2008.

[2] ISSACSON, E. and KELLER, H. B., *Analysis of numerical methods*, New York, John Wiley and Sons, 1982.

[3] ORTEGA, J. M., *Numerical analysis; a second course*, New York, Academic Press, 1972.



O livro *Álgebra Linear* da editora HARBRA, dos autores José Luiz Boldrini, Sueli I. Rodrigues Costa, Vera Lúcia F. F. Ribeiro e Henry G. Wetzler, contém um capítulo sobre processos iterativos aplicados à resolução de sistemas lineares. Vale a pena dar uma conferida.

**Exemplo 8. (Critério de geral de convergência)** Fizemos duas iterações com o método de Jacobi no **Exemplo 2** e vimos que o critério das linhas não era satisfeito, no **Exemplo 5**. Agora, vamos mostrar, utilizando o **Teorema 2**, que o método de Jacobi é

convergente quando aplicado ao sistema linear  $Ax = b$ , onde  $A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}$ ,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ e } b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ -2 \end{pmatrix}. \text{ De fato, a matriz de iteração é dada por}$$

$$C_J = \begin{pmatrix} 0 & -1/5 & -1/5 \\ 3/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}. \text{ Calculando-se } \det(C_J - \lambda I), \text{ obtemos o seguinte}$$

polinômio:  $p(\lambda) = -\lambda^3 - (\lambda/8) - 1/10$  (veja o exercício 8 da Atividade 1 do Módulo 1). Uma das raízes desse polinômio pode ser encontrada pelo método de Newton-Raphson ( $\lambda_1 = -0.375713238$ ) e as outras duas podem ser obtidas através da resolução de uma equação do segundo grau, após a utilização do algoritmo de Briot-Ruffini:

$$-(\lambda^3 + (\lambda/8) + 1/10) \div (\lambda + 0.375713238) = -(\lambda^2 - 0.375713238\lambda + 0.266160437).$$

Assim, as outras duas raízes de  $p$  são dadas por  $\lambda_2 = 0.187856619 + 0.480489674 i$  e  $\lambda_3 = 0.187856619 - 0.480489674 i$ . Note que  $\lambda_2 = a + bi$  e  $\lambda_3 = a - bi$ . Portanto,  $|\lambda_1| = 0.375713238$  e os módulos das raízes complexos são  $|\lambda_2| = |\lambda_3| = \sqrt{a^2 + b^2} \approx 0.51591$ . Dessa forma, pelo **Teorema 2**, o método de Jacobi será convergente. ■

## 4. Métodos Diretos baseados em Escalonamento de Matriz

Os métodos de escalonamento consistem em realizar operações elementares nas linhas da matriz  $A$  de um sistema linear,  $Ax = b$ , de modo a transformá-la em uma matriz triangular superior  $U$ . Uma matriz  $U$  é dita triangular superior quando os seus elementos satisfazem  $u_{ij} = 0$ , nos casos em que  $i > j$ .

O sistema linear  $Ax = b$  possui a mesma solução do sistema linear escalonado  $Ux = B$ , onde o vetor  $B$  foi obtido do vetor  $b$  por meio das mesmas operações elementares (aplicadas nas linhas da matriz do sistema) que transformaram  $A$  em  $U$ .

Para recordar, as operações elementares são 3:

(e1)  $L_i \leftrightarrow L_j$ : a linha  $i$  é trocada com a linha  $j$ ;

(e2)  $L_i \rightarrow L_i + t.L_j$ ,  $t \neq 0$ : a linha  $i$  é trocada pela soma da linha  $i$  com a linha  $j$  multiplicada por um número real  $t$  não nulo (se  $t$  for nulo então a matriz não sofrerá nenhuma alteração).

**Observação.** A operação anterior (e2) pode ser invertida, ou seja, para recuperar a linha  $i$  que foi alterada basta executar a seguinte operação:  $L_i \rightarrow L_i - t.L_j$ . Além disso, esse tipo de operação elementar não altera o valor do determinante da matriz.

(e3)  $L_i \rightarrow t.L_i$ ,  $t \neq 0$ : a linha  $i$  é trocada pela linha  $i$  multiplicada por um número real  $t$  não nulo.

**Observação:** A operação anterior (e3) também pode ser invertida, ou seja, para recuperar a linha  $i$  que foi alterada basta executar a seguinte operação:  $L_i \rightarrow (t^{-1}).L_i$ .

### 4.1 Método da Eliminação de Gauss

O Método da Eliminação de Gauss vai considerar, inicialmente, apenas a operação elementar,  $L_i \rightarrow L_i + t.L_j$ ,  $t \neq 0$  (e2), que não altera o determinante da matriz do sistema linear original,  $Ax = b$ . Se a matriz  $A$  for de ordem  $n$ , então serão necessárias  $n - 1$  etapas para se efetuar o escalonamento que transformará a matriz  $A$  em uma matriz triangular superior  $U$ .

Em cada etapa  $k$ ,  $1 \leq k \leq n - 1$ , o objetivo do Método de Eliminação de Gauss será o de anular os elementos  $a_{ik}$  que estão abaixo da posição do elemento da diagonal,  $a_{kk}$ . Portanto, para cada  $i$ ,  $k + 1 \leq i \leq n$ , será preciso escolher  $t = -m_{ik}$  de modo que a operação elementar  $L_i \rightarrow L_i - m_{ik}.L_k$  produza um novo valor para  $a_{ik}$ ,  $(a_{ik})^{(k)}$ , que seja nulo, ou seja,  $(a_{ik})^{(k)} = 0$ . De acordo com a operação elementar utilizada, tem-se que

$$0 = (a_{ik})^{(k)} = a_{ik} - m_{ik} a_{kk}. \text{ Assim, } m_{ik} = \frac{a_{ik}}{a_{kk}}.$$

Depois da etapa do escalonamento, vem a etapa conhecida como retro-solução, que consiste na resolução do sistema equivalente. Lembre-se de que a matriz dos coeficientes desse sistema linear equivalente é triangular superior.

Antes de analisar o caso geral, consideremos um caso particular no qual a matriz triangular superior,  $U$ , tem ordem 3. Nesse caso, o sistema linear oriundo do escalonamento é dado por:

$$\begin{aligned} u_{11} x_1 + u_{12} x_2 + u_{13} x_3 &= B_1; \\ u_{22} x_2 + u_{23} x_3 &= B_2; \\ u_{33} x_3 &= B_3. \end{aligned}$$

Assim, a solução desse sistema linear é obtida como segue:

$$\begin{aligned} x_3 &= B_3/u_{33}; \\ x_2 &= (1/u_{22})\{B_2 - u_{23} x_3\}; \\ x_1 &= (1/u_{11})\{B_1 - u_{12} x_2 - u_{13} x_3\}. \end{aligned}$$

Analogamente, no caso geral, a solução é dada por:

$$\begin{aligned} x_n &= B_n/u_{nn}; \\ x_i &= \frac{1}{u_{ii}}\left\{B_i - \sum_{j=i+1}^n u_{ij} x_j\right\}; \quad i \text{ variando de } n-1 \text{ até } 1. \end{aligned}$$

Dessa forma, podemos elaborar um algoritmo que contém duas partes. A primeira parte do algoritmo transformará a matriz  $A$  do sistema linear  $Ax = b$  em uma matriz triangular superior  $U$ . A segunda parte do algoritmo obterá a solução do sistema linear equivalente,  $Ux = B$ .

## Algoritmo: Eliminação de Gauss – Escalonamento e Retro-solução

O algoritmo que será apresentado a seguir está configurado para ser executado no *software* Octave. Observe que os comentários que são escritos no programa devem ser precedidos pelo símbolo de porcentagem: “%”.

Para não sobrecarregar a notação, no algoritmo exibido a seguir, sempre que uma operação elementar (do tipo e2) for realizada, as notações para os elementos da matriz e para os termos independentes serão as mesmas, isto é,  $a_{ij}$  e  $b_i$  respectivamente. Isso significa que, em vez de utilizarmos uma notação diferente para a matriz triangular superior ( $U$ ) e uma notação para o termo independente modificado com o escalonamento ( $B$ ), continuaremos a empregar a mesma notação original para os elementos da matriz ( $a(i,j)$ : notação usada no algoritmo) e para os termos independentes

(b(i): notação usada no algoritmo).

% Escalonamento de um sistema linear com n incógnitas

```
for k = 1: n - 1          % o índice de coluna k varia de 1 até n-1
    if(a(k,k) != 0)        % o termo da diagonal não é zero
        for i = k + 1:n    % i está variando de k + 1 até n, variando de 1 em 1.
            % cálculo do multiplicador; variável mult
            mult = (a(i,k))/a(k,k);
            if(mult != 0)   % se mult for diferente de zero
                %alteração do termo independente
                b(i) = b(i) - mult*b(k);
                %alteração das linhas
                for j = k+1:n    % j está variando de k+1 até n porque a(i,k) = 0
                    a(i,j) = a(i,j) - mult*a(k,j);
                end
            end
        end
    end
    %fim do “if”: mult diferente de zero
    %fim do laço para o índice i
else
    % se a(k,k) for zero
    fprintf('Impossível continuar com o escalonamento');
end
% fim do procedimento if-else
end
% fim do laço para o índice k; fim do escalonamento
```

%Retro-solução

```
x(n) = b(n)/a(n,n);      % n-ésima coordenada do vetor solução
for i = n-1:-1:1          % i variando de n-1e decrescendo uma unidade até chegar a 1
    soma = x(i+1)*a(i,i+1); %primeira parcela do somatório
    for j = (i+2):n
        soma = soma + x(j)*a(i,j); %soma das demais parcelas; j variando de i+2 até n
    end
    x(i) = (b(i) - soma)/a(i,i); % i-ésiam coordenada do vetor solução
end
```

**Exemplo 9. (Eliminação de Gauss)** Vamos utilizar o Método da Eliminação de Gauss

para resolver o seguinte sistema linear:  $Ax = b$ , onde  $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}$ ,  $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$  e

$b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ . Primeiramente, faremos o escalonamento da matriz ampliada  $(A | b)$ .

**Observação:** Como os multiplicadores são úteis para uma importante fatoração da matriz  $A$  (isto ficará evidente no final deste exemplo), seus valores serão armazenados justamente naquelas posições da matriz para as quais o processo de escalonamento produziu elementos nulos. Na matriz ampliada, estas posições serão destacadas com os valores dos multiplicadores colocados entre parênteses.

Etapa 1 do escalonamento ( $k = 1$ ):

$$\begin{aligned} m_{21} &= a_{21}/a_{11}; & L_2 &\rightarrow L_2 - m_{21}L_1; \\ m_{31} &= a_{31}/a_{11}; & L_3 &\rightarrow L_3 - m_{31}L_1. \end{aligned}$$

Dessa forma, obtemos a seguinte matriz ampliada:

$$A^{(1)} = \left( \begin{array}{ccc|c} 3 & -4 & 1 & 1 \\ (m_{21}=1/3) & 10/3 & 5/3 & -1/3 \\ (m_{31}=4/3) & 16/3 & -13/3 & -4/3 \end{array} \right).$$

Etapa 2 do escalonamento ( $k = 2$ ):

$$m_{32} = a_{32}/a_{22}; \quad L_3 \rightarrow L_3 - m_{32}L_2.$$

Dessa forma, obtemos a seguinte matriz ampliada:

$$A^{(2)} = \left( \begin{array}{ccc|c} 3 & -4 & 1 & 1 \\ (m_{21}=1/3) & 10/3 & 5/3 & -1/3 \\ (m_{31}=4/3) & (m_{32}=8/5) & -7 & -4/5 \end{array} \right).$$

Agora efetuaremos a retro-solução.

Observe que a matriz triangular superior e o novo termo independente são dados por

$$U = \begin{pmatrix} 3 & -4 & 1 \\ 0 & 10/3 & 5/3 \\ 0 & 0 & -7 \end{pmatrix} \quad \text{e} \quad B = \begin{pmatrix} 1 \\ -1/3 \\ -4/5 \end{pmatrix}. \quad \text{Assim, o vetor solução do sistema linear}$$



tem as seguintes coordenadas:

$$\begin{aligned}x_3 &= B_3/u_{33} = 4/35; \\x_2 &= (1/u_{22})\{B_2 - u_{23} x_3\} = -11/70; \\x_1 &= (1/u_{11})\{B_1 - u_{12} x_2 - u_{13} x_3\} = 3/35.\end{aligned}$$



**Observação:** Se considerarmos uma matriz triangular inferior,  $L$ , formada pelos multiplicadores que foram obtidos ao longo do processo de escalonamento e colocarmos

1 na diagonal desta matriz, obteremos:  $L = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 8/5 & 1 \end{pmatrix}$ .

É fácil ver que  $LU = A$ . De fato,

$$LU = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 8/5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 & -4 & 1 \\ 0 & 10/3 & 5/3 \\ 0 & 0 & -7 \end{pmatrix} = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix} = A.$$

O teorema mais geral sobre esse resultado, conhecido como decomposição LU, pode ser encontrado em:



[4] FRANCO, Neide Bertoldi, *Cálculo Numérico*, São Paulo, Pearson Referências Prentice Hall, 2006.

## Esforço Computacional

No algoritmo anterior (Eliminação de Gauss), vamos contar o número de operações de adição (subtração) e de multiplicação (divisão). Esse valor nos dará uma ideia do esforço computacional para se resolver um sistema linear com matriz dos coeficientes de ordem  $n$ .

### Adição e subtração no escalonamento

Observe que o índice  $k$  varia de 1 até  $n - 1$ . Além disso, quando o índice  $i$  varia de  $k + 1$  até  $n$ , uma subtração é realizada na alteração do termo independente e uma subtração é realizada para cada índice  $j$  variando de  $k + 1$  até  $n$ , na alteração das linhas da matriz. Note que não estamos computando aquelas operações que produzirão elementos nulos. Por essa razão, o algoritmo apresenta o índice  $j$  variando de  $k + 1$  até  $n$ , em vez de variar de  $k$  até  $n$ . Seja  $n_{ae}$  o número total de operações de adições e subtrações efetuadas no escalonamento, então

$$n_{ae} = \sum_{k=1}^{n-1} \{[n-(k+1)+1]+1\}[n-(k+1)+1] = \sum_{k=1}^{n-1} [(n-k)+1][n-k] \rightarrow$$

$$\rightarrow n_{ae} = \sum_{k=1}^{n-1} [(n-k)^2 + (n-k)] = \sum_{k=1}^{n-1} [(n^2 + n) - (2n+1)k + k^2].$$

Sabendo que  $\sum_{k=1}^{n-1} k = \frac{(n-1)n}{2}$  e  $\sum_{k=1}^{n-1} k^2 = \frac{(n-1)(2n-1)n}{6}$ , então  $n_{ae} = n^3/3 - n/3$ .

### Multiplicação e divisão no escalonamento

Observe que o índice  $k$  varia de 1 até  $n - 1$ . Além disso, quando o índice  $i$  varia de  $k + 1$  até  $n$ , uma divisão é realizada no cálculo do multiplicador e uma multiplicação na alteração do termo independente e, para cada índice  $j$  variando de  $k + 1$  até  $n$ , uma multiplicação é realizada na alteração das linhas da matriz. Seja  $n_{me}$  o número total de operações de multiplicações e divisões efetuadas no escalonamento, então

$$n_{me} = \sum_{k=1}^{n-1} \{[n-(k+1)+1]+2\}[n-(k+1)+1] = \sum_{k=1}^{n-1} [(n-k)+2][n-k] \rightarrow$$

$$\rightarrow n_{me} = \sum_{k=1}^{n-1} [(n-k)^2 + 2(n-k)] = \sum_{k=1}^{n-1} [(n-k)^2 + (n-k)] + \sum_{k=1}^{n-1} n-k.$$

Utilizando as informações anteriores, obtemos que  $n_{me} = n^3/3 + n^2/2 - (5n)/6$ .

### Adição e subtração na retro-solução

Quando o índice  $i$  varia de  $n-1$  até 1, uma subtração é realizada no cálculo da coordenada  $x(i)$  ao final da variação do índice  $j$ , de  $i + 2$  até  $n$ , e para cada variação do índice  $j$  é efetuada uma adição no cálculo da variável soma. Seja  $n_{ar}$  o número total de operações de adições e subtrações efetuadas na retro-solução, então

$$n_{ar} = \sum_{i=1}^{n-1} \{[n-(i+2)+1]+1\} = \sum_{i=1}^{n-1} (n-i) = n(n-1) - \frac{(n-1)n}{2} = \frac{n^2}{2} - \frac{n}{2}.$$

### Multiplicação e divisão na retro-solução

Uma divisão é efetuada no cálculo da coordenada  $x(n)$ . Além disso, quando o índice  $i$  varia de  $n - 1$  até 1, uma multiplicação é efetuada no cálculo da variável soma (primeira parcela) e uma divisão é realizada no cálculo da coordenada  $x(i)$  ao final da variação do índice  $j$ : de  $i + 2$  até  $n$ . Note que para cada variação do índice  $j$  é efetuada uma multiplicação no cálculo da variável soma (demais parcelas do somatório). Seja  $n_{mr}$  o número total de operações de multiplicações e divisões efetuadas na retro-solução, então

$$n_{mr} = 1 + \sum_{i=1}^{n-1} \{[n-(i+2)+1]+2\} = \sum_{i=1}^{n-1} (n+1-i) = (n+1)(n-1) - \frac{(n-1)n}{2}.$$

Assim,  $n_{mr} = 1 + \frac{n^2}{2} + \frac{n}{2} - 1 = n^2/2 + n/2.$

Juntando as informações sobre o número de operações efetuadas no escalonamento e na retro-solução, obtemos os seguintes valores:

$$n_{ae} + n_{me} = [n^3/3 - n/3] + [n^3/3 + n^2/2 - (5n)/6] = 2n^3/3 + n^2/2 - 7n/6;$$

$$n_{ar} + n_{mr} = n^2/2 - n/2 + [n^2/2 + n/2] = n^2.$$

Portanto, no escalonamento são efetuadas  $\frac{2n^3}{3} + \frac{n^2}{2} - \frac{7n}{6}$  operações e na retro-solução são executadas  $n^2$  operações, sendo tais operações de adição, subtração, multiplicação e divisão.

## Decomposição LU e inversão de matriz

A técnica apresentada a seguir pode ser aplicada em qualquer matriz  $A$  de ordem  $n$ , desde que  $A = LU$ . Note que  $\det(A) = \det(L) \cdot \det(U) = 1 \cdot \det(U) = u_{11} \cdot u_{22} \dots u_{nn}$ , pois a matriz  $L$  é triangular inferior com elementos unitários na diagonal e a matriz  $U$  é triangular superior. Portanto, a matriz  $A$  terá inversa,  $A^{-1}$ , se ao final do escalonamento – método de Eliminação de Gauss – não aparecer nenhum elemento nulo na diagonal.

Note que  $A \cdot A^{-1} = I$  (matriz identidade). Então, multiplicando-se a matriz  $A$  pela primeira coluna de  $A^{-1}$ , obtém-se a primeira coluna da matriz identidade; multiplicando-se a matriz  $A$  pela segunda coluna de  $A^{-1}$ , obtém-se a segunda coluna da matriz identidade, e assim por diante. Portanto, para se obter  $A^{-1}$ , basta resolvermos  $n$  sistemas lineares do tipo  $Ax^{(j)} = b^{(j)}$ , onde  $A$  é uma matriz de ordem  $n$ ,  $x^{(j)}$  é o vetor que representa a coluna  $j$  da matriz  $A^{-1}$  e  $b^{(j)}$  é a coluna  $j$  da matriz identidade, ou seja, a coordenada  $j$  do vetor  $b^{(j)}$  é igual a 1 e as demais coordenadas são iguais a zero.

Como  $A = LU$  então o sistema  $Ax = b$  é equivalente ao sistema  $LUx = b$ . Utilizando um vetor auxiliar,  $y = Ux$ , podemos resolver o sistema anterior em duas etapas:

(1ª) Resolução do sistema triangular inferior:  $Ly = b$ ;

(2ª) Resolução do sistema triangular superior:  $Ux = y$ .

**Exemplo 10. (Inversão de matriz)** Vamos obter a inversa da matriz  $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}$  dada no **Exemplo 9**. Conforme os cálculos efetuados naquele exemplo,

tem-se que  $L = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 8/5 & 1 \end{pmatrix}$  e  $U = \begin{pmatrix} 3 & -4 & 1 \\ 0 & 10/3 & 5/3 \\ 0 & 0 & -7 \end{pmatrix}$ . O primeiro sistema

linear, com  $b^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ , já foi resolvido e obtivemos  $x^{(1)} = \begin{pmatrix} 3/35 \\ -11/70 \\ 4/35 \end{pmatrix}$ , que é a

primeira coluna da matriz inversa. Vamos resolver os outros dois sistemas utilizando a decomposição LU.

### Sistema 2

$$\text{(sistema triangular inferior)} \quad \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 8/5 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -8/5 \end{pmatrix};$$

$$\text{(s. triang. sup.)} \quad \begin{pmatrix} 3 & -4 & 1 \\ 0 & 10/3 & 5/3 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -8/5 \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6/35 \\ 13/70 \\ 8/35 \end{pmatrix}.$$

### Sistema 3

$$\text{(sistema triangular inferior)} \quad \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 4/3 & 8/5 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix};$$

$$\text{(s. triang. sup.)} \quad \begin{pmatrix} 3 & -4 & 1 \\ 0 & 10/3 & 5/3 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/7 \\ 1/14 \\ -1/7 \end{pmatrix}.$$

Conclusão: a matriz inversa é dada por  $A^{-1} = \begin{pmatrix} 3/35 & 6/35 & 1/7 \\ -11/70 & 13/70 & 1/14 \\ 4/35 & 8/35 & 1/7 \end{pmatrix}$ . ■

## 4.2 Método da Eliminação de Gauss com Pivoteamento Parcial

O Método de Eliminação de Gauss com Pivoteamento Parcial foi desenvolvido para aperfeiçoar o Método de Eliminação de Gauss em relação a duas questões: (i) realização

do escalonamento com a possibilidade de se trocar as linhas da matriz do sistema linear;  
(ii) diminuição dos efeitos de erros de arredondamento.

Para motivar o estudo desse método, vejamos um exemplo que o professor José Bezerra Leite apresentou, na primeira parte da disciplina Cálculo Numérico, para os alunos da turma de 1986 do curso de Matemática da Unesp de Rio Claro. Esta disciplina foi dividida com o professor José Manoel Balthazar, também professor do Departamento de Matemática.

**Exemplo 11. (Pivoteamento Parcial)** Vamos resolver o sistema linear

$$\begin{aligned}0.0001x_1 + 1x_2 &= 1; \\ 1x_1 + 1x_2 &= 2;\end{aligned}$$

com o Método da Eliminação de Gauss, de duas maneiras diferentes: sem permutação e com permutação das equações. Nas duas resoluções, todos os cálculos serão efetuados considerando-se aritmética de ponto flutuante com 3 dígitos e arredondamento simétrico, ou seja, todos os números deverão ser representados da seguinte forma:  $0.a_1a_2a_3 \times 10^s$ ,  $s \in \mathbb{Z}$ . Por exemplo, o número 0.9998 será arredondado para 1.000 e representado por  $0.1 \times 10^1$  e o número 1.0001 também será arredondado para 1.000 e terá a mesma representação anterior.

Aplicando-se o Método da Eliminação de Gauss ao sistema linear, sem troca de linhas, obtemos  $m_{21} = (0.1 \times 10^1)/(0.1 \times 10^{-3}) = 1 \times 10^4 = 0.1 \times 10^5$ . Efetuando-se a operação elementar  $L_2 \rightarrow L_2 - m_{21}L_1$ , a matriz ampliada será dada por

$$\left( \begin{array}{cc|c} 0.1 \times 10^{-3} & 0.1 \times 10^1 & 0.1 \times 10^1 \\ (m_{21}=0.1 \times 10^5) & -0.1 \times 10^5 & -0.1 \times 10^5 \end{array} \right).$$

observe que

$$0.1 \times 10 - 0.1 \times 10^5 \times 0.1 \times 10^1 = 0.1 \times 10 - 0.01 \times 10^6 = 0.1 \times 10 - 0.1 \times 10^5 = 0.00001 \times 10^5 - 0.1 \times 10^5 = 0.000 - 0.1 \times 10^5 = -0.1 \times 10^5;$$

$$0.2 \times 10 - 0.1 \times 10^5 \times 0.1 \times 10^1 = 0.2 \times 10 - 0.01 \times 10^6 = 0.2 \times 10 - 0.1 \times 10^5 = 0.00002 \times 10^5 - 0.1 \times 10^5 = 0.000 - 0.1 \times 10^5 = -0.1 \times 10^5.$$

Portanto, o método de Eliminação de Gauss vai gerar a seguinte solução  $x_2 = 1.000$  e  $x_1 = 0.000$ , que não é a solução do sistema linear original.

Trocando-se as equações obtemos o sistema equivalente:

$$\begin{aligned}1x_1 + 1x_2 &= 2; \\ 0.0001x_1 + 1x_2 &= 1.\end{aligned}$$

Aplicando-se o Método da Eliminação de Gauss ao sistema linear anterior, obtemos  $m_{21} = (0.1 \times 10^{-3}) / (0.1 \times 10^1) = 0.1 \times 10^{-3}$ . Após a operação elementar  $L_2 \rightarrow L_2 - m_{21} \cdot L_1$ , a matriz ampliada será dada por

$$\left( \begin{array}{cc|c} 0.1 \times 10^1 & 0.1 \times 10^1 & 0.2 \times 10^1 \\ (m_{21} = 0.1 \times 10^{-3}) & 0.1 \times 10^1 & 0.1 \times 10^1 \end{array} \right).$$

observe que

$$0.1 \times 10 - 0.1 \times 10^{-3} \times 0.1 \times 10^1 = 0.1 \times 10 - 0.1 \times 10^{-3} = 0.1 \times 10 - 0.00001 \times 10 = 0.1 \times 10^1 - 0.000 = 0.1 \times 10^1;$$

$$0.1 \times 10 - 0.1 \times 10^{-3} \times 0.2 \times 10^1 = 0.1 \times 10 - 0.02 \times 10^{-2} = 0.1 \times 10 - 0.00002 \times 10 = 0.1 \times 10^1 - 0.000 = 0.1 \times 10^1.$$

Portanto, o método de Eliminação de Gauss com Pivoteamento vai gerar a seguinte solução  $x_2 = 1.000$  e  $x_1 = 1.000$ , que é a solução do sistema linear original com três casas decimais de precisão. ■

Outros exemplos desse tipo podem ser encontrado no livro da Márcia A. Gomes Ruggiero e Vera Lúcia da Rocha Lopes:



*Cálculo Numérico: Aspectos Teóricos e Computacionais*, 2ª Edição, São Paulo, Pearson Education do Brasil, 1996. Eu utilizei esse livro (1ª Ed.) em minha graduação e hoje, durante as aulas de cálculo numérico, o usufruto continua ativo.

## Estratégia de Pivoteamento Parcial

O Método da Eliminação de Gauss com Pivoteamento Parcial vai considerar as operações elementares: (e1)  $L_i \leftrightarrow L_j$  e (e2)  $L_i \rightarrow L_i + t \cdot L_j$ ,  $t \neq 0$ . A primeira operação inverte o sinal do determinante e a segunda não altera o determinante da matriz do sistema linear original,  $Ax = b$ . Se a matriz  $A$  for de ordem  $n$  então serão necessárias  $n - 1$  etapas para se efetuar o escalonamento que transformará a matriz  $A$  em uma matriz triangular superior  $U$ .

Em cada estágio  $k$ ,  $1 \leq k \leq n - 1$ , o objetivo do Método de Eliminação de Gauss com Pivoteamento será o de anular os elementos  $a_{ik}$  que estão abaixo da posição do elemento da diagonal,  $a_{kk}$ . Porém, esse elemento que vai aparecer na diagonal (chamado de pivô) deve ter valor absoluto elevado quando comparado com os demais valores absolutos dos elementos da matriz. Esse procedimento ameniza os problemas relacionados a erros de arredondamento.

Para evitar maiores complicações provenientes de possíveis trocas de colunas, vamos apresentar uma estratégia de pivoteamento parcial que consiste em calcular, em cada

estágio  $k$  que antecede o processo de escalonamento, o valor máximo dos módulos dos elementos da coluna  $k$ , da diagonal para baixo. Esse valor máximo será representado por  $P_k = \text{máximo } \{|a_{ik}|, k \leq i \leq n\}$ . Se  $|a_{i_{\max},k}|$  for o elemento máximo e  $i_{\max} > k$  então, antes de efetuarmos o escalonamento, a linha  $i_{\max}$  deverá ser trocada com a linha  $k$ . Se o elemento com o valor módulo já estiver na diagonal então seguimos direto para o processo de escalonamento, calculando-se os multiplicadores da mesma forma que no Método de Eliminação de Gauss. O processo referente à retro-solução é exatamente o mesmo já visto anteriormente.

**Exemplo 12. (Pivoteamento Parcial)** Vamos utilizar o Método da Eliminação de Gauss com Pivoteamento Parcial para escalonar a matriz  $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}$ .

**Observação:** Como os multiplicadores são úteis para uma importante fatoração da matriz  $A$  (isto ficará evidente no final deste exemplo), seus valores serão armazenados justamente naquelas posições da matriz para as quais o processo de escalonamento produziu elementos nulos. Essas posições serão destacadas com os valores dos multiplicadores colocados entre parênteses. Caso ocorra troca de linhas devido ao pivoteamento parcial, os valores dos multiplicadores deverão ser trocados de lugar também, seguindo o mesmo critério do pivoteamento.

**Observação:** Para guardar as trocas de linhas efetuadas durante a aplicação do método, vamos utilizar uma variável  $\tau$ , que, inicialmente, assumirá o valor zero ( $\tau := 0$ ), e um vetor de índices  $ind = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ . Toda vez que ocorrer troca de linhas, a variável  $\tau$  será acrescida de uma unidade ( $\tau := \tau + 1$ ) e o vetor de índices terá as suas coordenadas permutadas de acordo com o pivoteamento parcial.

Estágio 1 ( $k = 1$ ):

Pivoteamento

$$P_1 = \text{máximo } \{|a_{i1}|, 1 \leq i \leq 3\} = |a_{i_{\max},1}| = 4; \quad i_{\max} = 3 > 1 = k.$$

Assim, a linha 3 deve ser trocada com a linha 1. Logo,  $ind = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$  e  $\tau = 1$ .

Agora, temos que considerar a seguinte matriz:  $\begin{pmatrix} 4 & 0 & -3 \\ 1 & 2 & 2 \\ 3 & -4 & 1 \end{pmatrix}$ .

Escalonamento (os elementos da matriz anterior continuarão sendo denotados por  $a_{ij}$ )

$$\begin{aligned}m_{21} &= a_{21}/a_{11}; & L_2 &\rightarrow L_2 - m_{21}.L_1; \\m_{31} &= a_{31}/a_{11}; & L_3 &\rightarrow L_3 - m_{31}.L_1.\end{aligned}$$

Dessa forma obtemos a seguinte matriz:

$$A^{(1)} = \begin{pmatrix} 4 & 0 & -3 \\ (m_{21}=1/4) & 2 & 11/4 \\ (m_{31}=3/4) & -4 & 13/4 \end{pmatrix}.$$

Estágio 2 ( $k = 2$ ):

Pivoteamento

$$P_2 = \text{máximo } \{|a_{i2}|, 2 \leq i \leq 3\} = |a_{i_{\max},2}| = 4; \quad i_{\max} = 3 > 2 = k.$$

Assim, a linha 3 deve ser trocada com a linha 2. Logo,  $ind = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$  e  $\tau = 2$ .

Agora, temos que considerar a seguinte matriz:  $\begin{pmatrix} 4 & 0 & -3 \\ (3/4) & -4 & 13/4 \\ (1/4) & 2 & 11/4 \end{pmatrix}$ . Observe que

os valores dos multiplicadores foram trocados de lugar, porém continuam armazenados naquelas posições que seriam ocupadas pelos elementos que foram zerados pelo processo de escalonamento.

Escalonamento (os elementos da matriz anterior continuarão sendo denotados por  $a_{ij}$ )

$$m_{32} = a_{32}/a_{22}; \quad L_3 \rightarrow L_3 - m_{32}.L_2.$$

Dessa forma obtemos a seguinte matriz:

$$A^{(2)} = \begin{pmatrix} 4 & 0 & -3 \\ (3/4) & -4 & 13/4 \\ (1/4) & (m_{32}=-1/2) & 35/8 \end{pmatrix}.$$



Considerando-se os resultados obtidos no exemplo anterior, podemos fazer as seguintes observações.

**Observação:** No processo de pivoteamento e escalonamento ocorreram duas trocas de linhas ( $\tau = 2$ ); assim,  $\det(A) = \det(A^{(2)}).(-1)^\tau = 4.(-4).(35/8).(-1)^2 = -70$ .



**Observação:** Se considerarmos uma matriz triangular inferior,  $L$ , formada pelos multiplicadores que foram obtidos ao longo do processo de pivoteamento e escalonamento e colocarmos 1 na diagonal dessa matriz, obteremos:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix}. \text{ Note que } LU = \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 & -3 \\ 0 & -4 & 13/4 \\ 0 & 0 & 35/8 \end{pmatrix} =$$

$$\begin{pmatrix} 4 & 0 & -3 \\ 3 & -4 & 1 \\ 1 & 2 & 2 \end{pmatrix} = PA, \text{ onde } P \text{ é o produto de matrizes de permutação (veja a}$$

observação abaixo). Na prática, não precisamos da matriz de permutação mas, sim, do vetor de índices,  $ind = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$ , obtido no final da aplicação do método da Eliminação de

Gauss com Pivoteamento Parcial. A análise que devemos fazer desse vetor é a seguinte: a linha 3 da matriz original será a primeira linha da matriz  $PA$ , a linha 1 da matriz original será a segunda linha da matriz  $PA$  e a linha 2 da matriz original será a terceira linha da matriz  $PA$ .



Saiba Mais

**Observação:** Uma matriz de permutação é obtida da matriz identidade a partir de uma, e apenas uma, troca de linhas. Dessa forma, toda matriz de permutação possui inversa. No **Exemplo 12**, foram executadas duas trocas de linhas. Primeiro, a linha 1 foi trocada com a linha 3; depois, a linha 2 foi trocada com a linha 3. Fazendo cada uma destas trocas de linhas na matriz identidade, obteremos:

$$P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ e } P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \text{ A matriz } P \text{ é dada por } P_{23}P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note que o efeito de multiplicar uma matriz de permutação por uma matriz  $A$  é o de produzir uma permutação na matriz  $A$  com as mesmas características da permutação que foi realizada na identidade:

$$PA = P_{23}P_{13}A = P_{23}P_{13} \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix} = P_{23} \begin{pmatrix} 4 & 0 & -3 \\ 1 & 2 & 2 \\ 3 & -4 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & -3 \\ 3 & -4 & 1 \\ 1 & 2 & 2 \end{pmatrix}.$$

$$PA = P_{23}P_{13}A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 4 & 0 & -3 \\ 3 & -4 & 1 \\ 1 & 2 & 2 \end{pmatrix}.$$



### Saiba Mais

**Observação:** Como  $PA = LU$ , então o sistema linear  $Ax = b$  é equivalente ao sistema  $PAx = Pb$ , pois  $P$  é invertível (produto de matrizes invertíveis), que por sua vez é equivalente ao sistema  $LUx = Pb$ . Utilizando um vetor auxiliar,  $y = Ux$ , podemos resolver o sistema linear anterior em duas etapas:

(1ª) Resolução do sistema triangular inferior:  $Ly = Pb$ ;

(2ª) Resolução do sistema triangular superior:  $Ux = y$ .

### Exercício 2. (Decomposição LU com pivoteamento parcial e inversão de matrizes)

Obtenha a inversa da matriz  $A = \begin{pmatrix} -3 & 4 & -1 \\ -1 & -2 & -2 \\ -4 & 0 & 3 \end{pmatrix}$  utilizando a decomposição LU obtida do Método da Eliminação de Gauss com Pivoteamento Parcial.

### Solução do Exercício 2:

Utilizando-se o **Exemplo 12**, pode-se concluir que  $L = \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix}$  e  $U =$

$\begin{pmatrix} -4 & 0 & 3 \\ 0 & 4 & -13/4 \\ 0 & 0 & -35/8 \end{pmatrix}$ . Para se obter a inversa da matriz  $A$ , precisaremos resolver 3

sistemas lineares do tipo  $Ax^{(j)} = b^{(j)} \leftrightarrow PAx^{(j)} = Pb^{(j)} \leftrightarrow LUx^{(j)} = Pb^{(j)}$ , onde  $x^{(j)}$  é o vetor que representa a coluna  $j$  da matriz  $A^{-1}$  e  $Pb^{(j)}$  é a coluna  $j$  da matriz identidade permutada de acordo com o pivoteamento parcial (controlado pelo vetor de índices).

Como o vetor de índices, que controla as trocas de linhas, é dado por  $ind = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$  então,

para qualquer vetor  $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ , tem-se que  $Pv = \begin{pmatrix} v_3 \\ v_1 \\ v_2 \end{pmatrix}$ .

### Sistema 1

(triangular inferior)  $\begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = P \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1/2 \end{pmatrix};$

$$\text{(triang. superior)} \begin{pmatrix} -4 & 0 & 3 \\ 0 & 4 & -13/4 \\ 0 & 0 & -35/8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1/2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -3/35 \\ 11/70 \\ -4/35 \end{pmatrix}.$$

### Sistema 2

$$\text{(triangular inferior)} \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = P \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix};$$

$$\text{(triang. superior)} \begin{pmatrix} -4 & 0 & 3 \\ 0 & 4 & -13/4 \\ 0 & 0 & -35/8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -6/35 \\ -13/70 \\ -8/35 \end{pmatrix}.$$

### Sistema 3

$$\text{(triangular inferior)} \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = P \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -3/4 \\ -5/8 \end{pmatrix};$$

$$\text{(triang. superior)} \begin{pmatrix} -4 & 0 & 3 \\ 0 & 4 & -13/4 \\ 0 & 0 & -35/8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -3/4 \\ -5/8 \end{pmatrix} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1/7 \\ -1/14 \\ 1/7 \end{pmatrix}.$$

$$\text{Conclusão: a matriz inversa é dada por } A^{-1} = - \begin{pmatrix} 3/35 & 6/35 & 1/7 \\ -11/70 & 13/70 & 1/14 \\ 4/35 & 8/35 & 1/7 \end{pmatrix}. \quad \square$$

## **5. Aplicação – Discretização da equação do calor para uma barra homogênea**

A equação do calor para uma barra homogênea de comprimento  $L$  é dada por

$$u_t(x, t) = Ku_{xx}(x, t) + f(x, t), \quad x \in I = (x_0, x_f), \quad x_f - x_0 = L \quad \text{e} \quad t \in J = (t_0, T), \quad t_0 \geq 0, \quad K > 0,$$

com condições de fronteira:  $u(x_0, t) = g_1(t)$ ,  $t \in J$ ;  $u(x_f, t) = g_2(t)$ ,  $t \in J$ ; e com condição inicial dada por  $u(x, t_0) = q(x)$ ,  $x \in [x_0, x_f]$ .

A discretização no tempo vai utilizar o método de Crank – Nicolson e no espaço vai utilizar diferenças finitas centradas. Nas aproximações dadas a seguir, a notação  $O(h^p) = R \cdot h^p$ , onde  $R$  é uma constante.

Vamos considerar pontos igualmente espaçados no intervalo  $J$ , particionando-o em  $n_t$  subintervalos. Assim,  $h_t = \Delta t = (T-t_0)/n_t$  e  $t_j = t_0 + jh_t$ ,  $1 \leq j \leq n_t$ . Também iremos considerar pontos igualmente espaçados no intervalo  $I$ , particionando-o em  $n_x$  subintervalos, assim  $h_x = \Delta x = (x_f - x_0)/n_x$  e  $x_i = x_0 + ih_x$ ;  $1 \leq i \leq n_x$ .

Em primeiro lugar, deduziremos uma fórmula para aproximar a derivada segunda de uma função  $V(x)$ . Para isso, utilizaremos os seguintes desenvolvimentos de Taylor:

$$V(x+h) = V(x) + V'(x)h + V''(x)h^2/2! + V'''(c)h^3/3!, \text{ c entre x e x+h;}$$

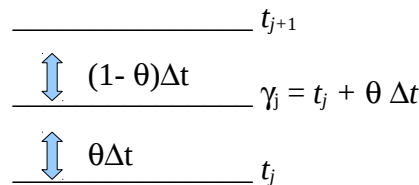
$$V(x-h) = V(x) - V'(x)h + V''(x)h^2/2! + V'''(d)h^3/3!, \text{ d entre x-h e x.}$$

Portanto,  $V(x+h) + V(x-h) = 2V(x) + V''(x)h^2 + O(h^3)$ . Assim,

$$V''(x) \approx (1/h^2) \{V(x-h) - 2V(x) + V(x+h)\}.$$

Se  $h = h_x$  e  $V(x_i) = u(x_i, t_j)$  então  $u_{xx}(x_i, t_j) \approx \frac{1}{h_x^2} \{u(x_{i-1}, t_j) - 2u(x_i, t_j) + u(x_{i+1}, t_j)\}$ . Essa última expressão será denotada por  $\frac{1}{h_x^2} \{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}\}$ .

A figura abaixo representa um passo de tempo partindo-se de  $t_j$  e chegando-se a  $t_{j+1}$ ;  $\Delta t = t_{j+1} - t_j$  e  $\gamma_j = t_j + \theta \Delta t$  é um valor intermediário entre  $t_j$  e  $t_{j+1}$ , com  $0 < \theta < 1$ .



**Figura 1 – Módulo 2. Esquema de discretização do Método de Crank - Nicolson**

Seja  $U(t)$  uma função diferenciável, então

$$\begin{aligned} U(t_{j+1}) &= U(\gamma_j + (1-\theta)h_t) = U(\gamma_j) + (1-\theta)h_t U'(\gamma_j) + O_1((h_t)^2) \quad \text{e} \\ U(t_j) &= U(\gamma_j - \theta h_t) = U(\gamma_j) + (-\theta)h_t U'(\gamma_j) + O_2((h_t)^2), \end{aligned}$$

onde  $O_1((h_t)^2) = U''(c_j).(h_t)^2/2!$  e  $O_2((h_t)^2) = U''(d_j).(h_t)^2/2!$ ;  $c_j$  está entre  $\gamma_j$  e  $t_{j+1}$ ;  $d_j$  entre  $t_j$  e  $\gamma_j$ .

Dessa forma,

$$U(t_{j+1}) - U(t_j) = h_t U'(\gamma_j) + O((h_t)^2) \rightarrow [U(t_{j+1}) - U(t_j)]/h_t \approx U'(\gamma_j).$$

Além disso,  $\theta U(t_{j+1}) + (1 - \theta) U(t_j) = U(\gamma_j) + \theta O_1 + (1 - \theta) O_2 = U(\gamma_j) + O_3((h_t)^2)$ .  
Portanto,  $\theta U(t_{j+1}) + (1 - \theta) U(t_j) \approx U(\gamma_j)$ .

Considerando-se  $t = \gamma_j$  e  $x = x_i$  na equação do calor, obtém-se que

$$u_t(x_i, \gamma_j) = K u_{xx}(x_i, \gamma_j) + f(x_i, \gamma_j).$$

De acordo com os cálculos anteriores e supondo que a função  $u_{xx}$  seja derivável no tempo, obtemos

$$[u(x_i, t_{j+1}) - u(x_i, t_j)]/h_t \approx K\{\theta u_{xx}(x_i, t_{j+1}) + (1 - \theta) u_{xx}(x_i, t_j)\} + \theta f(x_i, t_{j+1}) + (1 - \theta) f(x_i, t_j).$$

Portanto,

$$u_{i,j+1} - u_{i,j} = h_t \theta K \frac{1}{h_x^2} (u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1}) + h_t (1 - \theta) K \frac{1}{h_x^2} (u_{i-1,j} - 2u_{i,j} + u_{i+1,j}) \\ + h_t \theta f(x_i, t_{j+1}) + h_t (1 - \theta) f(x_i, t_j).$$

Seja  $\rho = K \frac{h_t}{h_x^2} > 0$ , então

$$u_{i,j+1} - u_{i,j} = \rho \{ \theta (u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1}) + (1 - \theta) (u_{i-1,j} - 2u_{i,j} + u_{i+1,j}) \} \\ + h_t \theta f(x_i, t_{j+1}) + h_t (1 - \theta) f(x_i, t_j) \rightarrow$$

$$- \rho \theta u_{i-1,j+1} + (1 + 2\rho \theta) u_{i,j+1} - \rho \theta u_{i+1,j+1} = \rho (1 - \theta) u_{i-1,j} + (1 - 2\rho (1 - \theta)) u_{i,j} + \rho (1 - \theta) u_{i+1,j} \\ + h_t \theta f(x_i, t_{j+1}) + h_t (1 - \theta) f(x_i, t_j).$$

Quando  $\theta = 1/2$  obtemos o método de Crank – Nicolson:

$$- \rho u_{i-1,j+1} + (2 + 2\rho) u_{i,j+1} - \rho u_{i+1,j+1} = \rho u_{i-1,j} + (2 - 2\rho) u_{i,j} + \rho u_{i+1,j} \\ + h_t (f(x_i, t_{j+1}) + f(x_i, t_j)).$$

Observe que os termos que aparecem do lado esquerdo são as incógnitas e os que aparecem do lado direito são termos conhecidos. Precisamos utilizar a condição inicial da equação do calor para gerar o vetor inicial  $u_j = (u_{1,j}, u_{2,j}, \dots, u_{n,j})$ ,  $j = 0$ ;  $n = n_x - 1$ . As condições de fronteira são utilizadas para ajustar a primeira e a última equação do sistema linear, com vetor de incógnitas  $u_{j+1} = (u_{1,j+1}, u_{2,j+1}, \dots, u_{n,j+1})$ . A matriz desse sistema linear é estritamente diagonal dominante:  $|2 + 2\rho| > 2\rho = |-\rho| + |-\rho|$ . Portanto, o sistema pode ser resolvido pelo Método de Gauss–Seidel. Como a matriz é tridiagonal, então é bastante simples a resolução do sistema linear por meio da decomposição LU.

**Exemplo 13. (Crank-Nicolson; caso particular)** Neste exemplo, vamos considerar um problema proposto no livro de *Análise Numérica* de Richard Burden e J. Douglas Faires (veja as referências da Seção 3.3). A equação do calor é dada por

$$u_t(x, t) = u_{xx}(x, t), \quad x \in I = (0, 1), \quad t \in J = (0, 0.5),$$

sujeita às condições

$$(\text{fronteira}) \quad u(0, t) = 0; u(1, t) = 0, \quad \forall t \in J;$$

$$(\text{inicial}) \quad u(x, 0) = \text{sen}(\pi x), \quad x \in [0, 1].$$

A solução analítica dessa equação do calor é dada por  $u(x, t) = \exp(-\pi^2 x) \text{sen}(\pi x)$ .

Vamos obter aproximações de  $u(x, t)$  no tempo  $t = 0.5$  e nos pontos  $x = x_i$ , igualmente espaçados com espaçamento  $h_x = 0.1$ . Os passos de tempo terão comprimento  $h_t = 0.01$ .

Observe que  $\rho = K \frac{h_t}{h_x^2} = 1$ ;  $n_x = 10$  e  $n_t = 50$ . Em cada passo de tempo, temos que resolver um sistema linear do seguinte tipo

$$\begin{pmatrix} 2(1+\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\rho & 2(1+\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\rho & 2(1+\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\rho & 2(1+\rho) & -\rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\rho & 2(1+\rho) & -\rho & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\rho & 2(1+\rho) & -\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho & 2(1+\rho) & -\rho & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1+\rho) & -\rho & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1+\rho) & -\rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1+\rho) \end{pmatrix} \begin{pmatrix} u_{1,j+1} \\ u_{2,j+1} \\ u_{3,j+1} \\ u_{4,j+1} \\ u_{5,j+1} \\ u_{6,j+1} \\ u_{7,j+1} \\ u_{8,j+1} \\ u_{9,j+1} \end{pmatrix} =$$

$$\begin{pmatrix} 2(1-\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\rho & 2(1-\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\rho & 2(1-\rho) & -\rho & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\rho & 2(1-\rho) & -\rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\rho & 2(1-\rho) & -\rho & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\rho & 2(1-\rho) & -\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho & 2(1-\rho) & -\rho & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1-\rho) & -\rho & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1-\rho) & -\rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 2(1-\rho) \end{pmatrix} \begin{pmatrix} u_{1,j} \\ u_{2,j} \\ u_{3,j} \\ u_{4,j} \\ u_{5,j} \\ u_{6,j} \\ u_{7,j} \\ u_{8,j} \\ u_{9,j} \end{pmatrix} + \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \\ \Psi_5 \\ \Psi_6 \\ \Psi_7 \\ \Psi_8 \\ \Psi_9 \end{pmatrix}.$$

A primeira e a última coordenada do vetor  $\Psi$  englobam valores de fronteira e de termos de fonte; as demais coordenadas somente envolvem termos de fonte:

$$\Psi_1 = \rho(u_{0,j} + u_{0,j+1}) + h_t(f(x_1, t_{j+1}) + f(x_1, t_j)) = \rho(g_1(t_j) + g_1(t_{j+1})) + h_t(f(x_1, t_{j+1}) + f(x_1, t_j));$$

$$\Psi_i = h_t(f(x_i, t_{j+1}) + f(x_i, t_j)); \quad 1 \leq i \leq n-1;$$

$$\Psi_n = \rho(u_{n+1,j} + u_{n+1,j+1}) + h_t(f(x_{n+1}, t_{j+1}) + f(x_{n+1}, t_j)) = \rho(g_2(t_j) + g_2(t_{j+1})) + h_t(f(x_{n+1}, t_{j+1}) + f(x_{n+1}, t_j)).$$

No exemplo em questão, todas as coordenadas do vetor  $\Psi$  são nulas porque as condições de fronteira são nulas e o termo de fonte é nulo.

Utilizando a condição inicial,  $u(x, 0) = \sin(\pi x)$ , obtemos o seguinte vetor

$$\begin{pmatrix} u_{1,0} \\ u_{2,0} \\ u_{3,0} \\ u_{4,0} \\ u_{5,0} \\ u_{6,0} \\ u_{7,0} \\ u_{8,0} \\ u_{9,0} \end{pmatrix} = \begin{pmatrix} 0.30902 \\ 0.58779 \\ 0.80902 \\ 0.95106 \\ 1.00000 \\ 0.95106 \\ 0.80902 \\ 0.58779 \\ 0.30902 \end{pmatrix}.$$

Observe que a matriz do sistema linear é tridiagonal, ou seja, os elementos não nulos da matriz estão na diagonal inferior (estes valores são todos iguais a -1), na diagonal principal (estes valores são todos iguais a 4) e na diagonal superior (estes valores são todos iguais a -1). Portanto, para efeitos computacionais, esses valores podem ser armazenados em 3 vetores apenas, o que gera uma economia muito grande relacionada à memória do computador, já que a quantidade de elementos nulos é muito maior do que a quantidade de elementos não nulos.

A solução do sistema linear produzirá os seguintes valores:

Tabela de valores – **Exemplo 13** –  
 Solução aproximada (Crank-Nicolson;  $T = 0.5$ ) e solução analítica

$x_i$	$u_i$ (Crank - Nicolson)	$\exp(-\pi^2 T) \sin(\pi x_i)$ (solução analítica)
0	0	0
0.1	0.00230512	0.0022224
0.2	0.00438461	0.0042273
0.3	0.00603489	0.0058184
0.5	0.00709444	0.0068399
0.5	0.00745954	0.0071919
0.6	0.00709444	0.0068399
0.7	0.00603489	0.0058184
0.8	0.00438461	0.0042273
0.9	0.00230512	0.0022224
1.0	0	0



## 6. Atividades do Módulo 2

### Atividade 1 – Adquirindo e testando conhecimentos através da resolução da lista de exercícios

#### Enunciado da atividade

Em praticamente todos os exercícios desta lista você precisará lidar com sistemas lineares e matrizes; portanto, é fundamental que você recorde a teoria vista sobre este assunto na disciplina de Álgebra Linear. Essa lista contém exercícios referentes ao conteúdo do Módulo 2: Método Iterativo de Gauss - Jacobi e de Gauss – Seidel; Método da Eliminação de Gauss e Decomposição LU. Você tem que se esforçar para tentar fazer todos os exercícios.

#### Segunda Lista de Exercícios de Cálculo Numérico

1) Seja  $A = (a_{ij})$  uma matriz de ordem  $n$ , onde  $a_{ii} = -2$ ;  $a_{i,i-1} = 1$ ;  $a_{i,i+1} = 1$  e os demais elementos são nulos. Por exemplo, para  $n = 4$ ,  $A = \begin{pmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{pmatrix}$ ;  $A$  é uma matriz tridiagonal.

i) Mostre que o critério de Sassenfeld é válido, independentemente da ordem da matriz. Sugestões: (a) Calcule  $\beta_1$  e  $\beta_2$  e verifique que são menores do que 1; (b) Suponha que  $\beta_{i-1} < 1$  e prove que  $\beta_i < 1$ .

ii) Considere  $n = 3$ ,  $b^t = (-1, 0, -0.5)$  e  $[x^{(0)}]^t = (0, 0, 0)$ . Faça duas iterações com o método de Gauss-Seidel e calcule o erro relativo. Justifique a convergência desse método.

iii) O método de Jacobi também será convergente. Considere  $n = 3$  e calcule o módulo dos autovalores da matriz de iteração,  $C$ , do método de Jacobi. Os autovalores são as raízes do polinômio característico  $p(\lambda) = \det(C - \lambda I)$ ;  $I$  é a matriz identidade.

2) Considere o sistema linear  $Ax = b$ , com

$$A = \begin{pmatrix} -1.0 & 0.5 & -0.1 & 0.1 \\ 0.2 & -0.6 & -0.2 & -0.1 \\ -0.1 & -0.2 & -1.5 & 0.2 \\ -0.1 & -0.3 & -0.2 & -1.0 \end{pmatrix}, \quad b = \begin{pmatrix} 0.2 \\ -2.6 \\ 1.0 \\ -2.5 \end{pmatrix} \quad \text{e} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

i) Verifique se o critério de Sassenfeld é satisfeito.

ii) Faça duas iterações com o método de Jacobi; use  $x^{(0)} = (0, 0, 0, 0)^T$ .

iii) Utilize  $x^{(2)}$ , calculado no item anterior, como valor inicial e faça uma iteração com o método de Gauss-Seidel. Calcule o erro relativo.

3) Sejam  $A = \begin{pmatrix} 2 & 1 & 3 \\ 0 & -1 & 1 \\ 1 & 0 & 3 \end{pmatrix}$ ;  $x^t = (x_1 \ x_2 \ x_3)$  e  $b^t = (9 \ 1 \ 3)$ . Troque a 1ª linha com a 3ª linha; depois troque a 1ª coluna com a terceira coluna.

(i) Verifique se o critério das linhas é satisfeito.

(ii) Verifique se o critério de Sassenfeld é satisfeito.

(iii) Monte o esquema numérico que certamente convergirá e faça uma iteração partindo de  $[x^{(0)}]^t = (8.8 \ -2.7 \ -1.8)$ .

(iv) Calcule o erro relativo.

4) “Se o Critério das Linhas é satisfeito, então o Critério de Sassenfeld também é satisfeito”. Verifique esse fato, considerando uma matriz,  $A$ ,  $2 \times 2$ . Lembre-se de que

$$L_i = \frac{1}{|a_{ii}|} \sum_{j=1, j \neq i}^n |a_{ij}|, \quad 1 \leq i \leq n$$

$$\beta_i = \frac{1}{|a_{ii}|} \sum_{j=1}^{i-1} |a_{ij}| \beta_j + \frac{1}{|a_{ii}|} \sum_{j=i+1}^n |a_{ij}|, \quad 2 \leq i \leq n, \text{ e } \beta_1 = L_1.$$

5) Um aluno, após efetuar algumas iterações de um método iterativo para sistemas lineares, chamou o professor e pediu para ele conferir os cálculos. O professor constatou que algumas contas estavam erradas, mas, em vez de pedir para o aluno refazê-las, apenas disse para ele continuar os cálculos e prestar mais atenção, pois assim ele obterá a aproximação correta do sistema linear. Como justificar a atitude do professor? Será que contas erradas podem conduzir a resultados corretos? Use os conhecimentos de Cálculo Numérico para explicar o fenômeno ocorrido.

6) Dado o sistema linear

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

verifique que o critério das linhas não é satisfeito. No entanto, o método de Jacobi (e também o de Gauss-Seidel) é convergente. Para demonstrar essa afirmação, proceda como segue.

i) Faça o gráfico de cada uma das retas representadas nas equações acima e verifique que o ponto de interseção entre elas,  $(3/2, 3/2)$ , é a solução do sistema linear.

ii) Isole  $x_1$  na primeira equação e  $x_2$  na segunda equação e escreva o método de Jacobi.

iii) Considere  $x^{(0)} = (0, 0)$ . Mostre que  $x^{(2k)} = ([x_1]^{(2k)}, [x_2]^{(2k)})$  é tal que  $[x_2]^{(2k)} = [x_1]^{(2k)}$ , ou seja, os pontos da sequência que possuem índices pares estão sobre a reta  $y = x$ .

iv) Sabendo que  $x^{(1)} = (3, 1)$ , mostre que os pontos de índices ímpares estão sobre a reta  $y = 2 - (1/3)x$ .

Sugestão: Calcule o coeficiente angular da reta que passa pelos pontos  $x^{(2k-1)}$  e  $x^{(2k-3)}$  e obtenha o seguinte valor:

$$\{[x_2]^{(2k-1)} - [x_2]^{(2k-3)}\} / \{[x_1]^{(2k-1)} - [x_1]^{(2k-3)}\} = -1/3.$$

Lembre-se de que  $[x_1]^{(2k-4)} = [x_2]^{(2k-4)}$ . Além disso, observe que

$$[x_1]^{(2k-1)} = 3 - [x_2]^{(2k-2)} = 2 - (1/3)[x_1]^{(2k-3)}.$$

Portanto,  $[x_1]^{(2k-1)} - [x_1]^{(2k-3)} = 2 - (4/3)[x_1]^{(2k-3)}$ . Contas análogas são feitas para  $[x_2]^{(2k-1)}$ .

v) Faça iterações até que se obtenha erro relativo menor do que  $0.5 \times 10^{-1}$ .

vi) Coloque os pontos sobre as respectivas retas e perceba a convergência para  $(3/2, 3/2)$ .

Observações: (I) a sequência de índices pares terá termo geral  $x_n = 2 - (1/3)x_{n-1}$ , com  $x_0 = 0$ . Pode-se mostrar que  $x_n = 2 \cdot \{1 - 1/3 + 1/3^2 - 1/3^3 \dots\}$ , que é uma série alternada, portanto, a sequência com termo geral  $x_n$  é convergente. (II) A sequência de índices ímpares é do tipo  $z_n = 2 - (1/3)z_{n-1}$ , com  $z_0 = 1$ . Pode-se mostrar que a sequência de termo geral  $z_n - x_n$  é convergente e o seu limite vale zero. Logo a sequência com termo geral  $z_n = x_n + (z_n - x_n)$  é convergente e possui o mesmo limite da sequência com termo geral  $x_n$ , que é igual a  $3/2$ .

7) Um procedimento para melhorar soluções de sistemas lineares acometidas por erros de arredondamento é o chamado refinamento de solução, explicado a seguir. Seja

$$Ax = b \quad (1)$$

um sistema linear com  $n$  equações e  $n$  incógnitas que tem como solução o vetor  $x^*$ . Resolvendo o sistema (1), obtém-se o vetor  $x^{(1)}$  que, devido a erros de arredondamento, é tal que  $x^* - x^{(1)} = e^{(1)}$ , onde  $e^{(1)}$  é o vetor erro. Assim,  $x^* = x^{(1)} + e^{(1)}$ . Observe que

$$b = Ax^* = Ax^{(1)} + Ae^{(1)}. \quad (2)$$

Logo,  $b - Ax^{(1)} = Ae^{(1)}$ . O vetor  $r^{(1)} = b - Ax^{(1)}$  é chamado de resíduo. Com a finalidade de obter  $e^{(1)}$ , resolve-se o seguinte sistema linear (com a mesma matriz do sistema dado em (1)!):

$$Ae^{(1)} = r^{(1)}. \quad (3)$$

A solução encontrada (com erro  $e^{(2)}$ ) é denotada por  $\bar{e}^{(1)} = e^{(1)} - e^{(2)}$ . Assim,  $x^* = x^{(1)} + e^{(1)} = x^{(1)} + \bar{e}^{(1)} + e^{(2)}$ . Dessa forma, é encontrada uma nova aproximação para  $x^*$  (com erro  $e^{(2)}$ ):

$$x^{(2)} = x^{(1)} + \bar{e}^{(1)}. \quad (4)$$

Os passos anteriores (2 a 4) podem ser repetidos até que o resíduo  $r^{(k)}$  (depois de  $k$  refinamentos) esteja próximo do vetor nulo, ou seja,  $\|r^{(k)}\| < \varepsilon$  (a norma do resíduo é inferior a uma tolerância dada).

i) Considere o sistema linear  $\begin{pmatrix} 16.0 & 5.0 \\ 3.0 & 2.5 \end{pmatrix} x = \begin{pmatrix} 21.0 \\ 5.5 \end{pmatrix}$ . Trabalhando com arredondamento para dois dígitos em todas as operações, o método de Eliminação de Gauss produziu a solução  $x = \begin{pmatrix} 1.0 \\ 0.94 \end{pmatrix}$ . A matriz escalonada é  $\begin{pmatrix} 16.0 & 5.0 \\ 0 & 1.6 \end{pmatrix}$ , onde foi utilizado o multiplicador  $m_{21} = 0.19$ . Resolva o sistema (3) de maneira eficiente (como se fosse resolver um sistema de grande porte:  $n$  muito grande, com mínimo esforço computacional). Faça apenas um refinamento da solução e obtenha  $x^{(2)}$ . Calcule o resíduo  $r^{(2)}$ .

ii) Os multiplicadores e a matriz escalonada podem ser armazenados na própria matriz  $A$ ? Justifique.

8) Dado o sistema  $Ax = B$ , com  $x^t = (x_1, x_2, x_3)$  e  $B^t = (3, 1.1, 0)$ , seja  $A^{(1)} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0.1 & 1 \\ 0 & -1 & 1 \end{pmatrix}$  a matriz obtida após o pivoteamento parcial que trocou a primeira linha com a terceira linha da matriz  $A$ . Os multiplicadores do primeiro passo são  $m_{21} = -1/4$  e  $m_{31} = 3/4$ .

i) Escalone a matriz  $A^{(1)}$  usando a técnica de pivoteamento parcial.

ii) Exiba o vetor dos termos independentes que é utilizado na decomposição LU.

iii) Obtenha a matriz  $A$ : use as matrizes  $L$  (triangular inferior) e  $U$  (triangular superior) resultantes dos processos de pivoteamento e escalonamento efetuados acima.

9) Dado o sistema  $Ax = B$ , com  $x^t = (x_1, x_2, x_3)$  e  $B^t = (7, 1, -4)$ , seja  $A^{(1)} = \begin{pmatrix} 4 & 0 & -3 \\ 0 & 2 & 11/4 \\ 0 & -4 & 13/4 \end{pmatrix}$  a matriz obtida após o pivoteamento parcial que trocou a primeira linha com a terceira linha da matriz  $A$ . Os multiplicadores do primeiro passo são  $m_{21} = -1/4$  e  $m_{31} = 3/4$ .

i) Escalone a matriz  $A^{(1)}$  usando a técnica de pivoteamento parcial.

ii) Exiba o vetor dos termos independentes que é utilizado na decomposição LU.

iii) Obtenha a matriz  $A$ : use as matrizes  $L$  (triangular inferior) e  $U$  (triangular superior) resultantes dos processos de pivoteamento e escalonamento efetuados acima.

10) Na decomposição  $LU$ , associada ao sistema linear  $Ax = B$ , são resolvidos dois sistemas lineares: primeiro um sistema triangular inferior  $Ly = B$  e, em seguida, um triangular superior  $Ux = y$ . O algoritmo associado ao sistema superior é dado abaixo; considerou-se uma matriz de ordem  $N$ .

```
x(N) = y(N)/a(N,N); (Obs.: x(N) é a N-ésima coordenada do vetor x)
for i = 1 : N-1
    I = N-i;
    (Obs.: as duas linhas seguintes correspondem ao somatório  $\sum_{j=I+1}^N x(j) \cdot a(I)(j)$ )
    soma = x(I+1)*a(I, I+1);
    for j = (I+2) : N
        soma = soma + x(j)*a(I,j);
    end
    x(I) = (y(I) - soma)/a(I,I);
end
```

Lembrando-se de que o sistema triangular inferior é resolvido de cima para baixo e que a matriz  $L$  possui todos os elementos da diagonal iguais a 1, escreva o algoritmo associado ao sistema triangular inferior.

11) Quer-se determinar uma matriz  $B$  tal que  $A.B = C$ , em que

$$A = \begin{pmatrix} 3 & -3 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & 2 \end{pmatrix} \quad \text{e} \quad C = \begin{pmatrix} 4 & -1 & 9 \\ 1 & -5 & 3 \\ 0 & 1 & 8 \end{pmatrix}.$$

Exiba as matrizes  $L$  e  $U$  e o termo independente utilizados na resolução do sistema linear que irá gerar a terceira coluna da matriz  $B$ . Não precisa resolver o sistema linear.

12) O método implícito de Crank-Nicolson aplicado à equação do calor  $u_t = u_{xx}$ , onde  $t > 0$  e  $0 < x < 1$ , é dado por:

$$-r.u_{i-1,j+1} + (2 + 2r).u_{i,j+1} - r.u_{i+1,j+1} = b_j \equiv r.u_{i-1,j} + (2 - 2r).u_{i,j} + r.u_{i+1,j}, \quad (C-N)$$

onde  $r$  é uma constante dada por  $r = \Delta t / (\Delta x)^2$ ,  $\Delta t = t_{j+1} - t_j$  e  $\Delta x = x_i - x_{i-1}$ .

Assim, conhecidas as aproximação no tempo  $t_j$ ,  $u_{i,j} \approx u(x_i, t_j)$ , as aproximações no tempo  $t_{j+1}$  serão obtidas resolvendo-se o sistema linear com equações do tipo (C-N) anterior. Explique por que o método de decomposição  $LU$  é mais adequado para a resolução desse sistema linear do que o método de Eliminação de Gauss. Observe que, em cada passo de tempo, o sistema será do tipo  $A.u(t) = b(t)$ . Apenas as incógnitas ( $u$ ) e o termo independente ( $b$ ) variam com o tempo.

## Atividade 2 – Método de Jacobi e de Gauss - Seidel

**Prezado(a) aluno(a),**

Você poderá encontrar vários problemas práticos relacionados a sistemas lineares na página da Faculdade de Matemática da UFU: [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) (Laboratório - Unidade 2).

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre os Métodos de Jacobi e de Gauss - Seidel: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 2), ambos localizados em [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Resolva o seguinte exercício:

Sejam  $A = \begin{pmatrix} 5 & 1 & 1 \\ 3 & -4 & 1 \\ 1 & 1 & -2 \end{pmatrix}$ ;  $x^t = (x_1 \ x_2 \ x_3)$  e  $b^t = (7 \ 0 \ 0)$ .

(a) Mostre que o Critério de Sassenfeld é satisfeito.

(b) Faça duas iterações com o método de Gauss – Seidel partindo do vetor inicial  $x^{(0)} = (0.95 \ 0.95 \ 0.95)^t$ .

(c) Calcule o erro relativo:  $\frac{\|x^{(2)} - x^{(1)}\|_{\max}}{\|x^{(2)}\|_{\max}}$ . Lembre-se de que  $\|x\|_{\max} = \max\{|x_1|, |x_2|, |x_3|\}$ .

(d) Monte a matriz  $C$  do método de Gauss – Jacobi e calcule os seus autovalores. Conclua que o Método de Jacobi também será convergente. Observação: os autovalores de  $C$  são as raízes do polinômio (característico) dado por  $p(\lambda) = \det(C - \lambda I)$ ;  $I$  é a matriz identidade. O polinômio característico de  $C$  é  $p(\lambda) = -\lambda^3 - (\lambda/8) - 1/10$ . (Veja o exercício 8 da Lista 1).

### Informações sobre a Atividade 2

(1) Critério de Sassenfeld.

$$\beta_1 = 1/5 + 1/5 = 2/5; \quad \beta_2 = (3/4). \beta_1 + 1/4 = 11/20; \quad \beta_3 = (1/2). \beta_1 + (1/2). \beta_2 = 19/40.$$

(2) Iterações com Gauss – Seidel.

$$x^{(1)} = (1.02 \ 1.0025 \ 1.01125)^t; \quad x^{(2)} = (0.99725 \ 1.00075 \ 0.999)^t.$$

(3) Erro Relativo.

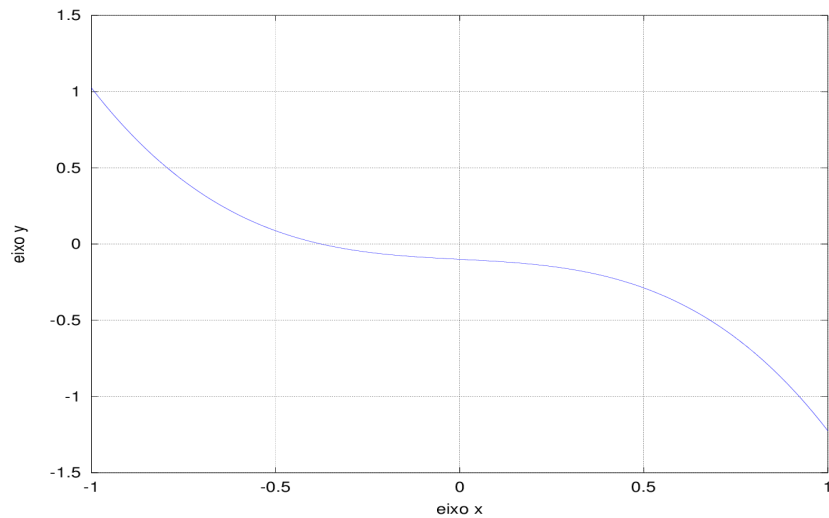
$$ER^{(2)} = \frac{\|x^{(2)} - x^{(1)}\|_{\max}}{\|x^{(2)}\|_{\max}} = 0.02275/1.00075 = 0.02273295 \approx 0.023.$$

(4) Matriz de Iteração do Método de Jacobi:  $C = \begin{pmatrix} 0 & -1/5 & -1/5 \\ 3/4 & 0 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$ .

(5) Obtenção do autovalor real de  $C$ , que é a raiz real de  $p(\lambda) = -\lambda^3 - (\lambda/8) - 1/10$ .

Através de um gráfico (veja abaixo) pode-se encontrar o intervalo  $[-1/2 \ 0]$  no qual  $p$  muda de sinal.

Gráfico (item 5)



Utilizando o Método de Newton – Raphson com valor inicial  $-0.25$  obtemos a seguinte aproximação para a raiz:  $\lambda_1 = -0.375713238$ .

(6) Briot – Ruffini.

$$-(\lambda^3 + (\lambda/8) + 1/10) \div (\lambda + 0.375713238) = -(\lambda^2 - 0.375713238\lambda + 0.266160437).$$

Assim, as outras duas raízes de  $p$  são dadas por  $\lambda_2 = 0.187856619 + 0.480489674 i$  e  $\lambda_3 = 0.187856619 - 0.480489674 i$ .

(7) Conclusão.

O módulo do autovalor real é dado por  $|\lambda_1| = 0.375713238$ ; os módulos dos autovalores complexos são dados por  $|\lambda_2| = |\lambda_3| \approx 0.51591$ . Portanto, os módulos de todos os autovalores de  $C$  são menores do que 1, o que garante a convergência do método de Jacobi.

### Atividade 3 – Eliminação de Gauss e Decomposição LU

**Prezado(a) aluno(a),**

Você já deve ter visto na disciplina de Álgebra Linear a definição de matriz inversa ( $A \cdot A^{(-1)} = I$ ). Em Cálculo Numérico, o método da Decomposição LU, obtido do Método da Eliminação de Gauss, fornecerá a inversa de qualquer matriz  $A$  quadrada de ordem  $n$ , que tenha determinante diferente de zero. O cálculo do determinante da matriz  $A$  também é fornecido pelo método da Decomposição LU. Sabendo-se que  $L$  é uma matriz triangular inferior, com elementos diagonais iguais a 1, e  $U$  é uma matriz triangular



superior, então  $\det(A) = \det(U)$ , se  $A = LU$  ou  $\det(A) = \det(U) \cdot (-1)^r$ , se  $PA = LU$ , onde  $r$  é o número de trocas de linhas que foram executadas ao longo do processo de escalonamento. Note que  $\det(U) = u_{11} \cdot u_{22} \dots u_{nn}$ .

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre o método da Eliminação de Gauss e Decomposição LU: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 2), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Obtenha a inversa da matriz  $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}$ . Utilize o método de

Decomposição LU obtido do Método de Eliminação de Gauss com Pivoteamento Parcial. Nesse caso você terá que resolver três sistemas lineares do tipo  $LU = Pb$ , onde  $L$  é a matriz dos multiplicadores (que são permutados de acordo com as trocas de linhas efetuadas);  $U$  é a matriz triangular superior obtida após o escalonamento da matriz  $A$  e  $Pb$  é o vetor  $b$  (colunas da matriz Identidade) permutado de acordo com as trocas de linhas efetuadas no Pivoteamento Parcial.

### Informações sobre a Atividade 3

(1) Pivoteamento e Escalonamento.

Partindo-se de  $A = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}$ , obtemos no primeiro estágio:

(pivoteamento)  $\begin{pmatrix} 4 & 0 & -3 \\ 1 & 2 & 2 \\ 3 & -4 & 1 \end{pmatrix}$ ; (escalonamento)  $A^{(1)} = \begin{pmatrix} 4 & 0 & -3 \\ (1/4) & 2 & 11/4 \\ (3/4) & -4 & 13/4 \end{pmatrix}$  e no

segundo estágio obtemos:

(pivoteam.)  $\begin{pmatrix} 4 & 0 & -3 \\ (3/4) & -4 & 13/4 \\ (1/4) & 2 & 11/4 \end{pmatrix}$ ; (escalonamento)  $A^{(2)} = \begin{pmatrix} 4 & 0 & -3 \\ (3/4) & -4 & 13/4 \\ (1/4) & (-1/2) & 11/4 \end{pmatrix}$ .

(2) Matriz Triangular Inferior.

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix}.$$

(3) Matriz Triangular Superior:  $U = \begin{pmatrix} 4 & 0 & -3 \\ 0 & -4 & 13/4 \\ 0 & 0 & 11/4 \end{pmatrix}.$

(4) Termos Independentes.

Foram executadas duas permutações de linhas: trocou-se a linha 1 com a linha 3 e depois trocou-se a linha 2 com a linha 3. Dessa forma, os termos independentes são:

$$b^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ (sistema 1); } \quad b^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ (sistema 2)} \quad \text{e} \quad b^{(3)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ (sistema 3).}$$

(5) Resolução dos sistemas lineares.

Após a resolução dos sistemas lineares (utilizado a decomposição LU) obtemos:

$$A^{-1} = \begin{pmatrix} 3/35 & 6/35 & 1/7 \\ -11/70 & 13/70 & 1/14 \\ 4/35 & 8/35 & -1/7 \end{pmatrix}.$$

## 7. Atividades suplementares



### Atividade 1

O código exibido a seguir foi desenvolvido na linguagem de programação do OCTAVE. A sua tarefa será testar o código utilizando o mesmo *software* (OCTAVE) ou desenvolver as suas próprias rotinas utilizando a linguagem de programação que lhe seja mais conveniente.

### Algoritmo do método de Crank – Nicolson; Problema proposto no **Exemplo 13 da Seção 5**

```
% Algoritmo para escalonar uma matriz tridiagonal de ordem N - referente ao problema
% de conducao de calor em uma barra unidimensional linear homogênea de
% comprimento L
```

% resolucao do sistema linear  $Au = B$  utilizando decomposiçao LU; discretizaçao da  
% equacao:  $u_t = Ku_{xx} + f$ .

% A discretizacao foi baseada em Diferencas finitas centrais:

%  $u_{xx} \sim (u_{i-1} - 2u_i + u_{i+1})/H^2$

% no tempo foi utilizado o metodo de Crank-Nicolson

% t pertencente a  $J = (t_0, T)$

%Entrada de dados

$t_0 = 0.0$ ;

$T = 0.5$ ;

$nt = 50$ ; %subdivisões do intervalo J

% Numero de subintervalos de  $I = [x_0, x_f]$ ;  $x_f - x_0 = L$

$x_0 = 0$ ;

$x_f = 1$ ;

$ns = 10$ ;

% Termo de fonte de calor

%chamar a funcao  $f_{\text{fon}}(x,t)$ ;

%Condiccao de fronteira no extremo inferior da barra

% usar a funcao  $g_e(x_0,t) = g_{e\_fr}(t)$ ;

%Condiccao de fronteira no extremo superior da barra

% usar a funcao  $g_d(x_f, t) = g_{d\_fr}(t)$ ;

%condicao inicial:  $u(x, t_0) = q(x)$

% deve ser chamada a funcao  $q_{\text{inic}}(x)$ ;

%Difusibilidade térmica

$KK = 1.00$ ;

%Comprimento da barra

$L = 1.0$ ;

%Numero de incognitas

```

N = ns-1;

%Comprimento dos subintervalos
H = L/ns;

Ht = (T - t0)/nt;

%pontos da malha - eixo x

for i=1:ns+1
xx(i) = x0 + (i-1)*H;
end

% pontos da malha - eixo t
for j= 1:nt+1
tt(j) = t0 + (j-1)*Ht;
end

%parametros do metodo de C-N

ro = (KK*Ht)/(H*H);
ro_p1 = ro + 1.0;
one_ro = 1.0 - ro;

%Construcao dos elementos da diagonal principal e diagonais superior e inferior;
%construcao do vetor termo independente.

for i = 1 : N
    diag(i) = 2.0*ro_p1;
    ds(i) = -ro;
    di(i) = -ro;
end

%Decomposicao LU

for i = 1 : N-1
    mult = di(i)/diag(i);
    di(i) = mult;
    diag(i+1) = diag(i+1) - mult*ds(i);
end

```

```

%calculo do vetor inicial uj0 = u(x,t0)
for i = 1:N
    uj0(i) = q_inic(xx(i+1));
    uj(i) = 0;
end

%calculo do vetor da fronteira esquerda u_fe = ge_fr(t)
%calculo do vetor da fronteira direita u_fd = gd_fr(t)

for j = 1:nt+1
    u_fe(j) = ge_fr(tt(j));
    u_fd(j) = gd_fr(tt(j));
end

%resolucao dos sistemas lineares para cada tempo tj

for j = 1:nt
    %calculo do termo independente

    for i = 2:N-1

        B(i) = ro*uj0(i-1) + 2*one_ro*uj0(i) + ro*uj0(i+1) +
            Ht*(f_fon(xx(i+1),tt(j)) + f_fon(xx(i+1),tt(j+1)));
    end

    B(1) = ro*(u_fe(j+1) + u_fe(j)) + 2*one_ro*uj0(1) + ro*uj0(2) +
        Ht*(f_fon(xx(2),tt(j)) + f_fon(xx(2),tt(j+1)));

    B(N) = ro*(u_fd(j+1) + u_fd(j)) + ro*uj0(N-1) + 2*one_ro*uj0(N) +
        Ht*(f_fon(xx(N+1),tt(j)) + f_fon(xx(N+1),tt(j+1)));

    %Retro-solucao - Sistema Triangular Inferior

    y(1) = B(1);
    for i = 2:N
        y(i) = B(i) - di(i-1)*y(i-1);
    end
    %FIM

```

```
%Retro-solucao - Sistema Triangular Superior
```

```
uj(N) = y(N)/diag(N);  
for k = 1:N-1  
    uj(N-k) = (y(N-k) - (ds(N-k)*uj(N-k+1)))/diag(N-k);  
end
```

```
%atualizacao do vetor uj0
```

```
for i = 1:N  
    uj0(i) = uj(i);  
end
```

```
end
```

```
%calcula de uma solucao exata
```

```
for i = 1:N  
    u_exata(i) = exp(-pi*pi*T)*sin(pi*xx(i+1));  
end
```

```
u_exata    %imprime na tela do computador o vetor u_exata
```

```
uj          %imprime na tela do computador a solucao do sistema: vetor uj
```

```
% FIM do código
```

**Observações:** As funções utilizadas na rotina anterior foram definidas em quatro arquivos com os conteúdos descritos a seguir.

(1) arquivo q\_inic.m (condição inicial):

```
function g = q_inic(x)  
    g = sin(pi*x);  
end
```

(2) arquivo ge\_fr.m (condição de fronteira esquerda):

```
function g = ge_fr(t)  
    g = 0;  
end
```

(3) arquivo gd\_fr.m (condição de fronteira direita):

```
function g = gd_fr(t)
    g = 0;
end
```

(4) arquivo f\_fon.m (função relacionada ao termo de fonte):

```
function g = f_fon(x,t)
    g = 0;
end
```

### Atividade 2

Modifique o algoritmo anterior (Atividade 1) de modo a resolver o mesmo problema proposto no **Exemplo 13** utilizando o Método de Gauss – Seidel.

### Atividade 3

Modifique o algoritmo da Atividade 1 anterior de modo a resolver o seguinte problema:

$$u_t(x, t) = Ku_{xx}(x, t) + f(x, t),$$

$$x \in I = (x_0, x_f) = (0, 1); \quad x_f - x_0 = L = 1; \quad t \in J = (t_0, T) = (0, 1); \quad K = 1 \quad \text{e} \quad f(x, t) = 2,$$

com condições de fronteira:  $u(x_0, t) = g_1(t) = 0$ ;  $u(x_f, t) = g_2(t) = 0$ ,  $\forall t \in J$ ; e com condição inicial dada por  $u(x, t_0) = q(x) = \sin(\pi x) + x(1 - x)$ ,  $x \in [x_0, x_f]$ .

**Observação:** A solução analítica desse problema de valor de contorno é dada por  $u(x, t) = \exp(-\pi^2 t) \sin(\pi x) + x(1 - x)$ .

## Módulo 3

# Ajuste de curvas, Interpolação Polinomial e Integração Numérica

### 1. Introdução

**Prezado estudante, seja bem vindo.**

Neste módulo você irá estudar o Método dos Quadrados Mínimos, que é muito utilizado para aproximar uma função  $f(x)$ ,  $x \in I = [a, b]$ , ou para aproximar (ou ajustar) os dados de uma tabela com  $m$  pontos  $(x_k, y_k)$ ,  $1 \leq k \leq m$ , por meio de uma função pré-estabelecida  $\Phi(x)$ . No caso discreto (pontos dados por meio de uma tabela), se a função de ajuste for do tipo  $\Phi(x) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ , onde  $v = (v_1, v_2, \dots, v_n)$  é um vetor de parâmetros, então o Método dos Quadrados Mínimos consiste em resolver um sistema linear  $A\alpha = b$ , cuja solução é o vetor de parâmetros  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  que minimiza a função  $E = E(v) = \sum_{k=1}^m (\Phi(x_k) - y_k)^2$ . Os elementos da matriz  $A$  são

dados por  $a_{ij} = a_{ji} = \sum_{k=1}^m g_i(x_k) g_j(x_k)$  e as coordenadas do vetor  $b$  são dadas por  $b_i = \sum_{k=1}^m y_k g_i(x_k)$ . Dessa forma, os métodos desenvolvidos para resolver sistemas lineares serão bastante úteis neste tópico.

Também, neste módulo, você utilizará um polinômio, denominado polinômio interpolador, para aproximar uma função e, na sequência, você aprenderá a aproximar a integral de uma função a partir da integral de seu polinômio interpolador.

Dada uma função real  $f: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  e dados  $n + 1$  pontos,  $x_0, x_1, \dots, x_n$ , no intervalo fechado  $I$ , dizemos que o polinômio de grau igual a  $n$  ou de grau menor do que  $n$ ,  $p_n(x) = a_0 + a_1 x + \dots + a_n x^n$ , é o polinômio interpolador associado à função  $f$  se  $p_n(x_i) = f(x_i)$ , para todo  $i$ ,  $0 \leq i \leq n$ . Observe que as  $n + 1$  equações anteriores dão origem a um sistema linear de ordem  $n + 1$ . A matriz desse sistema linear é conhecida como matriz de Vandermonde e possui determinante dado por  $\det(V) =$



$\prod_{j=0}^{n-1} \prod_{i=j+1}^n (x_i - x_j) = \prod_{0 \leq j < i \leq n} (x_i - x_j)$ . Por exemplo, se  $n = 3$  então  $\det(V) = (x_1 - x_0) \cdot (x_2 - x_0) \cdot (x_3 - x_0) \cdot (x_2 - x_1) \cdot (x_3 - x_1) \cdot (x_3 - x_2)$ . Dessa forma, se os pontos forem distintos dois a dois, isto é,  $x_i \neq x_j$  para  $i \neq j$ , então  $\det(V) \neq 0$  e, portanto, existirá um único polinômio interpolador (pois o sistema linear possuirá uma única solução, que corresponde aos coeficientes do polinômio interpolador).

A resolução do sistema linear mencionado anteriormente não é a maneira mais adequada para se obter o polinômio interpolador porque a matriz de Vandermonde é mal-condicionada. Isto significa que pequenas perturbações (imprecisões causadas por erros de arredondamento) nos elementos da matriz podem gerar grandes perturbações na solução do sistema linear. Por essa razão, você aprenderá duas técnicas diferentes para se obter o polinômio interpolador, quando os pontos de interpolação forem distintos dois a dois. Uma das técnicas é devida a Lagrange (polinômio de Lagrange) e a outra é devida a Newton (polinômio de Newton com diferenças divididas).

No caso de três pontos distintos de interpolação,  $x_0, x_1$  e  $x_2$ , a função  $f$ , com valores  $f(x_0)$ ,  $f(x_1)$  e  $f(x_2)$ , terá o seguinte polinômio interpolador de Lagrange:

$$p_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x),$$

onde

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} ; \quad L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} ;$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} .$$

O polinômio de Newton será dado por:  $p_2(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1)$ , onde

$$d_0 = f(x_0); \quad d_1 = f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} ;$$

$$d_2 = f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} .$$

A integral do polinômio interpolador dará origem a uma regra de integração. Nesse curso, estudaremos apenas duas regras de integração: a Regra do Trapézio, correspondente à integral de um polinômio interpolador de grau 1, e a Regra de Simpson (1/3 de Simpson), correspondente à integral de um polinômio interpolador de grau 2.

Ao aproximar uma função por um polinômio será cometido um erro ( $E$ ), chamado de erro de interpolação. Digamos que  $f(x) = p_n(x) + E^{(n)}(x)$ . Então a integral de  $f$  pode ser

escrita como:  $\int_a^b f(x) dx = \int_a^b p_n(x) dx + \int_a^b E^{(n)}(x) dx$ . Nesse caso dizemos que  $E^{(n)}(x)$  é o erro de interpolação e  $\int_a^b E^{(n)}(x) dx$  é o erro de integração.

**Observação.** Não deixe de rever os tópicos relacionados a polinômios (leia, por exemplo, o material visto na disciplina Fundamentos da Matemática elementar I).

Considerando-se uma partição de pontos igualmente espaçados do intervalo  $[a, b]$ ,  $x_0, x_1, \dots, x_m$ , onde  $x_i = x_0 + ih$ ,  $h = (b - a)/m$ , as duas regras de integração, com os respectivos erros de integração, são dadas por

Regra Repetida do Trapézio:

$$I_{TR} = (h/2) \{f(x_0) + 2 \sum_{i=1}^{m-1} f(x_i) + f(x_m)\};$$

$$E_{TR} = -\frac{m h^3 f''(c)}{12}, c \in [a, b];$$

$m$  pode ser qualquer número inteiro positivo.

Regra Repetida de Simpson:

$$I_{SR} = (h/3) \{f(x_0) + 4 \sum_{i=1}^{\frac{m}{2}} f(x_{2i-1}) + 2 \sum_{i=1}^{\frac{m}{2}-1} f(x_{2i}) + f(x_m)\},$$

$$E_{SR} = -\frac{m h^5 f^{iv}(C)}{180}, C \in [a, b],$$

$m$  deve ser um número par.

## 2. Objetivos e conteúdo do Módulo 3

São dois os objetivos deste módulo: (i) aproximar uma função real utilizando ajuste de curvas ou interpolação polinomial. Em ajuste de curvas, estudaremos o Método dos Quadrados Mínimos, que é utilizado para aproximar funções ou para ajustar os pontos de uma tabela através de uma função pré-estabelecida. Em interpolação polinomial, estudaremos os polinômios de Lagrange e de Newton. (ii) aproximar a integral de uma função real pela integral do polinômio interpolador dessa função. Para você ficar familiarizado com os métodos numéricos apresentados nesse módulo, não deixe de resolver os exercícios propostos e não deixe de ler as referências bibliográficas indicadas neste texto e outras que você achar conveniente.

### Conteúdos básicos do Módulo 3

- Método dos Quadrados Mínimos: Caso Discreto do Modelo Linear; Redução do Modelo Não-Linear ao Modelo Linear; Caso Contínuo do Modelo Linear.
- Existência e unicidade do polinômio interpolador.
- Polinômio de Lagrange.
- Polinômio de Newton com diferenças divididas.
- Regra dos Trapézios.
- Regra de Simpson.

## 3. Ajuste de Curvas e o Método dos Quadrados Mínimos

### 3.1 Caso Discreto – Modelo Linear

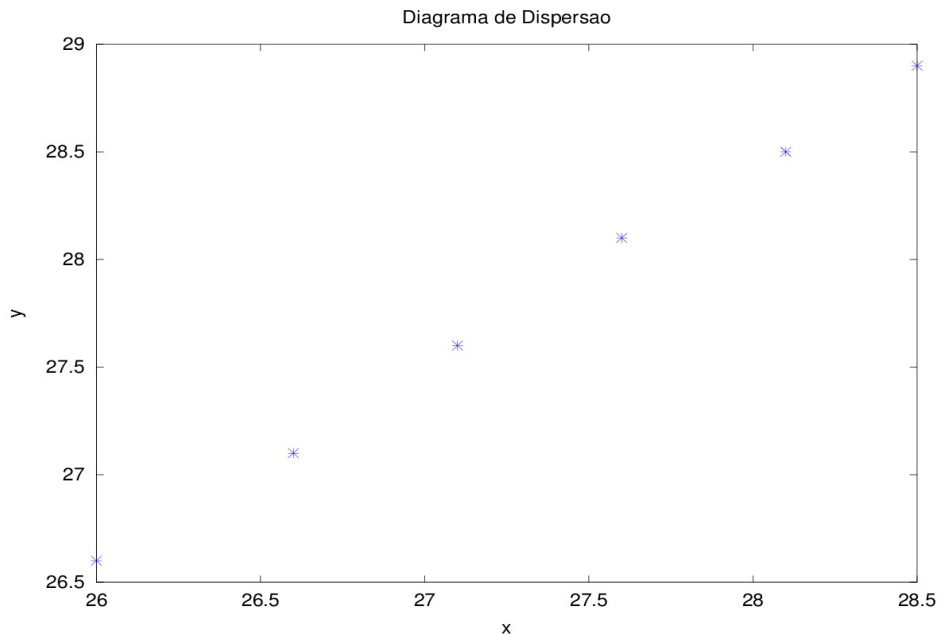
O caso discreto do Modelo Linear de Ajuste de Curvas consiste em ajustar os pontos de uma tabela com  $m$  pontos  $(x_k, y_k)$ ,  $1 \leq k \leq m$ , utilizando uma função  $\Phi(x; v_1, v_2, \dots, v_n) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ . Note que a função  $\Phi$  é uma transformação linear com respeito ao vetor de parâmetros  $v = (v_1, v_2, \dots, v_n)$ , isto é, dado um outro vetor qualquer de parâmetros,  $w = (w_1, w_2, \dots, w_n)$ , e qualquer escalar real  $\mu$ , tem-se que  $\Phi(v + \mu w) = \Phi(v) + \mu \Phi(w)$ .

As funções  $g_1, \dots, g_n$  que aparecem na função  $\Phi$  são obtidas a partir da análise do diagrama de dispersão da tabela dada. Isto é, os pontos  $(x_k, y_k)$  são colocados sobre o plano cartesiano, o que dará origem a uma curva (ou uma reta). É a partir da análise do tipo de curva proveniente do diagrama de dispersão que as funções anteriores são definidas. Veja o próximo exemplo.

**Exemplo 1. (Diagrama de Dispersão)** Considere a tabela:

$x_k$	$x_1 = 26$	$x_2 = 26.6$	$x_3 = 27.1$	$x_4 = 27.6$	$x_5 = 28.1$	$x_6 = 28.5$
$y_k$	$y_1 = 26.6$	$y_2 = 27.1$	$y_3 = 27.6$	$y_4 = 28.1$	$y_5 = 28.5$	$y_6 = 28.9$

Existe, claramente, um comportamento linear dos pontos exibidos no diagrama de dispersão desta tabela. Tal análise nos leva a acreditar que a função de ajuste seja uma reta, ou seja,  $\Phi(x) = v_1 + v_2 x$ . Portanto,  $g_1(x) = 1$  e  $g_2(x) = x$ . ■



**Figura 1 – Módulo 3. Diagrama de Dispersão – Tabela do Exemplo 1**

## O Método dos Quadrados Mínimos

O Método dos Quadrados Mínimos para o caso discreto do Modelo Linear de Ajuste de Curvas consiste em determinar o vetor ótimo de parâmetros  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  que irá minimizar a função erro:  $E(v) = \sum_{k=1}^m (\Phi(x_k) - y_k)^2$ , onde  $\Phi(x; v_1, v_2, \dots, v_n) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ . Note que  $E$  é uma função de  $n$  variáveis reais, ou seja,  $E: \mathbb{R}^n \rightarrow \mathbb{R}$ .

Mostraremos que o vetor ótimo de parâmetros,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ , é aquele que anula o gradiente da função  $E$ :  $\nabla E(\alpha) = \left( \frac{\partial E}{\partial \alpha_1}, \frac{\partial E}{\partial \alpha_2}, \dots, \frac{\partial E}{\partial \alpha_n} \right) = (0, 0, \dots, 0)$ . Observe que  $\frac{\partial E(v)}{\partial v_i} = \sum_{k=1}^m \frac{\partial ([\Phi(x_k) - y_k]^2)}{\partial v_i}$ , para cada  $i$  variando de 1 até  $n$ .

**Observação:** Lembre-se de que, ao fazer a derivada parcial em relação à variável  $v_i$ , todas as outras coordenadas do vetor  $v = (v_1, v_2, \dots, v_n)$  permanecem constantes.

Utilizando a definição da função  $\Phi$  e a Regra da Cadeia para efetuar a derivação anterior, obtemos que

$$\begin{aligned} \frac{\partial E(v)}{\partial v_i} &= \sum_{k=1}^m \frac{\partial ([v_1 g_1(x_k) + \dots + v_i g_i(x_k) + \dots + v_n g_n(x_k) - y_k]^2)}{\partial v_i} \rightarrow \\ &\rightarrow \frac{\partial E(v)}{\partial v_i} = \sum_{k=1}^m \{ 2[(\sum_{j=1}^n v_j g_j(x_k)) - y_k] g_i(x_k) \} . \end{aligned}$$

Portanto,

$$\frac{\partial E(\alpha)}{\partial \alpha_i} = 0 \leftrightarrow \sum_{k=1}^m \sum_{j=1}^n \alpha_j g_j(x_k) g_i(x_k) = \sum_{k=1}^m y_k g_i(x_k).$$

A igualdade anterior pode ser reescrita como segue:

$$\sum_{j=1}^n \alpha_j \sum_{k=1}^m g_j(x_k) g_i(x_k) = \sum_{k=1}^m y_k g_i(x_k) , \quad 1 \leq i \leq n.$$

Se você ainda não percebeu que isso é um sistema linear nas incógnitas  $\alpha_1, \alpha_2, \dots, \alpha_n$ , então considere o caso particular  $n = 2$  e note que as equações anteriores são dadas por:

$$\begin{aligned} \alpha_1 \sum_{k=1}^m g_1(x_k) g_1(x_k) + \alpha_2 \sum_{k=1}^m g_2(x_k) g_1(x_k) &= \sum_{k=1}^m y_k g_1(x_k) ; \\ \alpha_1 \sum_{k=1}^m g_1(x_k) g_2(x_k) + \alpha_2 \sum_{k=1}^m g_2(x_k) g_2(x_k) &= \sum_{k=1}^m y_k g_2(x_k) . \end{aligned}$$

Nesse caso, o sistema dois por dois possui matriz de coeficientes  $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ , onde

$$a_{11} = \sum_{k=1}^m g_1(x_k) g_1(x_k); \quad a_{12} = a_{21} = \sum_{k=1}^m g_2(x_k) g_1(x_k); \quad a_{22} = \sum_{k=1}^m g_2(x_k) g_2(x_k) \text{ e}$$

vetor de termos independentes  $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ , onde  $b_1 = \sum_{k=1}^m y_k g_1(x_k)$  e

$b_2 = \sum_{k=1}^m y_k g_2(x_k)$ . No caso geral, o sistema  $n \times n$  possui matriz  $A$  de ordem  $n$ ,

simétrica, tal que  $a_{ij} = a_{ji} = \sum_{k=1}^m g_i(x_k) g_j(x_k)$ ; e vetor de termos independentes  $b$ ,

com coordenadas  $b_i = \sum_{k=1}^m y_k g_i(x_k)$ ,  $1 \leq i \leq n$ .

**Exercício 1. (Matriz Hessiana da função  $E(v)$ )** Sabendo que  $\frac{\partial E(v)}{\partial v_i} = 2 \sum_{k=1}^m [v_1 g_1(x_k) + \dots + v_j g_j(x_k) + \dots + v_n g_n(x_k) - y_k] g_i(x_k)$  obtenha a matriz Hessiana da função  $E(v)$ , a qual é composta pelas derivadas de ordem 2  $\frac{\partial^2 E(v)}{\partial^2 v_j v_i}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq n$ .

**Solução do Exercício 1:**

Note que

$$\frac{\partial^2 E(v)}{\partial^2 v_j v_i} = 2 \sum_{k=1}^m \frac{\partial \{ [v_1 g_1(x_k) + \dots + v_j g_j(x_k) + \dots + v_n g_n(x_k) - y_k] g_i(x_k) \}}{\partial v_j}.$$

Assim,  $\frac{\partial^2 E(v)}{\partial^2 v_j v_i} = 2 \sum_{k=1}^m g_j(x_k) g_i(x_k) = 2 a_{ij} = \frac{\partial^2 E(v)}{\partial^2 v_i v_j}$ . Portanto, a Matriz Hessiana da função  $E(v)$  é  $H(E(v)) = 2A$ , onde  $A$  é a matriz do sistema linear (apresentado antes do **Exercício 1**) oriundo do Método dos Quadrados Mínimos. □

**Exercício 2. ( $A$  é uma matriz simétrica e definida positiva)** Sejam  $g_1(x)$ ,  $g_2(x)$ , ...,  $g_n(x)$  funções linearmente independentes, ou seja,  $v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x) = 0$ , para todo  $x$  real, se, e somente se,  $v_1 = v_2 = \dots = v_n = 0$ . Então a matriz  $A$  (e, portanto, a matriz  $H(E(v))$ ) é definida positiva, isto é,  $w^t A w > 0$  para todo  $w$ , vetor coluna não nulo com  $n$  coordenadas reais.

**Solução do Exercício 2:**

Seja  $w$  um vetor coluna não nulo com  $n$  coordenadas reais. Então, de acordo com a hipótese do exercício,  $S_k = w_1 g_1(x_k) + w_2 g_2(x_k) + \dots + w_n g_n(x_k) \neq 0$ . Portanto,  $(S_k)^2 > 0$ .

Assim,  $0 < (S_k)^2 = \left( \sum_{i=1}^n w_i g_i(x_k) \right) \left( \sum_{j=1}^n w_j g_j(x_k) \right) = \sum_{i=1}^n \sum_{j=1}^n w_i g_i(x_k) w_j g_j(x_k)$  (na

última igualdade, basta usar indução finita). Note que  $\sum_{k=1}^m (S_k)^2 > 0$ , logo:

$$\begin{aligned} 0 < \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^n w_i g_i(x_k) w_j g_j(x_k) &= \sum_{i=1}^n w_i \sum_{j=1}^n \left( \sum_{k=1}^m g_i(x_k) g_j(x_k) \right) w_j \\ &= \sum_{i=1}^n w_i \left( \sum_{j=1}^n a_{ij} w_j \right) = w^t A w. \end{aligned}$$

Portanto, a matriz  $A$  é definida positiva. □

**Observação:** O desenvolvimento de Taylor da função  $E: \mathbb{R}^n \rightarrow \mathbb{R}$  em torno do ponto  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  é dado por

$$E(w) = E(\alpha + (w - \alpha)) = E(\alpha + v) = E(\alpha) + \nabla E(\alpha) \cdot v + (1/2!)v^t \cdot H(E(\alpha + \theta v)) \cdot v,$$

onde  $\theta$  é um número real pertencente ao intervalo aberto  $(0, 1)$ .

Este resultado pode ser encontrado em livros de Análise no  $\mathbb{R}^n$ , por exemplo:



[1] LIMA, Elon Lages, *Curso de Análise*. Volume 2 (Projeto Euclides). Rio de Janeiro, Instituto de Matemática Pura e Aplicada, CNPq, 2000.

### Referências

**Observação:** Como a matriz Hessiana da função  $E$  é definida positiva, ou seja,  $v^t \cdot H(E(\alpha + \theta v)) \cdot v > 0$ , e  $\nabla E(\alpha) = 0$  então  $E(w) > E(\alpha)$ , qualquer que seja  $w \in \mathbb{R}^n$ . Dessa forma,  $E(\alpha)$  é o valor mínimo global da função  $E$ .



### Síntese

O Método dos Quadrados Mínimos para o caso discreto do Modelo Linear de Ajuste de Curvas consiste em determinar o vetor ótimo de parâmetros  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  que irá minimizar a função erro:  $E(v) = \sum_{k=1}^m (\Phi(x_k) - y_k)^2$ , onde  $\Phi(x; v_1, v_2, \dots, v_n) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ . Para isso, basta resolver o sistema linear  $A\alpha = b$ , onde os elementos da matriz  $A$  são dados por  $a_{ij} = a_{ji} = \sum_{k=1}^m g_i(x_k) g_j(x_k)$  e as coordenadas do vetor  $b$  são dadas por  $b_i = \sum_{k=1}^m y_k g_i(x_k)$ .

**Exemplo 2. (Ajuste linear)** Considere a tabela dada no **Exemplo 1**:

$x_k$	$x_1 = 26$	$x_2 = 26.6$	$x_3 = 27.1$	$x_4 = 27.6$	$x_5 = 28.1$	$x_6 = 28.5$
$y_k$	$y_1 = 26.6$	$y_2 = 27.1$	$y_3 = 27.6$	$y_4 = 28.1$	$y_5 = 28.5$	$y_6 = 28.9$

Vamos obter os parâmetros ótimos da seguinte função de ajuste:  $\Phi(x) = \alpha_1 + \alpha_2 x$ . Nesse caso,  $g_1(x) = 1$  e  $g_2(x) = x$ . Os elementos da matriz e os termos independentes do sistema linear proveniente do Método dos Quadrados Mínimos são dados por:

$$a_{11} = \sum_{k=1}^6 g_1^2(x_k) = \sum_{k=1}^6 1 = 6 ; \quad a_{12} = a_{21} = \sum_{k=1}^6 g_2(x_k) g_1(x_k) = \sum_{k=1}^6 x_k = 163.9 ;$$

$$a_{22} = \sum_{k=1}^6 g_2^2(x_k) = \sum_{k=1}^6 x_k^2 = 4481.59 ; \quad b_1 = \sum_{k=1}^6 y_k g_1(x_k) = \sum_{k=1}^6 y_k = 166.8 \quad \text{e}$$

$$b_2 = \sum_{k=1}^6 y_k g_2(x_k) = \sum_{k=1}^6 y_k x_k = 4560.48 .$$

Resolvendo o sistema linear  $A\alpha = b$  obtemos:  $\alpha_1 = 2.527155336$  e  $\alpha_2 = 0.925180407$ . Portanto, função de ajuste é dada por  $\Phi(x) = 2.527155336 + 0.925180407 x$ . ■

### 3.2 Caso Discreto – Modelo Não Linear

O Modelo de Ajuste de Curvas é dito não linear sempre que a função de ajuste não estiver no formato  $\Phi(x; v_1, v_2, \dots, v_n) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ .

**Exemplo 3. (Crescimento populacional de Uberlândia)** Considere os dados da Tabela abaixo, que apresenta o crescimento populacional de Uberlândia no período de 1950 a 2010, de acordo com o censo demográfico realizado pelo IBGE.

$t$ (anos)	1950	1960	1970	1980	1991	2000	2010
$y$ (Habitantes $\times 10^3$ )	54.984	88.282	126.112	240.961	367.062	501.214	604.013

Fonte: IBGE – Censos Demográficos 1950, 1960, 1970, 1980, 1991, 2000 e 2010

Modelos de crescimento populacional sugerem que o ajuste de curva seja do tipo exponencial. No diagrama de dispersão exibido a seguir (**Fig. 2**), em vez de utilizarmos os pontos  $(t, y)$ , consideramos os pontos  $(x, y)$ , onde  $x = (t - 1950)/10$ . O diagrama sugere um comportamento de crescimento exponencial com  $\Phi(x; v_1, v_2) = v_1 \exp(v_2 x)$ . ■

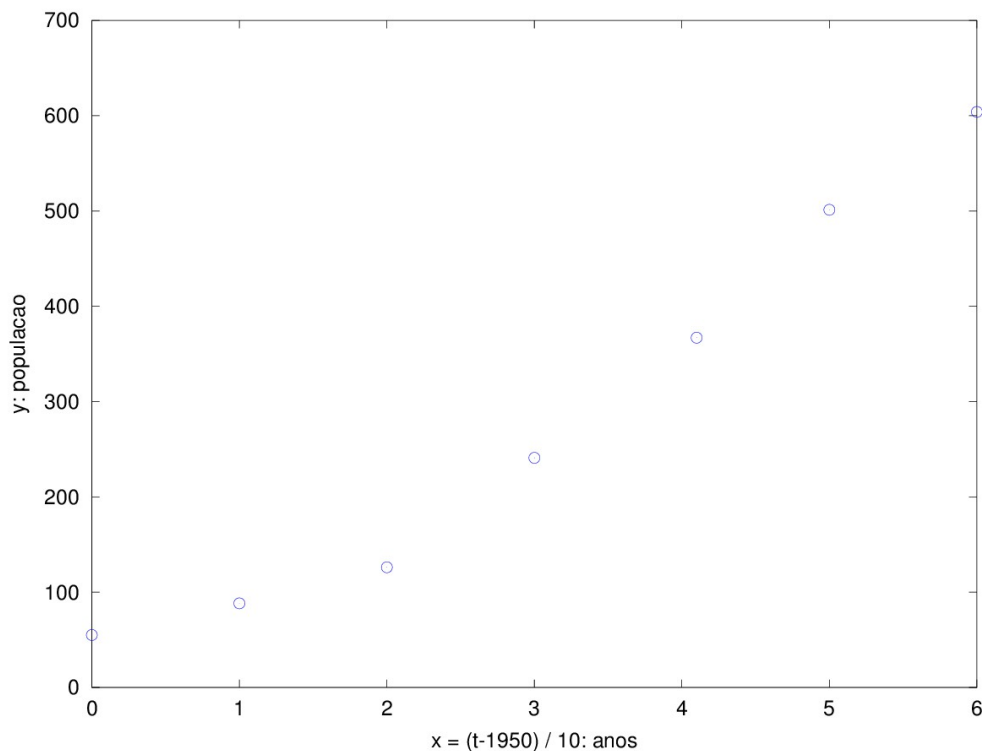
No modelo não linear, a minimização da função erro  $E(v) = \sum_{k=1}^m (\Phi(x_k) - y_k)^2$  conduziria à resolução de um sistema não linear. Para contornar esta dificuldade, apresentaremos uma estratégia de linearização. A solução do modelo linearizado irá aproximar a solução do modelo não linear.

A linearização do modelo é a seguinte: dada uma tabela de pontos  $(x_k, y_k)$ ,  $1 \leq k \leq m$ , onde  $y_k$  é aproximado por uma função  $\Phi(x_k; v_1, v_2, \dots, v_n)$ , a qual não é linear nos



parâmetros  $v_1, v_2, \dots, v_n$ , procuramos uma função  $L$  tal que  $L(\Phi(x_k)) = w_1 g_1(x) + w_2 g_2(x) + \dots + w_n g_n(x)$ . Assim,  $L(\Phi)$  é linear nos novos parâmetros  $w_1, w_2, \dots, w_n$ . Obtida a linearização  $L$ , a próxima etapa consiste em considerar uma nova tabela com pontos  $(x_k, z_k)$ , onde  $z_k = L(y_k)$ ,  $1 \leq k \leq m$ . O Método dos Quadrados Mínimos aplicado ao problema linearizado produzirá um sistema linear:  $Aw = B$ , onde os elementos de  $A$  são dados por  $a_{ij} = a_{ji} = \sum_{k=1}^m g_i(x_k) g_j(x_k)$  e as coordenadas do vetor  $B$  são dadas por  $b_i = \sum_{k=1}^m z_k g_i(x_k)$ .

**Observação:** Não existe uma receita para a obtenção da função de linearização  $L$ . A escolha de  $L$  dependerá do formato da função  $\Phi$ .



**Figura 2 – Módulo 3. Diagrama de Dispersão do Exemplo 2 – Crescimento Exponencial**

**Exemplo 4. (Algumas funções de linearização)** Considere as seguintes funções de ajuste e suas respectivas linearizações:

(1)  $\Phi(x) = 1/(a + cx)$ ;  $L(\Phi(x)) = 1/(\Phi(x)) = a + cx$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Nesse caso, os parâmetros de  $\Phi$  e  $L(\Phi(x))$  são os mesmos:  $v_1 = a$  e  $v_2 = c$ .

(2)  $\Phi(x) = x/(a + cx)$ ;  $L(\Phi(x)) = 1/(\Phi(x)) = a(1/x) + c$ ;  $g_1(x) = 1/x$  e  $g_2(x) = 1$ . Nesse caso, os parâmetros de  $\Phi$  e  $L(\Phi(x))$  são os mesmos:  $v_1 = a$  e  $v_2 = c$ .

(3)  $\Phi(x) = 20 + [1/(a + cx)]$ ;  $L(\Phi(x)) = 1/(\Phi(x) - 20) = a + cx$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Nesse caso, os parâmetros de  $\Phi$  e  $L(\Phi(x))$  são os mesmos:  $v_1 = a$  e  $v_2 = c$ .

(4)  $\Phi(x) = y^*/(c \cdot e^{(-ax)} + 1)$ , onde  $y^*$  é uma constante dada;  $a$  e  $c$  são parâmetros a serem determinados;  $v_1 = c$  e  $v_2 = a$ ;  $L(\Phi(x)) = \ln([y^* - \Phi(x)]/\Phi(x)) = \ln(c) - ax$ . Assim, os novos parâmetros são  $w_1 = \ln(v_1)$  e  $w_2 = -v_2$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Note que:  $v_1 = \exp(w_1)$  e  $v_2 = -w_2$ .

(5)  $\Phi(x) = 20 + \frac{1}{a\sqrt{1+cx^2}}$ ;  $v_1 = a$  e  $v_2 = c$ ;  $L(\Phi(x)) = 1/(\Phi(x) - 20)^2 = a^2(1 + cx^2) = a^2 + a^2cx^2$ . Assim, os novos parâmetros são  $w_1 = (v_1)^2$  e  $w_2 = (v_1)^2v_2$ ;  $g_1(x) = 1$  e  $g_2(x) = x^2$ . Note que:  $v_1 = \pm\sqrt{w_1}$  e  $v_2 = w_2/w_1$ .

(6)  $\Phi(x) = ac^x = a \cdot \exp(x \cdot \ln(c))$ ;  $v_1 = a$  e  $v_2 = c$ ;  $L(\Phi(x)) = \ln(\Phi(x)) = \ln(a) + x \cdot \ln(c)$ . Assim, os novos parâmetros são  $w_1 = \ln(v_1)$  e  $w_2 = \ln(v_2)$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Note que:  $v_1 = \exp(w_1)$  e  $v_2 = \exp(w_2)$ .

(7)  $\Phi(x) = ax^c = a \exp(c \cdot \ln(x))$ ;  $v_1 = a$  e  $v_2 = c$ ;  $L(\Phi(x)) = \ln(\Phi(x)) = \ln(a) + c \cdot \ln(x)$ . Assim, os novos parâmetros são  $w_1 = \ln(v_1)$  e  $w_2 = v_2$ ;  $g_1(x) = 1$  e  $g_2(x) = \ln(x)$ . Note que:  $v_1 = \exp(w_1)$  e  $v_2 = w_2$ . ■

### Problema 1. (Ajuste não linear – crescimento populacional de Uberlândia)

Considere o crescimento populacional de Uberlândia no período de 1950 a 2000, de acordo com a Tabela do **Exemplo 3**.

$t$ (anos)	1950	1960	1970	1980	1991	2000
$y$ (Habitantes $\times 10^3$ )	54.984	88.282	126.112	240.961	367.062	501.214

Fonte: IBGE – Censos Demográficos 1950, 1960, 1970, 1980, 1991 e 2000.

(i) Em vez dos pontos  $(t, y)$ , considere os pontos  $(x, y)$ , onde  $x = (t - 1950)/10$ . Faça um ajuste dos pontos desta nova tabela utilizando uma família de curvas exponenciais do tipo  $\Phi(x; v_1, v_2) = v_1 \cdot \exp(v_2 \cdot x)$ . Estime a população de Uberlândia em  $t = 2010$ .

(ii) Seja  $y^* = 838.916$  (estimativa obtida no item (i) anterior). Considere uma família de curvas do tipo logística  $\Phi(x) = y^*/(c \cdot e^{(-ax)} + 1)$  para ajustar os pontos da tabela, conforme o item (i) anterior. Obtenha uma estimativa da população de Uberlândia em  $t = 2010$ .

(iii) Levando-se em conta os itens anteriores, compare os valores mínimos da função

$$\text{erro } E = \sum_{k=1}^m (\Phi(x_k) - y_k)^2 .$$

### Solução do Problema 1:

(i) Vamos utilizar a tabela com pontos  $x$  e  $y$ :

$x =$ $(t - 1950)/10$	0	1	2	3	4.1	5
$y$ (Habitantes $\times 10^3$ )	54.984	88.282	126.112	240.96 1	367.062	501.214

Como  $\Phi(x; v_1, v_2) = v_1 \cdot \exp(v_2 \cdot x)$ , então a linearização é dada por  $L(\Phi(x)) = \ln(\Phi(x)) = \ln(v_1) + x \cdot v_2$ . Assim, os novos parâmetros são  $w_1 = \ln(v_1)$  e  $w_2 = v_2$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Note que:  $v_1 = \exp(w_1)$  e  $v_2 = w_2$ .

A tabela que será utilizada no modelo linearizado tem pontos  $(x_k, z_k)$ , onde  $z_k = \ln(y_k)$ ,  $1 \leq k \leq 6$ . Considerando-se 4 casas decimais, obtemos os seguintes valores:  $z_1 = 4.0070$ ;  $z_2 = 4.4805$ ;  $z_3 = 4.8372$ ;  $z_4 = 5.4846$ ;  $z_5 = 5.9055$  e  $z_6 = 6.2170$ .

O sistema linear oriundo da aplicação do Método dos Quadrados Mínimos no problema linearizado é dado por:

$$A = \begin{pmatrix} 6.00 & 15.10 \\ 15.10 & 55.81 \end{pmatrix}, \text{ onde } a_{11} = \sum_{k=1}^m g_1(x_k)g_1(x_k); \quad a_{12} = a_{21} = \sum_{k=1}^m g_2(x_k)g_1(x_k);$$

$$a_{22} = \sum_{k=1}^m g_2(x_k)g_2(x_k) \quad \text{e} \quad b = \begin{pmatrix} 30.9318 \\ 85.9063 \end{pmatrix} \quad (\text{considerando apenas 4 casas decimais}),$$

$$\text{onde } b_1 = \sum_{k=1}^m z_k g_1(x_k) \quad \text{e} \quad b_2 = \sum_{k=1}^m z_k g_2(x_k) .$$

A solução do sistema linear considerando 9 casas decimais é dada por  $w_1 = 4.016115915$  e  $w_2 = 0.452665722$ . Portanto,  $v_1 = \exp(w_1) = 55.48517762$  e  $v_2 = w_2$ . Logo, se  $t = 2010$  então  $x = (2010 - 1950)/10 = 6$ . Dessa forma, a população de Uberlândia estimada para 2010 é dada por  $\Phi(6) = v_1 \cdot \exp(v_2 \cdot 6) \approx 838.916$ , ou seja, a população estimada é de aproximadamente oitocentos e trinta e oito mil novecentos e dezesseis habitantes.

(ii) Seja  $y^* = 838.916$ . Como  $\Phi(x) = y^*/(c \cdot e^{(-ax)} + 1)$  então  $v_1 = c$  e  $v_2 = a$ ;  $L(\Phi(x)) = \ln([y^* - \Phi(x)]/\Phi(x)) = \ln(c) - ax$ . Assim, os novos parâmetros são  $w_1 = \ln(v_1)$  e  $w_2 = -v_2$ ;  $g_1(x) = 1$  e  $g_2(x) = x$ . Note que:  $v_1 = \exp(w_1)$  e  $v_2 = -w_2$ .

A tabela que será utilizada no modelo linearizado tem pontos  $(x_k, z_k)$ , onde  $z_k = \ln([y^* - y_k]/y_k)$ ,  $1 \leq k \leq 6$ . Considerando-se 5 casas decimais, obtemos os seguintes valores:  $z_1 = 2.65728$ ;  $z_2 = 2.14038$ ;  $z_3 = 1.73204$ ;  $z_4 = 0.90888$ ;  $z_5 = 0.25114$  e  $z_6 = -0.39487$ .

A matriz do sistema linear oriundo da aplicação do Método dos Quadrados Mínimos no problema linearizado é exatamente a mesma do item anterior e o vetor de termos independentes é dado por:

$$b = \begin{pmatrix} 7.29485 \\ 7.38642 \end{pmatrix} \text{ (considerando apenas 5 casas decimais), onde } b_1 = \sum_{k=1}^m z_k g_1(x_k) \text{ e } b_2 = \sum_{k=1}^m z_k g_2(x_k).$$

A solução do sistema linear considerando 10 casas decimais é dada por  $w_1 = 2.7664066958$  e  $w_2 = -0.6161319304$ . Portanto,  $c = v_1 = \exp(w_1) = 15.90139269$  e  $a = v_2 = -w_2 = 0.6161319304$ . Logo, se  $t = 2010$ , então  $x = (2010 - 1950)/10 = 6$ . Dessa forma, a população de Uberlândia estimada para 2010 é dada por  $\Phi(6) = 838.916/(c \cdot e^{(-ax)} + 1) \approx 601.631$  mil habitantes, ou seja, a população estimada é de aproximadamente seiscentos e um mil seiscentos e trinta e um habitantes.

(iii) No primeiro caso, o valor mínimo da função erro,  $E$ , é aproximadamente 1947.9533. No segundo caso, o valor mínimo da função erro é aproximadamente 820.5534. Isso ajuda a justificar por que a segunda estimativa foi melhor do que a primeira. Lembre-se que o valor correto (de acordo com o censo realizado pelo IBGE) para a população de Uberlândia, em 2010, é dado por 604.013 mil habitantes. ■


## Teste de alinhamento

Na dúvida entre duas ou mais funções de linearização do modelo relacionado ao caso não linear do Método dos Quadrados Mínimos, sempre que possível, é aconselhável utilizar a técnica conhecida como teste de alinhamento.

Dadas uma tabela de pontos  $(x_k, y_k)$ ,  $1 \leq k \leq m$ ,  $y_k$  sendo aproximado por uma função  $\Phi(x_k; v_1, v_2)$ , que não é linear em seus parâmetros, e uma linearização  $L$ , tal que  $L(\Phi(x_k)) = w_1 + w_2 g_2(x)$ , o teste de alinhamento consiste em verificar se o diagrama de dispersão dos pontos  $(g_2(x_k), L(y_k))$ ,  $1 \leq k \leq m$ , assemelha-se a uma reta. Quanto mais alinhados estiverem os pontos do diagrama, mais adequada será a linearização.

**Observação:** Todas as funções de linearização exibidas no **Exemplo 4** podem ser avaliadas de acordo com o teste de alinhamento.

**Observação:** O próximo exemplo foi retirado da lista de exercícios da seguinte referência:

 [2] RUGGIERO, M.A.G., LOPES, V.L.R., *Cálculo Numérico: Aspectos Teóricos e Computacionais*, 2ª Edição, São Paulo, Pearson Education do Brasil, 1996.

**Exemplo 5. (Teste de alinhamento)** Vamos aplicar o teste de alinhamento no seguinte problema que foi proposto na lista de exercícios da referência [2] acima: “Dada a tabela

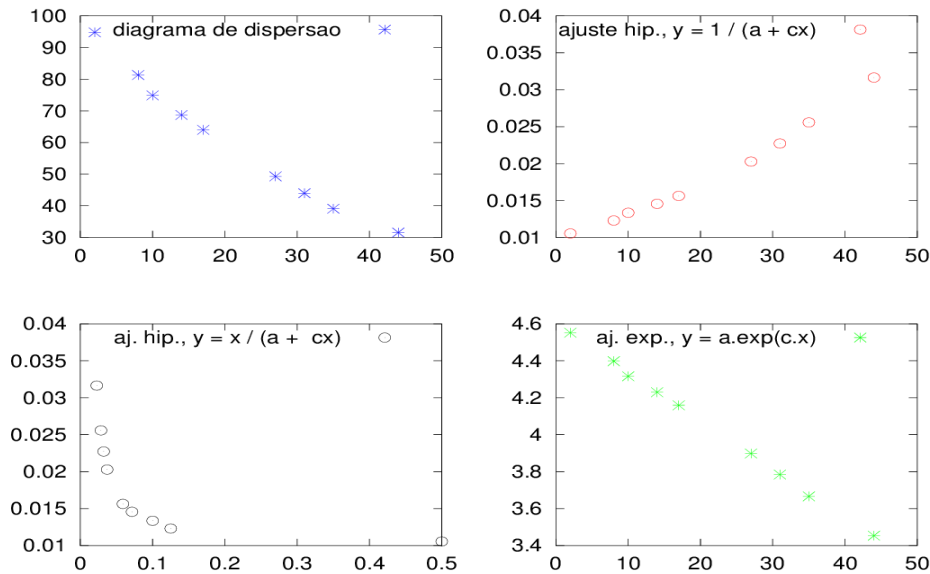
$x_k$	2	5	8	10	14	17	27	31	35	44
$y_k$	94.8	98.7	81.3	74.9	68.7	64	49.3	44	39.1	31.6

e as funções:  $\Phi_1(x) = 1/(a + cx)$ ;  $\Phi_2(x) = x/(a + cx)$ ;  $\Phi_3(x) = a.exp(x.c)$ ; qual é a melhor família de curvas para ajustar estes pontos?”

As linearizações utilizadas são aquelas exibidas no **Exemplo 4**, itens 1, 2 e 6 respectivamente. Dessa forma, temos que fazer os seguintes diagramas de dispersão:

$$(1) (x_k, 1/y_k); \quad (2) (1/x_k, 1/y_k); \quad (3) (x_k, \ln(y_k)); \quad 1 \leq k \leq 10.$$

A **Figura 3** exibe o teste de alinhamento para as três situações anteriores e, também, o diagrama de dispersão dos pontos da tabela original.



**Figura 3 – Módulo 3. Teste de alinhamento – Exemplo 5**

Para ficar mais evidente o alinhamento, o segundo ponto da tabela, (5, 98.8), foi retirado. Analisando-se os diagramas produzidos pelo teste de alinhamento, fica claro que o ajuste exponencial irá produzir o melhor ajuste dos pontos. Se a dúvida persistir, o ideal é aplicar o Método dos Quadrados Mínimos no problema linearizado e comparar o valor mínimo de cada uma das funções erro  $E = \sum_{k=1}^m (\Phi(x_k) - y_k)^2$ . ■

### 3.3 Caso Contínuo – Modelo Linear

No caso contínuo do Método dos Quadrados Mínimos, Modelo Linear, o objetivo é o de encontrar a melhor aproximação para uma função contínua  $f: [a, b] \rightarrow \mathbb{R}$  dentre uma família de curvas do tipo  $\Phi(x; v_1, v_2, \dots, v_n) = v_1 g_1(x) + v_2 g_2(x) + \dots + v_n g_n(x)$ . A melhor aproximação é no sentido de minimizar a seguinte função de  $n$  variáveis reais:

$$E(v_1, v_2, \dots, v_n) = \int_a^b |\Phi(x) - f(x)|^2 dx \quad .$$

Analogamente ao Caso Discreto do Modelo Linear, visto na Seção 3.1, se as funções  $g_1(x)$ ,  $g_2(x)$ ,  $\dots$ ,  $g_n(x)$  forem linearmente independentes, então o vetor ótimo de parâmetros,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ , que minimiza a função  $E$  é a solução do sistema linear

$A\alpha = b$  onde os elementos da matriz  $A$  são dados por  $a_{ij} = a_{ji} = \int_a^b g_i(x)g_j(x) dx$  e as coordenadas do vetor  $b$  são dadas por  $b_i = \int_a^b f(x)g_i(x) dx$ ;  $1 \leq i \leq n$  e  $1 \leq j \leq n$ .

**Exercício 3. (Caso Contínuo – Modelo Linear)** Seja  $f: [0, 2\pi] \rightarrow \mathbb{R}$  uma função contínua. Obtenha os parâmetros ótimos da família  $\Phi(x) = v_0 + v_1 \cos(x) + v_2 \cos(2x) + \dots + v_n \cos(nx) + w_1 \sin(x) + w_2 \sin(2x) + \dots + w_n \sin(nx)$ ,  $x \in [0, 2\pi]$ , que melhor aproxime a função  $f$  no sentido do Método dos Quadrados Mínimos.

### Solução do Exercício 3:

Primeiramente faremos um caso particular considerando-se  $n = 2$ . Dessa forma, a matriz dos coeficientes do sistema linear oriundo do Método dos Quadrados Mínimos possuirá ordem 5 (note que a quantidade de parâmetros é igual a  $2n + 1$ ):

$$A = \begin{pmatrix} (1, 1) & (1, \cos(x)) & (1, \cos(2x)) & (1, \sin(x)) & (1, \sin(2x)) \\ (\cos(x), 1) & (\cos(x), \cos(x)) & (\cos(x), \cos(2x)) & (\cos(x), \sin(x)) & (\cos(x), \sin(2x)) \\ (\cos(2x), 1) & (\cos(2x), \cos(x)) & (\cos(2x), \cos(2x)) & (\cos(2x), \sin(x)) & (\cos(2x), \sin(2x)) \\ (\sin(x), 1) & (\sin(x), \cos(x)) & (\sin(x), \cos(2x)) & (\sin(x), \sin(x)) & (\sin(x), \sin(2x)) \\ (\sin(2x), 1) & (\sin(2x), \cos(x)) & (\sin(2x), \cos(2x)) & (\sin(2x), \sin(x)) & (\sin(2x), \sin(2x)) \end{pmatrix}.$$

O vetor de termos independentes, nesse caso, será dado por:

$$b = \begin{pmatrix} (1, f(x)) \\ (\cos(x), f(x)) \\ (\cos(2x), f(x)) \\ (\sin(x), f(x)) \\ (\sin(2x), f(x)) \end{pmatrix}.$$

Tanto os elementos da matriz  $A$  quanto os elementos do vetor  $b$  são produtos internos entre funções definidas no intervalo  $[0, 2\pi]$ , conforme a notação usual:  $(g(x), h(x)) =$

$$\int_0^{2\pi} g(x)h(x) dx, \text{ quaisquer que sejam as funções } g \text{ e } h.$$

No caso geral, a matriz do sistema linear possuirá ordem  $2n + 1$ , que é exatamente a quantidade de parâmetros, e o vetor de termos independentes terá essa mesma quantidade de coordenadas. A matriz do sistema linear e o vetor de termos independentes serão análogos aos exibidos no caso particular, no qual foi considerado  $n = 2$ .

A linha 1 (coluna 1) da matriz  $A$  e a coordenada 1 do vetor  $b$  estão associadas à função  $g_0(x) = 1$  e ao parâmetro  $v_0$ ; a linha  $i + 1$  (coluna  $i + 1$ ) da matriz  $A$  e a coordenada  $i +$

1 do vetor  $b$  estão associadas à função  $g_i(x) = \cos(ix)$  e ao parâmetro  $v_i$ ;  $1 \leq i \leq n$ ; a linha  $j + n + 1$  (coluna  $j + n + 1$ ) da matriz  $A$  e a coordenada  $j + n + 1$  do vetor  $b$  estão associadas à função  $g_j(x) = \sin(jx)$  e ao parâmetro  $w_j$ ;  $1 \leq j \leq n$ .

A partir das identidades trigonométricas exibidas a seguir, mostraremos que a matriz  $A$  será diagonal com  $a_{11} = 2\pi$  e  $a_{\ell\ell} = \pi$ ,  $2 \leq \ell \leq 2n + 1$ .

Sabendo que

$$\begin{aligned}\cos(x + y) &= \cos(x)\cos(y) - \sin(x)\sin(y); \\ \cos(x - y) &= \cos(x)\cos(y) + \sin(x)\sin(y);\end{aligned}$$

$$\begin{aligned}\sin(x + y) &= \sin(x)\cos(y) + \sin(y)\cos(x); \\ \sin(x - y) &= \sin(x)\cos(y) - \sin(y)\cos(x);\end{aligned}$$

obtemos que

$$\begin{aligned}\cos(x)\cos(y) &= (1/2) [\cos(x + y) + \cos(x - y)]; \\ \sin(x)\sin(y) &= (1/2) [\cos(x - y) - \cos(x + y)]; \\ \sin(x)\cos(y) &= (1/2) [\sin(x + y) + \sin(x - y)].\end{aligned}$$

Além disso,  $\cos(2x) = \cos^2(x) - \sin^2(x) = 2.\cos^2(x) - 1$ , pois  $\sin^2(x) + \cos^2(x) = 1$ . Portanto,

$$\cos^2(x) = [1 + \cos(2x)]/2 \quad \text{e} \quad \sin^2(x) = [1 - \cos(2x)]/2.$$

Dessa forma,  $a_{11} = \int_0^{2\pi} 1^2 dx = 2\pi$  e os demais elementos da linha 1 são nulos porque

$$\int_0^{2\pi} \cos(ix) \cdot 1 dx = (1/i) [\sin(i.2\pi) - \sin(i.0)] = 0 \quad \text{e}$$

$$\int_0^{2\pi} \sin(jx) \cdot 1 dx = (-1/j) [\cos(j.2\pi) - \cos(j.0)] = 0; \quad 1 \leq i \leq n; \quad 1 \leq j \leq n.$$

Na linha  $i + 1$ ,  $1 \leq i \leq n$ , tem-se que o elemento da diagonal é dado por

$$\begin{aligned}\int_0^{2\pi} \cos^2(ix) dx &= \int_0^{2\pi} [0.5 + 0.5 \cos(2ix)] dx \\ &= 0.5 \times 2\pi + (0.5/2i) [\sin(2i.2\pi) - \sin(2i.0)] = \pi.\end{aligned}$$

Os demais elementos dessa linha são nulos porque

$$\int_0^{2\pi} \cos(ix) \cos(lx) dx = \frac{1}{2} \int_0^{2\pi} \cos((i+l)x) + \cos((i-l)x) dx = 0 \quad (\text{veja os}$$

cálculos efetuados após a obtenção de  $a_{11}$ ),  $1 \leq i \leq n$ ;  $1 \leq l \leq n$ ;  $i \neq l$ . Além disso,



$$\int_0^{2\pi} \cos(ix) \operatorname{sen}(jx) dx = \frac{1}{2} \int_0^{2\pi} \operatorname{sen}((j+i)x) + \operatorname{sen}((j-i)x) dx = 0 \quad (\text{veja os}$$

cálculos efetuados após a obtenção de  $a_{11}$ ),  $1 \leq i \leq n$ ;  $1 \leq j \leq n$ ;  $i \neq j$ . Também, tem-se que  $\int_0^{2\pi} \cos(ix) \operatorname{sen}(ix) dx = \frac{1}{2} \int_0^{2\pi} \operatorname{sen}(2ix) dx = 0$ .

Na linha  $j + n + 1$ ,  $1 \leq j \leq n$ , de forma análoga a cálculos já efetuados anteriormente, tem-se que o elemento da diagonal é dado por

$$\int_0^{2\pi} \operatorname{sen}^2(jx) dx = \frac{1}{2} \int_0^{2\pi} [1 - \cos(2jx)] dx = \pi.$$

Os demais elementos desta linha são nulos porque já foi mostrado que

$$\int_0^{2\pi} \cos(ix) \operatorname{sen}(jx) dx = 0, \text{ quaisquer que sejam } i \text{ e } j. \text{ Além disso,}$$

$$\int_0^{2\pi} \operatorname{sen}(jx) \operatorname{sen}(lx) dx = \frac{1}{2} \int_0^{2\pi} \cos((j-l)x) - \cos((j+l)x) dx = 0,$$

$$1 \leq j \leq n; 1 \leq l \leq n; j \neq l.$$

Dessa forma, é fácil concluir que as coordenadas do vetor ótimo de parâmetros são dadas por

$$v_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx ; \quad v_i = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(ix) dx ; \quad w_j = \frac{1}{\pi} \int_0^{2\pi} f(x) \operatorname{sen}(jx) dx ;$$

$$1 \leq i \leq n; 1 \leq j \leq n. \quad \square$$

Os parâmetros ótimos são conhecidos como coeficientes de Fourier. Para saber mais consulte a referência abaixo



[3] FIGUEIREDO, Djairo Guedes de, *Análise de Fourier e Equações Diferenciais Parciais*, Rio de Janeiro, IMPA, 2ª Ed, 1987.

**Referências**

## 4. Interpolação Polinomial


### 4.1 Existência e Unicidade do Polinômio Interpolador

Dada uma função real  $f: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  e dados  $n + 1$  pontos,  $x_0, x_1, \dots, x_n$ , no intervalo fechado  $I$ , dizemos que o polinômio de grau igual a  $n$  ou de grau menor do que  $n$ ,  $p_n(x) = a_0 + a_1x + \dots + a_nx^n$ , é o polinômio interpolador associado à função  $f$  se  $p_n(x_i) = f(x_i)$ , para todo  $i$ ,  $0 \leq i \leq n$ . Observe que as  $n + 1$  equações anteriores dão origem a um sistema linear de ordem  $n + 1$ . A matriz desse sistema linear é conhecida como matriz de Vandermonde. Para o caso particular  $n = 3$ , a matriz de Vandermonde é

$$V = \begin{pmatrix} 1 & x_0 & x_0^2 & x_0^3 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \end{pmatrix} \text{ e o vetor de termos independentes é } B = \begin{pmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ f(x_3) \end{pmatrix}. \text{ Nesse caso, o}$$

determinante de  $V$  é igual a  $(x_1 - x_0).(x_2 - x_0).(x_3 - x_0).(x_2 - x_1).(x_3 - x_1).(x_3 - x_2)$ .

**Observação:** Para relembrar o conceito de determinante de uma matriz e suas propriedades, inclusive o cálculo de determinante pelo desenvolvimento de Laplace, consulte livros de Álgebra Linear.

 [4] BOLDRINI, J. L., COSTA, S. I. R., RIBEIRO, V. L. F. F. E WETZLER, H.G., *Álgebra Linear*, 3ª Edição, São Paulo, Harper & Row Referências do Brasil, 1980.

Retornando ao caso particular, considere as seguintes operações elementares efetuadas nas colunas da matriz de Vandermonde:

$$C_{4-k+1} \rightarrow C_{4-k+1} - x_0.C_{4-k}, \text{ com } 1 \leq k \leq 3.$$

Note que tais operações não alteram o determinante de  $V$ . A partir desse fato, e utilizando o desenvolvimento de Laplace, pela primeira linha, para o cálculo de determinantes, e a propriedade que afirma o seguinte: “se uma linha qualquer de uma matriz  $M$  for multiplicada por um número real  $t$ , não nulo, gerando uma matriz  $M'$ , então  $\det(M') = t \det(M)$ ”, obtemos:

$$\det(V) = \begin{vmatrix} 1 & x_0 & x_0^2 & x_0^3 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & (x_1 - x_0) & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) \\ 1 & (x_2 - x_0) & x_2(x_2 - x_0) & x_2^2(x_2 - x_0) \\ 1 & (x_3 - x_0) & x_3(x_3 - x_0) & x_3^2(x_3 - x_0) \end{vmatrix}$$

$$= \begin{vmatrix} (x_1 - x_0) & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) \\ (x_2 - x_0) & x_2(x_2 - x_0) & x_2^2(x_2 - x_0) \\ (x_3 - x_0) & x_3(x_3 - x_0) & x_3^2(x_3 - x_0) \end{vmatrix} = (x_1 - x_0).(x_2 - x_0).(x_3 - x_0). \begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix}.$$

Repetindo-se o mesmo argumento, porém considerando-se  $x_1$  no lugar de  $x_0$ , obtém-se que

$$\begin{aligned} \det(V) &= (x_1 - x_0).(x_2 - x_0).(x_3 - x_0). \begin{vmatrix} 1 & 0 & 0 \\ 1 & (x_2 - x_1) & x_2(x_2 - x_1) \\ 1 & (x_3 - x_1) & x_3(x_3 - x_1) \end{vmatrix} \\ &= (x_1 - x_0).(x_2 - x_0).(x_3 - x_0). \begin{vmatrix} (x_2 - x_1) & x_2(x_2 - x_1) \\ (x_3 - x_1) & x_3(x_3 - x_1) \end{vmatrix} \\ &= (x_1 - x_0).(x_2 - x_0).(x_3 - x_0).(x_2 - x_1).(x_3 - x_1). \begin{vmatrix} 1 & x_2 \\ 1 & x_3 \end{vmatrix} \\ &= (x_1 - x_0).(x_2 - x_0).(x_3 - x_0).(x_2 - x_1).(x_3 - x_1).(x_3 - x_2). \end{aligned}$$

No caso geral, será mostrado que  $\det(V) = \prod_{j=0}^{n-1} \prod_{i=j+1}^n (x_i - x_j) = \prod_{0 \leq j < i \leq n} (x_i - x_j)$ . Dessa forma, se os pontos forem distintos dois a dois, isto é,  $x_i \neq x_j$  para  $i \neq j$ , então  $\det(V) \neq 0$  e, portanto, existirá um único polinômio interpolador (pois o sistema linear possuirá uma única solução, que corresponderá aos coeficientes do polinômio interpolador).

**Teorema 1. (Existência e Unicidade do Polinômio Interpolador)** Dada uma função real  $f: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  e dados  $n + 1$  pontos distintos dois a dois,  $x_0, x_1, \dots, x_n$ , no intervalo fechado  $I$ , existe um único polinômio de grau até  $n$ ,  $p_n(x) = a_0 + a_1x + \dots + a_nx^n$ , tal que  $p_n(x_i) = f(x_i)$ , para todo  $i$ ,  $0 \leq i \leq n$ .

**Demonstração:** Utilizaremos indução finita sobre a dimensão da matriz de Vandermonde, a qual será denotada por  $\dim(V)$ .

Seja  $\dim(V) = m + 1 = 2$ . Então  $V = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix}$  e, portanto,  $\det(V) = (x_1 - x_0) \neq 0$ ; quaisquer que sejam  $x_0$  e  $x_1$  distintos.

Seja  $V$  uma matriz de ordem  $m + 1$  que tenha a mesma estrutura de uma matriz de Vandermonde, isto é, os seus elementos são construídos da mesma forma que os elementos da matriz de Vandermonde. Note que tal matriz tem elementos iguais a 1 na primeira coluna. Se a segunda coluna dessa matriz for composta pelo vetor  $v = (v_1, v_2, \dots, v_{m+1})$ , com coordenadas distintas duas a duas, então a coluna  $j$  desta mesma matriz, com  $2 < j \leq m + 1$ , será composta pelo vetor  $([v_1]^{j-1}, [v_2]^{j-1}, \dots, [v_{m+1}]^{j-1})$ .

Suponha que exista um número natural  $n > 1$  tal que toda matriz  $V$  de ordem  $m + 1$ , com a mesma estrutura de uma matriz de Vandermonde, possua determinante dado por

$$\prod_{j=1}^m \prod_{i=j+1}^{m+1} (v_i - v_j) = \prod_{1 \leq j < i \leq m+1} (v_i - v_j), \text{ para } 1 \leq m \leq n - 1.$$

Seja  $N = n + 1$ . Mostraremos que se  $\dim(V) = N$  então a fórmula do determinante continuará valendo, ou seja,

$$\begin{aligned} \det(V) &= \prod_{j=1}^n \prod_{i=j+1}^{n+1} (v_i - v_j) = \prod_{1 \leq j < i \leq n+1} (v_i - v_j) \leftrightarrow \\ &\leftrightarrow \det(V) = \prod_{j=0}^{n-1} \prod_{i=j+1}^n (x_i - x_j) = \prod_{0 \leq j < i \leq n} (x_i - x_j). \end{aligned}$$

Para isso, considere as seguintes operações elementares efetuadas nas colunas da matriz  $V$ :

$$C_{N-k+1} \rightarrow C_{N-k+1} - x_0 \cdot C_{N-k}, \text{ com } 1 \leq k \leq N - 1.$$

Note que tais operações não alteram o determinante de  $V$ . De acordo com a formação dos elementos desta matriz, após a execução de cada uma das operações elementares anteriores, a coluna  $C_{N-k+1}$  passará a ter, na posição  $i + 1$ ,  $0 \leq i \leq N - 1$ , o seguinte elemento:  $[x_i]^{N-k} - x_0 \cdot [x_i]^{N-k-1} = [x_i]^{N-k-1} (x_i - x_0)$ . Portanto, o primeiro elemento de cada uma das colunas anteriores ( $i = 0$ ) será nulo. Dessa forma, a linha  $i + 1$  da matriz  $V_1$  obtida após a execução das operações elementares será formada pelos elementos

$$1, (x_i - x_0), x_i(x_i - x_0), [x_i]^2(x_i - x_0), \dots, [x_i]^{N-2}(x_i - x_0).$$

Calculando o determinante de  $V_1$  pelo desenvolvimento de Laplace pela primeira coluna e utilizando a propriedade de determinante mencionada logo após a referência [3] de

Álgebra Linear, obtém-se que  $\det(V) = \det(V_1) = \prod_{i=1}^n (x_i - x_0) \cdot \det(V_2)$ , onde  $V_2$  é

uma matriz de ordem  $n$  que tem a mesma estrutura de uma matriz de Vandermonde. A segunda coluna dessa matriz é composta pelo vetor  $v = (x_1, x_2, \dots, x_n)$ . Portanto, pela

hipótese de indução,  $\det(V_2) = \prod_{j=1}^{n-1} \prod_{i=j+1}^n (x_i - x_j) = \prod_{1 \leq j < i \leq n} (x_i - x_j)$ . Portanto,

$$\det(V) = \prod_{i=1}^n (x_i - x_0) \prod_{1 \leq j < i \leq n} (x_i - x_j) = \prod_{0 \leq j < i \leq n} (x_i - x_j) = \prod_{j=0}^{n-1} \prod_{i=j+1}^n (x_i - x_j). \quad \square$$

**Exemplo 6. (Mal condicionamento da matriz de Vandermonde)** O Teorema 1 anterior apresentou as condições que garantiram a existência e a unicidade do polinômio interpolador. Porém, a resolução do sistema linear com a matriz de Vandermonde não é

a melhor opção para se obter o polinômio interpolador. O mal condicionamento desta matriz será presenciado neste exemplo que foi retirado da referência [2] da Seção 3.2: “Utilizando aritmética de ponto flutuante com 3 dígitos e o Método da Eliminação de Gauss (veja o **Exemplo 11** da Seção 4.2 do Módulo 2), o polinômio interpolador de grau 3 associado à função  $f$  tabelada abaixo:

$x_k$	0.1	0.2	0.3	0.4
$f(x_k)$	5	13	-4	-8

é dado por  $p_3(x) = -0.66 \times 10^2 + (0.115 \times 10^4)x - (0.505 \times 10^4)x^2 + (0.633 \times 10^4)x^3$ . Porém,  $p_3(0.4) = -6 \neq -8 = f(0.4)$ . Portanto, a resolução do sistema linear não está correta (imprecisão devida a erros de arredondamento), indicando o mal condicionamento (grande sensibilidade a erros de arredondamento) da matriz de Vandermonde.” ■

## 4.2 Polinômio de Lagrange

Sabendo que o polinômio interpolador é único, se os pontos de interpolação  $x_0, x_1, \dots, x_n$  forem distintos dois a dois, vamos obtê-lo de outra maneira que não envolva a resolução de um sistema linear.

Em vez de escrever o polinômio interpolador na base canônica,  $\{1, x, x^2, \dots, x^n\}$ , do espaço vetorial dos polinômios de grau até  $n$ ,  $\mathcal{P}_n$ , vamos encontrar uma outra base para este espaço vetorial. Como a dimensão de  $\mathcal{P}_n$  é  $n + 1$ , então qualquer base desse espaço deve ter  $n + 1$  polinômios linearmente independentes. Seja  $\mathcal{L}_n = \{L_0(x), L_1(x), L_2(x), \dots, L_n(x)\}$  esta nova base, de modo que o polinômio interpolador da função  $f$  seja escrito como  $p_n(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x) + \dots + f(x_n)L_n(x) = \sum_{j=0}^n f(x_j)L_j(x)$  e cada um dos polinômios de Lagrange,  $L_j(x)$ , possua grau  $n$ .

Baseando-se na definição do polinômio interpolador,  $p_n(x_i) = f(x_i)$ ,  $0 \leq i \leq n$ , é fácil concluir que os polinômios de Lagrange devem possuir as seguintes propriedades:

$$L_j(x_i) = 1, \text{ se } i = j \quad \text{e} \quad L_j(x_i) = 0, \text{ se } i \neq j.$$

Como o polinômio  $L_j(x)$  possui grau  $n$  e  $x_i$ ,  $0 \leq i \leq n$ ,  $i \neq j$ , são raízes desse polinômio ( $L_j(x_i) = 0$ , se  $i \neq j$ ), então o polinômio de Lagrange é dado por  $c.N_j(x_i)$ , onde  $c$  é uma constante não nula e  $N_j(x) = (x - x_0)(x - x_1) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n) = \prod_{i=0; i \neq j}^n (x - x_i)$ . Além disso,  $L_j(x_j) = 1$ . Dessa forma, basta considerar  $c = N_j(x_j) \neq 0$  (lembre-se que os pontos de interpolação são distintos dois a dois). Conclusão: os polinômios de Lagrange são dados por  $L_j(x) = N_j(x)/N_j(x_j)$ .

**Observação:** É comum utilizar o polinômio  $N(x) = \prod_{i=0}^n (x - x_i)$ , de grau  $n + 1$ , para escrever  $N_j(x) = N(x)/(x - x_j)$ , se  $x \neq x_j$  e  $N_j(x_j) = N'(x_j)$  (derivada de  $N(x)$  no ponto  $x = x_j$ ).

**Exemplo 7. (Polinômio de Lagrange)** Sejam  $x_0, x_1$  e  $x_2$  pontos de interpolação distintos dois a dois. A base com polinômios de Lagrange é dada por  $\mathcal{L}_2 = \{L_0(x), L_1(x), L_2(x)\}$ , onde

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} ; \quad L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} ;$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} .$$

O polinômio interpolador é dado por  $p_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x)$ . ■

**Exercício 4. (Interpolação de Lagrange)** Utilizando-se um número finito de parcelas da série de potências da função exponencial  $e^x$ , foi possível construir a tabela abaixo com alguns valores da função  $f(x) = \int_0^x e^{-t^2} dt$  :

$x$	0.5	0.6	0.8	0.9	1.1
$f(x)$	0.46128	0.53515	0.65767	0.70624	0.78006

Use um polinômio de Lagrange de grau 2 para aproximar o valor de  $f(0.78)$ .

#### Solução do Exercício 4:

O primeiro passo consiste na escolha dos pontos de interpolação. Como a função será aproximada no ponto  $x^* = 0.78$ , vamos escolher os pontos  $x_0, x_1$  e  $x_2$  que estiverem mais próximos de  $x^*$  de modo que o valor de  $|N(x^*)| = |x^* - x_0| \cdot |x^* - x_1| \cdot |x^* - x_2|$  seja o menor possível.

Existem duas opções para a escolha dos pontos: (i)  $x_0 = 0.5, x_1 = 0.6$  e  $x_2 = 0.8$  ou (ii)  $x_0 = 0.6, x_1 = 0.8$  e  $x_2 = 0.9$ . O menor valor de  $|N(x^*)|$  é obtido com a segunda opção (ii).

Obtidos os pontos de interpolação, o segundo passo consiste na construção do polinômio interpolador:

$$p_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x),$$

com

$$x_0 = 0.6, \quad x_1 = 0.8 \quad \text{e} \quad x_2 = 0.9;$$

$$f(x_0) = 0.53515, \quad f(x_1) = 0.65767 \quad \text{e} \quad f(x_2) = 0.70624.$$

Os polinômios de Lagrange foram exibidos acima, no **Exemplo 7**.

Finalmente, a aproximação desejada de  $f(0.78)$  é obtida com o cálculo do valor numérico do polinômio interpolador, ou seja,  $f(0.78) \approx p_2(0.78)$ . Como

$$L_0(x^*) = (0.78 - 0.8)(0.78 - 0.9)/(0.6 - 0.8)(0.6 - 0.9) = 0.04,$$

$$L_1(x^*) = (0.78 - 0.6)(0.78 - 0.9)/(0.8 - 0.6)(0.8 - 0.9) = 1.08,$$

$$L_2(x^*) = (0.78 - 0.6)(0.78 - 0.8)/(0.9 - 0.6)(0.9 - 0.8) = -0.12,$$

então  $f(0.78) \approx p_2(0.78) = 0.6469408$ . □



$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \rightarrow e^{-t^2} = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{n!} \rightarrow \int_0^x e^{-t^2} dt = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^x t^{2n} dt.$$

### 4.3 Polinômio de Newton com Diferenças Divididas

Analizando o exercício apresentado no final da seção anterior, podemos averiguar que não seria uma tarefa nada fácil conseguir reaproveitar os cálculos efetuados com um polinômio interpolador de Lagrange de grau  $n$  para exibir uma outra aproximação da função utilizando-se um polinômio interpolador de grau  $n + 1$ . Além disso, o custo computacional do cálculo do valor numérico do polinômio interpolador de Lagrange é elevado porque todos os polinômios da base do espaço vetorial  $\mathcal{P}_n$  possuem o mesmo grau ( $n$ ).

Com o intuito de diminuir o esforço computacional referente ao cálculo do valor numérico do polinômio interpolador e de produzir um esquema recursivo para a geração de polinômios interpoladores de grau 0 (polinômio constante) até grau  $n$ , uma nova base do espaço  $\mathcal{P}_n$  será considerada:

$$\mathcal{B} = \{1, (x - x_0), (x - x_0)(x - x_1), (x - x_0)(x - x_1)(x - x_2), \dots, (x - x_0)(x - x_1)\dots(x - x_{n-1})\}.$$

**Observação:** Utilizando indução finita, não é difícil mostrar que os polinômios da base  $\mathcal{B}$  são linearmente independentes, isto é,

$$d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + \dots + d_n(x - x_0)(x - x_1)\dots(x - x_{n-1}) = 0, \forall x \in \mathbb{R},$$

se, e somente se,  $d_i = 0$ ,  $\forall i$ ,  $0 \leq i \leq n$ .

Para isso, basta observar que se  $x = x_0$ , então  $d_0 = 0$ , pois todas as parcelas acima, a partir da segunda, tem o fator  $(x - x_0)$ . Suponha que  $d_i = 0$  para  $i < k \leq n$  (hipótese de

indução). Note que o fator  $(x - x_k)$  está presente em todas as parcelas a partir da parcela  $k + 1$ . Assim, se  $x = x_k$  então  $d_k(x_k - x_0)(x_k - x_1)\dots(x_k - x_{k-1}) = 0$ . Como os pontos de interpolação são distintos dois a dois, segue-se que  $d_k = 0$ .

O polinômio interpolador com diferenças divididas de grau  $n$  é dado por

$$p_n(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + \dots + d_n(x - x_0)(x - x_1)\dots(x - x_{n-1}).$$

Cada coeficiente da combinação linear anterior,  $d_i$ , é chamado de diferença dividida de ordem  $i$ . De acordo com a definição do polinômio interpolador, tem-se que

$$p_n(x_0) = f(x_0) \rightarrow d_0 = f(x_0);$$

$$p_n(x_1) = f(x_1) \rightarrow f(x_1) = f(x_0) + d_1(x_1 - x_0) \rightarrow d_1 = [f(x_1) - f(x_0)]/(x_1 - x_0);$$

$$p_n(x_2) = f(x_2) \rightarrow f(x_2) = f(x_0) + d_1(x_2 - x_0) + d_2(x_2 - x_0)(x_2 - x_1) \rightarrow$$

$$\rightarrow d_2 = \frac{1}{(x_2 - x_0)} \left[ \frac{f(x_2) - f(x_0)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \frac{(x_2 - x_0)}{(x_2 - x_1)} \right].$$

Somando e subtraindo  $f(x_1)$  no numerador da primeira parcela entre chaves e depois colocando em evidência o valor  $[f(x_1) - f(x_0)]/(x_2 - x_1)$ , obtém-se

$$d_2 = \frac{1}{(x_2 - x_0)} \left\{ \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_2 - x_1} \left[ \frac{x_2 - x_0}{x_1 - x_0} - 1 \right] \right\} \rightarrow$$

$$\rightarrow d_2 = \frac{1}{(x_2 - x_0)} \left[ \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right].$$

Dos casos analisados anteriormente, podemos definir o seguinte operador de diferenças divididas:

(0) Ordem zero:  $f[w] = f(w)$ , para todo  $w \in \text{dom}\{f\}$  (domínio da função  $f$ ).

**Observação:** a diferença dividida de ordem zero depende de apenas um argumento (ponto  $w$ ) e coincide com o valor da função no ponto.

(1) Ordem 1 (2 argumentos):  $f[w_1, w_2] = \frac{f[w_2] - f[w_1]}{(w_2 - w_1)}, \forall w_1 \in \text{dom}\{f\}, \forall w_2 \in \text{dom}\{f\}.$



(2) Ordem 2 (3 argumentos):  $f[w_1, w_2, w_3] = \frac{f[w_2, w_3] - f[w_1, w_2]}{(w_3 - w_1)}, \quad \forall w_i \in \text{dom}\{f\},$   
 $i \in \{1, 2, 3\}.$

Recursivamente, definimos o operador de diferenças divididas de ordem  $n$ :

$f[w_1, w_2, \dots, w_n, w_{n+1}] = \frac{f[w_2, \dots, w_{n+1}] - f[w_1, \dots, w_n]}{(w_{n+1} - w_1)},$  o qual depende de  $n + 1$  argumentos.

**Observação:** O numerador do operador de diferenças divididas de ordem  $n$  é uma diferença entre dois operadores de ordem  $n - 1$  e o denominador do operador de ordem  $n$  é a diferença entre os extremos dos argumentos deste operador, mais precisamente, a diferença entre o extremo superior dos argumentos e o extremo inferior dos argumentos.

Com essa notação, tem-se que  $d_2 = f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{(x_2 - x_0)}.$

Vamos mostrar que vale a seguinte propriedade geral:

$$d_i = f[x_0, x_1, x_2, \dots, x_i] = \frac{f[x_1, \dots, x_i] - f[x_0, \dots, x_{i-1}]}{(x_i - x_0)}, \quad 1 \leq i \leq n.$$

**Teorema 2. (1ª propriedade do operador de diferenças divididas)** Considere  $n + 1$  pontos distintos dois a dois. Utilizando a notação apresentada anteriormente, tem-se que

$$f[x_0, x_1, x_2, \dots, x_i] = f[x_{\sigma(0)}, x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(i)}], \quad 1 \leq i \leq n,$$

onde  $\sigma: \{0, 1, 2, \dots, i\} \rightarrow \{0, 1, 2, \dots, i\}$  é uma função bijetora ou uma permutação. Isso equivale a dizer que a ordem dos argumentos não altera o valor do operador de diferenças divididas.

**Demonstração:** Considere as seguintes bases do espaço vetorial  $\mathcal{P}_i$ :

$$\mathcal{B} = \{1, (x - x_0), (x - x_0)(x - x_1), (x - x_0)(x - x_1)(x - x_2), \dots, (x - x_0)(x - x_1)\dots(x - x_{i-1})\}$$

e

$$\mathcal{B}_\sigma = \{1, (x - x_{\sigma(0)}), (x - x_{\sigma(0)})(x - x_{\sigma(1)}), (x - x_{\sigma(0)})(x - x_{\sigma(1)})(x - x_{\sigma(2)}), \dots, (x - x_{\sigma(0)})(x - x_{\sigma(1)})\dots(x - x_{\sigma(i-1)})\}.$$

Escrevendo os polinômios interpoladores de grau  $i$ , em relação a cada uma das duas bases, obtemos:

$$p_i(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + \dots + d_i(x - x_0)(x - x_1)\dots(x - x_{i-1});$$

$$p_{i\sigma}(x) = D_0 + D_1(x - x_{\sigma(0)}) + D_2(x - x_{\sigma(0)})(x - x_{\sigma(1)}) + \dots + D_i(x - x_{\sigma(0)})(x - x_{\sigma(1)})\dots(x - x_{\sigma(i-1)});$$

onde  $d_i$  e  $D_i$  são as diferenças divididas de ordem  $i$  com relação a cada uma das bases  $\mathcal{B}$  e  $\mathcal{B}_\sigma$ , respectivamente.

Como o polinômio interpolador é único (**Teorema 1**, Seção 4.1), então o coeficiente de  $x^i$  deve ser o mesmo nas duas representações dos polinômios anteriores. Note que o coeficiente de  $x^i$  é igual a  $d_i$ , na primeira representação, e é igual a  $D_i$ , na segunda representação. Portanto, de acordo com a notação utilizada anteriormente,

$$f[x_0, x_1, x_2, \dots, x_i] = d_i = D_i = f[x_{\sigma(0)}, x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(i)}], \quad \forall i, 1 \leq i \leq n. \quad \square$$

**Observação:** O seguinte resultado será útil na demonstração do próximo teorema: “O coeficiente de  $x^m$  no polinômio  $(x - x_0)(x - x_1)\dots(x - x_m)$  é dado por  $-(x_0 + x_1 + \dots + x_m)$ .” A demonstração deste resultado pode ser feita por indução finita (verifique!).

O professor Tadashi Yokoyama, do Departamento de Matemática da Unesp – Rio Claro, substituindo o professor Bezerra, em uma aula de Cálculo Numérico, foi quem demonstrou para a minha turma de graduação em Matemática (turma de 1986) o teorema que será enunciado a seguir. Sou muito grato a todos os meus professores que, com talento para ensinar, souberam transmitir não apenas conhecimento mas, também, amor pela profissão. Profissão que ainda vem sendo exercida pelo professor Tadashi, conforme me informou a Maria Luiza Furlan Cardoso, Supervisora Técnica de Seção Técnica de Desenvolvimento e Administração de Recursos Humanos do IGCE - Campus de Rio Claro.

**Teorema 3. (2ª propriedade do operador de diferenças divididas)** Considere  $n + 1$  pontos distintos dois a dois. A diferença dividida de ordem  $i$  é dada por:

$$d_i = f[x_0, x_1, x_2, \dots, x_i] = \frac{f[x_1, \dots, x_i] - f[x_0, \dots, x_{i-1}]}{(x_i - x_0)}, \quad 1 \leq i \leq n.$$

**Demonstração:** Considere as seguintes bases do espaço vetorial  $\mathcal{P}_i$ :

$\mathcal{B} = \{1, (x - x_0), (x - x_0)(x - x_1), (x - x_0)(x - x_1)(x - x_2), \dots, (x - x_0)(x - x_1)\dots(x - x_{i-1})\}$   
e

$$\mathcal{B}_\sigma = \{1, (x - x_{\sigma(0)}), (x - x_{\sigma(0)})(x - x_{\sigma(1)}), (x - x_{\sigma(0)})(x - x_{\sigma(1)})(x - x_{\sigma(2)}), \dots, (x - x_{\sigma(0)})(x - x_{\sigma(1)})\dots(x - x_{\sigma(i-1)})\},$$

em que a permutação é definida como  $\sigma(k) = i - k$ , para todo  $k, 0 \leq k \leq i$ . Dessa forma,

$$\mathcal{B}_\sigma = \{1, (x - x_i), (x - x_i)(x - x_{i-1}), (x - x_i)(x - x_{i-1})(x - x_{i-2}), \dots, (x - x_i)(x - x_{i-1}) \dots (x - x_1)\}.$$

Escrevendo os polinômios interpoladores de grau  $i$ , em relação a cada uma das duas bases, obtemos:

$$p_i(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + \dots + d_i(x - x_0)(x - x_1) \dots (x - x_{i-1});$$

$$p_{i\sigma}(x) = D_0 + D_1(x - x_i) + D_2(x - x_i)(x - x_{i-1}) + \dots + D_i(x - x_i)(x - x_{i-1}) \dots (x - x_1);$$

onde  $d_i$  e  $D_i$  são as diferenças divididas de ordem  $i$  com relação a cada uma das bases  $\mathcal{B}$  e  $\mathcal{B}_\sigma$ , respectivamente.

Como o polinômio interpolador é único (**Teorema 1**, Seção 4.1), então  $p_i(x) - p_{i\sigma}(x) = 0$  (polinômio nulo), para todo número real  $x$ . Dessa forma, o coeficiente de  $x^{i-1}$  no polinômio nulo deve ser igual a zero. De acordo com as definições dos polinômios  $p_i(x)$  e  $p_{i\sigma}(x)$ , as únicas parcelas destes polinômios que contribuirão para a análise do coeficiente de  $x^{i-1}$  no polinômio nulo são:

$$d_{i-1}(x - x_0)(x - x_1) \dots (x - x_{i-2}); \quad d_i(x - x_0)(x - x_1) \dots (x - x_{i-1});$$

$$D_{i-1}(x - x_i)(x - x_{i-1}) \dots (x - x_2); \quad D_i(x - x_i)(x - x_{i-1}) \dots (x - x_1).$$

Portanto, o coeficiente de  $x^{i-1}$  no polinômio nulo é dado por  $d_{i-1} - d_i(x_0 + x_1 + \dots + x_{i-1}) - [D_{i-1} - D_i(x_1 + x_2 + \dots + x_i)]$  (veja a **observação** anterior ao enunciado deste teorema). Como tal valor deve ser zero e, pelo **Teorema 2**,  $d_i = D_i$ , então:

$$d_{i-1} - d_i(x_0 + x_1 + \dots + x_{i-1}) - [D_{i-1} - D_i(x_1 + x_2 + \dots + x_i)] = 0 \rightarrow$$

$$\rightarrow d_{i-1} - D_{i-1} + D_i[x_1 + x_2 + \dots + x_i - (x_0 + x_1 + \dots + x_{i-1})] = 0 \rightarrow$$

$$\rightarrow d_{i-1} - D_{i-1} + D_i(x_i - x_0) = 0 \rightarrow D_i = [1/(x_i - x_0)]\{D_{i-1} - d_{i-1}\}.$$

De acordo com o **Teorema 2**, a ordem dos argumentos não altera o valor do operador de diferenças divididas. Assim,

$$f[x_0, x_1, x_2, \dots, x_i] = \frac{f[x_i, \dots, x_1] - f[x_0, \dots, x_{i-1}]}{(x_i - x_0)} = \frac{f[x_1, \dots, x_i] - f[x_0, \dots, x_{i-1}]}{(x_i - x_0)}.$$

■

**Observação:** A recursividade do polinômio de Newton é evidenciada no seguinte resultado: Sejam  $\mathcal{B}$  e  $\mathcal{B}'$  bases dos espaços vetoriais  $\mathcal{P}_n$  e  $\mathcal{P}_{n+1}$ , respectivamente, onde

$$\mathcal{B} = \{1, (x - x_0), (x - x_0)(x - x_1), (x - x_0)(x - x_1)(x - x_2), \dots, (x - x_0)(x - x_1) \dots (x - x_{n-1})\};$$

$$\mathcal{B}' = \{1, (x - x_0), (x - x_0)(x - x_1), (x - x_0)(x - x_1)(x - x_2), \dots, (x - x_0)(x - x_1) \dots (x - x_{n-1}), (x - x_0)(x - x_1) \dots (x - x_{n-1})(x - x_n)\}.$$

O polinômio interpolador de grau  $n$ , escrito na base  $\mathcal{B}$ , é dado por

$$p_n(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + \dots + d_n(x - x_0)(x - x_1)\dots(x - x_{n-1}).$$

O polinômio interpolador de grau  $n + 1$ , escrito na base  $\mathcal{B}'$ , é dado por

$$p_{n+1}(x) = D_0 + D_1(x - x_0) + D_2(x - x_0)(x - x_1) + \dots + D_{n+1}(x - x_0)(x - x_1)\dots(x - x_n),$$

onde  $D_i = d_i$ ,  $0 \leq i \leq n$ . Basta observar que o polinômio, de grau  $n$ ,  $P(x) = D_0 + D_1(x - x_0) + D_2(x - x_0)(x - x_1) + \dots + D_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$  satisfaz, para cada  $i$ ,  $0 \leq i \leq n$ ,  $P(x_i) = p_{n+1}(x_i) = f(x_i)$ . Como o polinômio interpolador é único, segue-se que  $P(x) = p_n(x)$ . Portanto,  $p_{n+1}(x) = p_n(x) + f[x_0, x_1, x_2, \dots, x_n, x_{n+1}]N(x)$ . ■

**Exemplo 8. (Tabela de Diferenças Divididas)** Vamos construir a tabela de diferenças divididas associada ao problema dado no **Exercício 4** (Seção 4.2). A partir da função tabelada:

$x$	0.5	0.6	0.8	0.9	1.1
$f(x)$	0.46128	0.53515	0.65767	0.70624	0.78006

serão construídas, recursivamente, as diferenças divididas de ordem 0 até a diferença dividida de ordem 4. Note que a tabela contém 5 pontos:  $x_0 = 0.5$ ;  $x_1 = 0.6$ ;  $x_2 = 0.8$ ;  $x_3 = 0.9$  e  $x_4 = 1.1$ . Os valores das diferenças divididas (DD) serão exibidos em colunas, como segue:

$x$	$f(x)$ : DD0	DD1	DD2	DD3	DD4
$x_0 = 0.5$	$f[x_0] = 0.46128$	$f[x_0, x_1]$	$f[x_0, x_1, x_2]$	$f[x_0, x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3, x_4]$
$x_1 = 0.6$	$f[x_1] = 0.53515$	$f[x_1, x_2]$	$f[x_1, x_2, x_3]$	$f[x_1, x_2, x_3, x_4]$	
$x_2 = 0.8$	$f[x_2] = 0.65767$	$f[x_2, x_3]$	$f[x_2, x_3, x_4]$		
$x_3 = 0.9$	$f[x_3] = 0.70624$	$f[x_3, x_4]$			
$x_4 = 1.1$	$f[x_4] = 0.78006$				

Os cálculos das diferenças divididas de ordem  $k$  (0, 1, 2, 3 e 4) serão efetuados de acordo com a fórmula:  $f[w_1, w_2, \dots, w_k, w_{k+1}] = \frac{f[w_2, \dots, w_{k+1}] - f[w_1, \dots, w_k]}{(w_{k+1} - w_1)}$ .

Portanto, a tabela é dada por

$x$	$f(x)$ : DD0	DD1	DD2	DD3	DD4
$x_0 = 0.5$	$f[x_0] = 0.46128$	0.7387	$-0.4203\dots$	$-0.00666666675$	0.12555...
$x_1 = 0.6$	$f[x_1] = 0.53515$	0.6126	$-0.423$	0.06866...	
$x_2 = 0.8$	$f[x_2] = 0.65767$	0.4857	$-0.3886\dots$		
$x_3 = 0.9$	$f[x_3] = 0.70624$	0.3691			
$x_4 = 1.1$	$f[x_4] = 0.78006$				

**Observação:** Dados os valores de diferenças divididas de uma coluna  $k - 1$ , os valores das diferenças divididas da coluna  $k$  são obtidos da seguinte forma: agrupa-se, de dois em dois, os valores da coluna  $k - 1$  (como mostra a seta azul na tabela anterior). Digamos que o primeiro valor do agrupamento seja  $v_1$  e o segundo  $v_2$  (na coluna  $k - 1$ ,  $v_1$  está acima de  $v_2$ ). Assim, o numerador da respectiva razão que está na coluna  $k$  é a diferença  $v_2 - v_1$ . O denominador dessa razão é a diferença entre os extremos do respectivo operador de diferenças divididas. Por exemplo, se a diferença dividida for  $f[x_2, x_3, x_4]$  (que é de ordem 2, pois tem 3 argumentos), então  $v_1 = 0.4857$  e  $v_2 = 0.3691$ . Portanto,  $f[x_2, x_3, x_4] = (0.3691 - 0.4857)/(x_4 - x_2) = 0.1166/0.3 = -0.3886\dots$ . Se a diferença dividida for  $f[x_0, x_1, x_2, x_3]$  (que é de ordem 3, pois tem 4 argumentos), então  $v_1 = -0.4203\dots$  e  $v_2 = -0.423$ . Portanto,  $f[x_0, x_1, x_2, x_3] = (-0.423 + 0.4203\dots)/(x_3 - x_0) = -0.002666\dots/0.4 = -0.00666666675$ . ■

**Exercício 5. (Polinômio de Newton)** Como no Exercício 4 (Seção 4.2), encontre uma aproximação para  $f(0.78)$  utilizando um polinômio de Newton de grau 2. Depois, obtenha outra estimativa de  $f$ , no mesmo ponto, utilizando um polinômio de grau 3.

**Solução do Exercício 5:**

O primeiro passo consiste na escolha dos pontos de interpolação. Como a função será aproximada no ponto  $x^* = 0.78$ , vamos escolher os pontos  $x_0, x_1$  e  $x_2$  que estiverem mais próximos de  $x^*$  de modo que o valor de  $|N(x^*)| = |x^* - x_0| \cdot |x^* - x_1| \cdot |x^* - x_2|$  seja o menor possível. De acordo com o Exercício 4 (Seção 4.2), a melhor opção é escolher  $x_0 = 0.6$ ,  $x_1 = 0.8$  e  $x_2 = 0.9$ .

Obtido os pontos de interpolação, o segundo passo consiste na construção do polinômio interpolador:

$$p_2(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1).$$

De acordo com a escolha dos pontos de interpolação, deve-se desprezar o primeiro valor de  $x$  na tabela original, ou seja, em vez de  $x_0 = 0.5$ , o primeiro ponto de interpolação será denotado por  $x_0 = 0.6$ . Dessa forma, a escolha dos valores das diferenças divididas de ordem 0, 1 e 2, que serão utilizados na construção do polinômio interpolador de grau 2, deverá seguir o mesmo princípio, ou seja, na coluna de diferenças divididas de ordem  $k$  (0, 1 e 2), o primeiro valor tem que ser desprezado e o segundo será o valor considerado para  $d_k$ . Dessa forma,  $d_0 = 0.53515$ ,  $d_1 = 0.6126$  e  $d_2 = -0.423$ . Portanto,  $f(0.78) \approx p_2(0.78) = d_0 + d_1(0.78 - x_0) + d_2(0.78 - x_0)(0.78 - x_1) = 0.6469408$ .

Para a construção do polinômio interpolador de grau 3, os pontos de interpolação devem ser escolhidos de modo que  $x_0, x_1, x_2$  e  $x_3$  estejam próximos de  $x^* = 0.78$ , no sentido de tornar mínimo o valor de  $|N(x^*)| = |x^* - x_0| \cdot |x^* - x_1| \cdot |x^* - x_2| \cdot |x^* - x_3|$ . As duas opções de pontos são: (i)  $x_0 = 0.5, x_1 = 0.6, x_2 = 0.8$  e  $x_3 = 0.9$ , ou (ii)  $x_0 = 0.6, x_1 = 0.8, x_2 = 0.9$  e  $x_3 = 1.1$ . Dessa forma, o menor valor de  $|N(x^*)|$  será obtido na primeira opção.

**Observação:** Nas duas opções anteriores os seguintes valores se repetem: 0.6, 0.8 e 0.9. O valor 0.5 aparece somente na primeira opção (i) e o valor 1.1 aparece somente na segunda opção (ii). Portanto, para saber qual a opção que fornecerá o menor valor de  $|N(x^*)|$ , basta verificar qual dos dois seguintes valores é o menor:  $|x^* - 0.5|$  ou  $|x^* - 1.1|$ . Como  $x^* = 0.78$  então a primeira opção é a melhor.

O primeiro ponto de interpolação coincide com o primeiro valor de  $x$  na tabela original. Portanto, as diferenças divididas utilizadas na construção do polinômio de grau 3 são dadas por:  $d_0 = 0.46128$ ,  $d_1 = 0.7387$ ,  $d_2 = -0.4203...$  e  $d_3 = -0.00666666675$ .

O polinômio interpolador é dado por:

$$p_3(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + d_3(x - x_0)(x - x_1)(x - x_2).$$

Assim,  $f(0.78) \approx p_3(0.78) = 0.64693792$ . ■

## Regra dos Parênteses Encaixados

O cálculo do valor numérico do polinômio de Newton com diferenças divididas pode ser feito de uma maneira recursiva, que exigirá menos esforço computacional do que a avaliação usual que, simplesmente, substitui a variável  $x$  por um número na expressão do do polinômio. Esta maneira recursiva é conhecida como a Regra dos Parênteses Encaixados.

Como exemplo, considere um polinômio de grau 3,

$$p_3(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1) + d_3(x - x_0)(x - x_1)(x - x_2).$$

Observe que este polinômio pode ser reescrito da seguinte forma:

$$p_3(x) = \left( d_0 + (x - x_0) \left( d_1 + (x - x_1) \left( d_2 + (x - x_2) (d_3) \right) \right) \right).$$

Rekursivamente, para se obter o valor numérico de  $p_3(x)$ , os cálculos que aparecem na expressão do polinômio são efetuados do final (expressão que está entre parênteses vermelhos) para o início (expressão que está entre parênteses verdes). Para isso, seja  $V_3 = d_3$  (valor que está entre parênteses marrons). Na primeira iteração, o valor de  $V_3$  será alterado para  $V_2$  (valor da expressão que está entre parênteses vermelhos), tal que  $V_2 = d_2 + (x - x_2)V_3$ ; na segunda iteração, o valor de  $V_2$  será alterado para  $V_1$  (valor da expressão que está entre parênteses azuis), tal que  $V_1 = d_1 + (x - x_1)V_2$ , na terceira e última iteração, o valor de  $V_1$  será alterado para  $V_0$  (valor da expressão que está entre parênteses verdes), tal que  $V_0 = d_0 + (x - x_0)V_1$ .

No caso geral, o seguinte algoritmo obtém o valor numérico de um polinômio de Newton com diferenças divididas de grau  $n$ .

$V = d_n$ ; (diferença dividida de ordem  $n$ )

para  $i$  variando de  $n - 1$  até zero calcule:

$V := d_i + (x - x_i).V$ ; ( $d_i$  é a diferença dividida de ordem  $i$ )

fim

O último valor de  $V$  (quando  $i = 0$ ) é o valor numérico de  $p_n(x)$ .

**Exemplo 9. (Regra dos Parênteses Encaixados)** Vamos utilizar a Regra dos Parênteses Encaixados para calcular o valor numérico do polinômio de grau 3 encontrado no **Exercício 5** que foi resolvido nesta seção.

Naquele exercício, os pontos de interpolação utilizados foram:  $x_0 = 0.5$ ,  $x_1 = 0.6$ ,  $x_2 = 0.8$  e  $x_3 = 0.9$ ; as diferenças divididas utilizadas foram:  $d_0 = 0.46128$ ,  $d_1 = 0.7387$ ,  $d_2 = -0.4203\dots$  e  $d_3 = -0.006666666675$ . O polinômio será avaliado em  $x = x^* = 0.78$ .

Fazendo as iterações com a Regra dos Parênteses Encaixados, obtemos:

(iteração 0)

$$V = d_3 = -0.006666666675;$$

(iteração 1)

$$V := d_2 + (x - x_2).V = -0.4203\dots + (0.78 - 0.8).(-0.006666666675) = -0.4202;$$

(iteração 2)

$$V := d_1 + (x - x_1).V = 0.7387 + (0.78 - 0.6).(-0.4202) = 0.663064;$$

(iteração 3)

$$V := d_0 + (x - x_0).V = 0.46128 + (0.78 - 0.5).(0.663064) = 0.64693792.$$

Portanto,  $p_3(0.78) = 0.64693792$ . ■

## 4.4. Erro de Interpolação e Estimativas para o Erro

Nesta seção será apresentada a fórmula do erro de interpolação e algumas estimativas desse erro. A demonstração do teorema do erro de interpolação vai utilizar um importante resultado de análise na reta, o Teorema de Rolle, que foi mencionado na Seção 3 do Módulo 1. Para mais informações, as referências apresentadas naquela seção podem ser consultadas.

O enunciado do Teorema de Rolle é o seguinte: “Seja  $g$  uma função real, contínua no intervalo fechado  $[a, b]$  e derivável no intervalo aberto  $(a, b)$ . Se  $g(a) = g(b)$  então existe  $c \in (a, b)$  tal que  $g'(c) = 0$ .”

**Teorema 4. (Erro de Interpolação)** Seja  $f: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  uma função real que possui derivadas até a ordem  $n + 1$ . Considere  $n + 2$  pontos distintos dois a dois  $x_0, x_1, \dots, x_n$  e  $x$  no intervalo fechado  $I$ . O erro,  $E^{(n)}(x)$ , cometido ao se aproximar  $f(x)$  por  $p_n(x)$  é dado por  $E^{(n)}(x) = f(x) - p_n(x) = \frac{N(x)f^{(n+1)}(\xi(x))}{(n+1)!}$ , onde  $N(x) = (x - x_0)(x - x_1) \dots$

$(x - x_n) = \prod_{i=0}^n (x - x_i)$  e  $\xi(x)$  pertence ao intervalo aberto  $(a, b)$ .

**Demonstração:** Considere a seguinte função  $G: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  dada por  $G(t) = f(t) - p_n(t) - [f(x) - p_n(x)](N(t)/N(x))$ . Note que a função  $G$  possui derivada até a ordem  $n + 1$ , porque a função  $f$  tem esta propriedade e  $N(t)$  é um polinômio de grau  $n + 1$ . Além disso, a função  $G$  possui  $n + 2$  zeros:  $G(x_i) = 0$ ,  $0 \leq i \leq n$ , e  $G(x) = 0$ . Aplicando-se sucessivamente o Teorema de Rolle às funções  $G, G', G'', G''', \dots, G^{(n)}$  tem-se que  $G^{(k)}$  tem pelo menos  $n + 2 - k$  zeros,  $1 \leq k \leq n + 1$ . Portanto, aplicando-se o Teorema de Rolle à função  $G^{(n)}$ , obtém-se  $c = \xi(x) \in (a, b)$  tal que  $G^{(n+1)}(c) = 0$ . Porém,  $0 = G^{(n+1)}(c) = f^{(n+1)}(c) - [f(x) - p_n(x)]((n + 1)!/N(x))$ . Portanto,

$$E^{(n)}(x) = f(x) - p_n(x) = \frac{N(x)f^{(n+1)}(\xi(x))}{(n+1)!}. \quad \square$$

**Observação:** A derivada de ordem  $n + 1$  de um polinômio de grau  $n$  é zero e a derivada de ordem  $n + 1$  do polinômio  $x^{n+1}$  é igual a  $(n + 1)!$ .

### 1ª Estimativa para o Erro

Em geral, o valor de  $\xi(x) \in (a, b)$  não é conhecido. Assim, assumindo-se que  $f^{(n+1)}$  é uma função contínua no intervalo fechado  $[a, b]$ , podemos fazer a seguinte estimativa:



$$|E^{(n)}(x)| \leq \frac{|N(x)| \max\{|f^{(n+1)}(t)|; t \in [a, b]\}}{(n+1)!}.$$

**Observação:** Toda função real contínua definida em um intervalo fechado e limitado possui valor máximo e valor mínimo (Veja as referências [1 – 5] exibidas na Seção 3 do Módulo 1).

## 2ª Estimativa para o Erro

Se os pontos de interpolação forem igualmente espaçados, ou seja,  $x_i = x_0 + ih$ ,  $1 \leq i \leq n$  e  $h > 0$ , e  $x \in [x_0, x_n]$ , então a limitação superior para o erro de interpolação é dada por

$$|E^{(n)}(x)| \leq \frac{h^{n+1} \max\{|f^{(n+1)}(t)|; t \in [x_0, x_n]\}}{4(n+1)}.$$

**Demonstração:** Primeiramente, observe que se os pontos de interpolação forem igualmente espaçados, ou seja,  $x_i = x_0 + ih$ ,  $1 \leq i \leq n$  e  $h > 0$ , então  $x \in [x_0, x_n] \leftrightarrow x =$

$x_0 + uh$ , com  $u \in [0, n]$ . Dessa forma,  $N(x) = N(x_0 + uh) = \prod_{i=0}^n (x_0 + uh - x_i) \rightarrow$

$\rightarrow N(x_0 + uh) = \prod_{i=0}^n h(u - i) = h^{n+1} u(u-1)(u-2) \dots (u-n)$ . Dessa forma,

$$\frac{|N(x)|}{(n+1)!} = h^{n+1} \frac{|u|}{1} \frac{|u-1|}{2} \frac{|u-2|}{3} \dots \frac{|u-n|}{n+1}, \text{ com } u \in [0, n]. \text{ Assim, usando a}$$

primeira estimativa, obtém-se que

$$|E^{(n)}(x)| \leq h^{n+1} \frac{|u|}{1} \frac{|u-1|}{2} \frac{|u-2|}{3} \dots \frac{|u-n|}{n+1} \max\{|f^{(n+1)}(t)|; t \in [x_0, x_n]\}.$$

Vamos analisar a expressão anterior para alguns valores de  $n$ .

Para  $n = 1$  tem-se que  $0 \leq u \leq 1$ . Assim,  $p(u) = |u| |u-1| = u(1-u) = u - u^2$ ;  $p'(u) = 1 - 2u$  e  $p''(u) = -2 < 0$ . Note que  $p'(u) = 0 \leftrightarrow u = 1/2$ . Portanto, o ponto crítico  $u = 1/2$  é ponto de máximo da função  $p$ ; logo  $p(u) = |u| |u-1| \leq p(1/2) = 1/4$ . Portanto,  $|E^{(1)}(x)| \leq \frac{h^2 \max\{|f^{(2)}(t)|; t \in [x_0, x_1]\}}{4(2)}.$

Para  $n = 2$  tem-se que  $0 \leq u \leq 2$ . Consideremos dois casos: (I)  $0 \leq u \leq 1$  e (II)  $1 < u \leq 2$ .

No primeiro caso, tem-se que  $p(u) = |u| |u - 1| \leq 1/4$ . Além disso,  $|u - 2|/2 \leq 1$ . De fato,  $0 \leq u \leq 1 \rightarrow -2 \leq u - 2 \leq -1 < 0 \rightarrow |u - 2| = 2 - u \leq 2 \rightarrow |u - 2|/2 \leq 1$ .

$$\text{Portanto, } |E^{(2)}(x)| \leq \frac{h^3 \max \{|f^{(3)}(t)|; t \in [x_0, x_2]\}}{4(3)}.$$

No segundo caso,  $1 < u \leq 2$ . Assim,  $p(u) = |u - 1| |u - 2| = (u - 1)(2 - u) = -u^2 + 3u - 2$ ;  $p'(u) = -2u + 3$  e  $p''(u) = -2 < 0$ . Note que  $p'(u) = 0 \leftrightarrow u = 3/2$ . Portanto, o ponto crítico  $u = 3/2$  é ponto de máximo da função  $p$ ; logo  $p(u) = |u - 1| |u - 2| \leq p(3/2) = 1/4$ . Além disso,  $|u|/2 \leq 1$ . De fato,  $1 < u \leq 2 \rightarrow |u| = u \leq 2 \rightarrow |u|/2 \leq 1$ . Portanto,

$$|E^{(2)}(x)| \leq \frac{h^3 \max \{|f^{(3)}(t)|; t \in [x_0, x_1]\}}{4(3)}.$$

Agora, considere a seguinte hipótese de indução:  $\frac{|u|}{1} \frac{|u-1|}{2} \frac{|u-2|}{3} \dots \frac{|u-n|}{n+1} \leq \frac{1}{4} \frac{1}{n+1}$ ,

com  $u \in [0, n]$ . Agora, considere  $u \in [0, n + 1]$ ; **(1º Caso)**  $0 \leq u \leq n \rightarrow |u - (n + 1)| = (n + 1) - u \leq (n + 1)$ ; portanto,  $|u - (n + 1)|/(n + 1) \leq 1$ . Pela hipótese de indução, segue-se que

$$\frac{|u|}{1} \frac{|u-1|}{2} \frac{|u-2|}{3} \dots \frac{|u-n|}{n+1} \frac{|u-(n+1)|}{n+2} \leq \frac{1}{4} \frac{1}{n+2}.$$

**(2º Caso)**  $n < u \leq n + 1 \rightarrow p(u) = |u - n| |u - (n + 1)| = (u - n)[(n + 1) - u] = -u^2 + (2n + 1)u - n(n + 1) \rightarrow p'(u) = -2u + (2n + 1)$  e  $p''(u) = -2 < 0$ . Note que  $p'(u) = 0 \leftrightarrow u = n + 1/2$ . Portanto, o ponto crítico  $u = n + 1/2$  é ponto de máximo da função  $p$ ; logo  $p(u) \leq p(n + 1/2) = 1/4$ . Além disso, se  $0 \leq i \leq n - 1$ , então  $\frac{|u-i|}{i+2} \leq \frac{n+1-i}{i+2}$ . Observe, também, que  $0 \leq n - 1 - i \leq n - 1$ ; logo:

$$\frac{|u-(n-1-i)|}{(n-1-i)+2} \leq \frac{n+1-(n-1-i)}{n-i+1} = \frac{i+2}{n+1-i}.$$

Dessa forma, se  $n$  for par, então, para cada  $i$ ,  $0 \leq i < n/2$ , existirá o par correspondente  $n - 1 - i$  tal que  $\frac{|u-i|}{i+2} \frac{|u-(n-1-i)|}{(n-1-i)+2} \leq \frac{n+1-i}{i+2} \frac{i+2}{n+1-i} = 1$ . Se  $n$  for ímpar, então  $i = (n - 1)/2$  satisfaz:  $i = n - 1 - i \leftrightarrow i + 2 = n + 1 - i$ . Nesse caso, tem-se que  $\frac{|u-i|}{i+2} \leq \frac{n+1-i}{i+2} = 1$ . Para os demais pares de pontos  $(i, n - 1 - i)$ ,  $0 \leq i \leq (n - 3)/2$ , obtemos a mesma desigualdade exibida no início deste parágrafo. Portanto,

$$\frac{|u|}{1} \frac{|u-1|}{2} \frac{|u-2|}{3} \dots \frac{|u-n|}{n+1} \frac{|u-(n+1)|}{n+2} \leq \frac{1}{4} \frac{1}{n+2}.$$

Conclusão: por indução finita, tem-se que

$$|E^{(n)}(x)| \leq \frac{h^{n+1} \max\{|f^{(n+1)}(t)|; t \in [x_0, x_n]\}}{4(n+1)}.$$

### 3ª Estimativa para o Erro

Caso a expressão da função  $f$  não seja conhecida ou seja muito complexa, então podemos utilizar apenas a tabela de diferenças divididas para obter uma estimativa do erro de interpolação. Nesse caso, vale a seguinte estimativa:

$$|E^{(n)}(x)| \approx |N(x)| \max\{|DD_{n+1}|\},$$

onde  $|DD_{n+1}|$  é o valor absoluto de uma diferença dividida de ordem  $n + 1$ .

Esta estimativa é proveniente da observação feita após a demonstração do **Teorema 3** da Seção 4.3 e da primeira estimativa apresentada anteriormente. Dessa forma, sabendo-se que  $p_{n+1}(x) = p_n(x) + f[x_0, x_1, x_2, \dots, x_n, x_{n+1}]N(x)$ , suponha que  $x_{n+1} = x$ . Então  $f(x) = p_{n+1}(x) = p_n(x) + DD_{n+1}N(x)$ . Assim,  $E^{(n)}(x) = DD_{n+1}N(x)$ . Por outro lado, utilizando o **Teorema 4**, tem-se que  $\frac{N(x)f^{(n+1)}(\xi(x))}{(n+1)!} = E^{(n)}(x) = DD_{n+1}N(x)$ .

Portanto,  $\frac{f^{(n+1)}(\xi(x))}{(n+1)!} = DD_{n+1} = f[x_0, x_1, x_2, \dots, x_n, x]$ .

**Exemplo 10. (Estimativas do Erro de Interpolação)** Neste exemplo, nós iremos aproximar uma função conhecida por um polinômio interpolador de grau 2 e depois apresentaremos algumas estimativas do erro cometido. Os resultados serão comparados com o erro efetivo. O exemplo foi adaptado de um exercício do livro da professora Neide Bertoldi Franco (veja a referência [4] apresentada na Seção 4.1 do Módulo 2). A primeira vez que resolvi este exercício foi a pedido da professora Rosana Sueli da Motta Jafelice, que estava ministrando a disciplina de Cálculo Numérico para uma turma (se não me falha a memória) da Engenharia Química da Universidade Federal de Uberlândia. A professora estranhara que os resultados teóricos não eram compatíveis com os resultados numéricos. A meditação foi o caminho para refutar o paradoxo!

A função  $f(x) = \sqrt{x}$  está tabelada abaixo. A aproximação será feita em  $x^* = 1.12$ .

$x$	1.00	1.10	1.15	1.25
$f(x)$	1.00	1.048	1.072	1.118

Montando-se a tabela de diferenças divididas, obtemos:

$x$	$f(x)$ : DD0	DD1	DD2	DD3
$x_0 = 1.00$	$f[x_0] = 1.00$	0.48	0	-0.533...
$x_1 = 1.10$	$f[x_1] = 1.048$	0.48	-0.133...	
$x_2 = 1.15$	$f[x_2] = 1.072$	0.46		
$x_3 = 1.25$	$f[x_3] = 1.118$			

Vamos escolher os pontos  $x_0, x_1$  e  $x_2$  que estiverem mais próximos de  $x^* = 1.12$  de modo que o valor de  $|N(x^*)| = |x^* - x_0| \cdot |x^* - x_1| \cdot |x^* - x_2|$  seja o menor possível.

Seguindo o mesmo procedimento recomendado na observação que foi feita no **Exercício 5**, Seção 4.3, os pontos de interpolação são dados por:  $x_0 = 1.00$ ;  $x_1 = 1.10$  e  $x_2 = 1.15$  e  $|N(x^*)| = 0.000072$ . Assim, serão consideradas as seguintes diferenças divididas:  $d_0 = 1.00$ ,  $d_1 = 0.48$  e  $d_2 = 0$ .

Nesse caso, o polinômio interpolador é  $p_2(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)(x - x_1)$ . Como  $d_2 = 0$ , então o polinômio terá grau 1:  $p_2(x) = 1 + 0.48(x - 1.0)$ . Desse modo, a aproximação desejada é  $\sqrt{1.12} \approx p_2(1.12) = 1.0576$ . O erro efetivo é  $\sqrt{1.12} - 1.0576 \approx 0.70 \times 10^{-3}$ .

Observe que, se fossem escolhidos os pontos  $x_0 = 1.10$ ;  $x_1 = 1.15$  e  $x_2 = 1.25$ , obteríamos o polinômio  $p_2(x) = 1.048 + 0.48(x - 1.10) - (x - 1.10)(x - 1.15)0.133...$ . Nesse caso,  $|N(x^*)| = 0.000078$  e a aproximação seria dada por  $\sqrt{1.12} \approx p_2(1.12) = 1.05768$ . O erro efetivo é  $\sqrt{1.12} - 1.05768 \approx 0.62 \times 10^{-3}$ .

Esse é o primeiro resultado estranho. A escolha dos pontos foi feita de modo a obter o menor valor para o erro de interpolação. Porém, como mostrado anteriormente, a primeira aproximação (gerada pelo polinômio de grau 1) produziu um erro maior do que o erro produzido pela segunda aproximação, que foi gerada por um polinômio de grau 2. Por que isso ocorreu? Será que os cálculos efetuados na construção da tabela de diferenças divididas, e que geraram  $d_2 = 0$ , na primeira escolha dos pontos, estão corretos?

Sabendo-se que  $f^{(3)}(t) = (3/8)t^{-5/2}$  é uma função decrescente então  $\max\{|f^{(3)}(t)|, t \in [1, 1.25]\} = 3/8$ . Considerando-se  $x = 1.12$ ,  $N(t) = (t - x_0)(t - x_1)(t - x_2)$  e os pontos de interpolação escolhidos em cada uma das opções anteriores, segue da teoria do erro de interpolação que  $|E^{(2)}(x)| \leq \frac{|N(x)| \max\{|f^{(3)}(t)|; t \in [a, b]\}}{3!}$ . Portanto, na primeira opção,  $|E^{(2)}(x)| \leq 0.000072 \times (3/8) \times (1/6) = 0.45 \times 10^{-5}$  e, na segunda opção,  $|E^{(2)}(x)| \leq 0.000078 \times (3/8) \times (1/6) = 0.4875 \times 10^{-5}$ .

Observe, também, que  $|E^{(2)}(x)| \approx |N(x)| \max\{|DD_3|\} = |N(x)|0.533\dots$ . Assim, na primeira opção,  $|E^{(2)}(x)| \approx 0.000072 \times 0.533\dots = 0.384 \times 10^{-4}$  e, na segunda opção,  $|E^{(2)}(x)| \approx 0.000078 \times 0.533\dots = 0.416 \times 10^{-4}$ .

Esse é o segundo resultado estranho na resolução do exercício. A teoria nos diz que o valor absoluto do erro de interpolação deveria ser inferior a  $10^{-5}$ , porém os cálculos efetuados mostram que as estimativas obtidas, mesmo as referentes ao erro efetivo, são da ordem de  $10^{-3}$  ou  $10^{-4}$  e, portanto, superiores às estimativas previstas na teoria.

Depois de muito analisar o problema, descobri que todos os resultados aparentemente confusos foram gerados pela pouca precisão dos pontos da tabela. Note que todos os valores da função tabelada possuem apenas 3 casas decimais depois da vírgula. Seria muito surpreendente se conseguíssemos uma aproximação com precisão de 5 casas decimais com esses dados! ■

**Exercício 6. (Interpolação Linear com pontos igualmente espaçados).** A tabela abaixo exhibe os valores da função  $f(x) = x^5 \ln(x)$  em pontos igualmente espaçados. Mostre que a interpolação linear, nesse caso, sempre irá gerar erros inferiores a  $10^{-3}$ .

$x$	0.1	0.2	0.3	0.4	0.5
$y = f(x)$	$-0.23025851 \times 10^{-4}$	$-5.15020132 \times 10^{-4}$	$-29.25653915 \times 10^{-4}$	$-93.82817094 \times 10^{-4}$	$-216.60849 \times 10^{-4}$

### Solução do Exercício 6:

De acordo com a fórmula da estimativa de erros com pontos igualmente espaçados, tem-se que  $|E^{(1)}(x)| \leq \frac{h^2 \max\{|f^{(2)}(t)|; t \in [0.1, 0.5]\}}{4(2)}$ .

Note que  $h = 0.1$ ;  $f^{(1)}(x) = x^4[1 + 5\ln(x)]$  e  $f^{(2)}(x) = x^3[9 + 20\ln(x)]$ . Os pontos críticos  $x = x_c$  de  $f^{(2)}(x)$  são tais que:

$f^{(3)}(x)$  não existe ou  $f^{(3)}(x) = 0 \leftrightarrow x^2[47 + 60\ln(x)] = 0 \leftrightarrow x_c = 0$  ou  $x_c = \exp(-47/60) = 0.456880535$ . Portanto,

$\max\{|f^{(2)}(t)|, t \in [0.1, 0.5]\} = \max\{|f^{(2)}(x_c)|, |f^{(2)}(0.1)|, |f^{(2)}(0.5)|\} = |f^{(2)}(0.456880535)| = 0.635794414$ .

Dessa forma,  $|E^{(1)}(x)| \leq (0.1)^2 \times 0.635794414 \times 0.125 \approx 0.7947 \times 10^{-3}$ .

Por exemplo, utilizando o polinômio de Lagrange sobre os pontos de interpolação  $x_0 = 0.4$  e  $x_1 = 0.5$ , obtemos  $p_1(0.45) = -0.015521833$  (Verifique!). Como  $f(0.45) = -0.014734712$  então  $f(0.45) - p_1(0.45) \approx 0.7871 \times 10^{-3}$ . ■

## 4.5. Interpolação Inversa

Um importante teorema sobre função inversa, consequência do Teorema do Valor Intermediário (**Teorema 3**, Seção 3 do Módulo 1), provavelmente comentado em disciplinas de Cálculo Diferencial e demonstrado em um curso de Análise na Reta, diz que toda função  $f$  real, contínua e injetora, definida em um intervalo  $I$ , é monótona (crescente ou decrescente) e tem como imagem um intervalo  $J = f(I)$ . Além disso, a função inversa,  $f^{-1}: J \rightarrow I$  é uma função contínua.

Em determinados problemas, precisamos usar um polinômio interpolador para aproximar a inversa de uma dada função. Acontece que, em muitos casos, a função pode ser dada por meio de uma tabela. Assim, como saber se a função representada por uma tabela possui inversa? É simples. Basta supor que a função tabelada seja contínua e considerar apenas os pontos da tabela nos quais podemos afirmar que a função é monótona (crescente ou decrescente). A interpolação inversa só terá validade no intervalo onde a função é estritamente crescente ou estritamente decrescente.

**Exercício 7. (Interpolação Inversa).** Dada a tabela

$x$	1	2	3	4	5	6
$y = f(x)$	0.841	0.909	0.141	-0.757	-0.959	-0.279

obtenha  $x^*$  tal que  $f(x^*) = 0$ . Use um polinômio de grau 2 e apresente uma estimativa do erro.

**Solução do Exercício 7:**

(i) Escolha dos pontos da interpolação inversa.

Como o problema exige que seja feita uma estimativa do erro de interpolação, então o polinômio interpolador deve ser o de Newton. A análise do erro vai precisar das diferenças divididas de ordem 3, pois será utilizado um polinômio interpolador de grau 2. Assim, os pontos escolhidos para a interpolação inversa são: 0.909; 0.141;  $-0.757$  e  $-0.959$ , correspondentes a valores decrescentes da função tabelada.

(ii) Construção da Tabela de Diferenças Divididas.

Tabela de diferenças divididas ( $g = f^{-1}(y)$ )

$y$	$x = g(y)$ : DD0	DD1	DD2	DD3
$y_0 = 0.909$	$g[y_0] = 2$	-1.3020833...	- 0.113143809	- 1.927860373
$y_1 = 0.141$	$g[y_1] = 3$	- 1.113585746	3.488099367	
$y_2 = - 0.757$	$g[y_2] = 4$	- 4.95049505		
$y_3 = - 0.959$	$g[y_3] = 5$			

(iii) Escolha dos pontos que serão utilizados na construção do polinômio de grau 2.

Dado  $y^* = 0$ , a melhor escolha dos pontos é aquela que fornecerá o menor valor para  $|N(y^*)| = |y^* - y_0| \cdot |y^* - y_1| \cdot |y^* - y_2|$ . Nesse caso, os pontos escolhidos são:  $y_0 = 0.909$ ;  $y_1 = 0.141$  e  $y_2 = - 0.757$ .

(iv) Construção do polinômio de grau 2.

$$p_2(y) = d_0 + d_1(y - y_0) + d_2(y - y_0)(y - y_1);$$

onde

$$y_0 = 0.909; y_1 = 0.141; y_2 = - 0.757; d_0 = 2; d_1 = - 1.3020833...; d_2 = - 0.113143809.$$

(v) Cálculo do valor numérico do polinômio de grau 2.

$$p_2(0) = 3.169092221 = x^*.$$

(vi) Estimativa do erro.

$$|E^{(2)}(0)| \approx |0 - y_0| |0 - y_1| |0 - y_2| \max\{|DD3|\} = 0.909 \times 0.141 \times 0.757 \times 1.927860373.$$

Portanto,  $|E^{(2)}(0)| \approx 0.187048595$ . ■

## 5. Integração Numérica

Ao aproximar uma função  $f$  por um polinômio de grau  $n$ ,  $p_n(x)$ , será cometido um erro de interpolação. De acordo com as hipóteses do **Teorema 4** (Seção 4.4),  $f(x) = p_n(x) +$

$E^{(n)}(x)$ , onde  $E^{(n)}(x) = \frac{N(x)f^{(n+1)}(\xi(x))}{(n+1)!}$ . Então a integral de  $f$  pode ser escrita como:

$\int_a^b f(x) dx = \int_a^b p_n(x) dx + \int_a^b E^{(n)}(x) dx$ . Nesse caso, dizemos que  $\int_a^b E^{(n)}(x) dx$  é o erro de integração.

## 5.1 Regra do Trapézio: Simples e Repetida

Na Regra do Trapézio, a função  $f: [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  será aproximada pelo polinômio de Lagrange de grau 1, sobre os pontos de interpolação  $x_0 = a$  e  $x_1 = b$ . Seja  $h = x_1 - x_0$  e suponha que  $f$  tenha derivada segunda contínua em  $[a, b]$ . Dessa forma,

$$\int_a^b f(x) dx = \int_{x_0}^{x_1} p_1(x) dx + \int_{x_0}^{x_1} \frac{(x-x_0)(x-x_1)f^{(2)}(\xi(x))}{2!} dx,$$

onde  $\xi(x) \in (a, b)$ ,  $\forall x \in [a, b]$ ; portanto,  $f^{(2)}(x_{\min}) \leq f^{(2)}(\xi(x)) \leq f^{(2)}(x_{\max})$  (lembre-se de que toda função contínua definida em um intervalo fechado assume valor mínimo e valor máximo – referências [1 – 5] apresentadas na Seção 3, Módulo 1).

A Regra do Trapézio,  $I_T$ , será obtida da integral de  $p_1(x) = f(x_0)L_0(x) + f(x_1)L_1(x)$ , onde  $L_0(x) = \frac{(x-x_1)}{(x_0-x_1)}$  ;  $L_1(x) = \frac{(x-x_0)}{(x_1-x_0)}$ . É usual utilizar a seguinte troca de variáveis:

$x = x_0 + uh$ . Assim,  $\int_{x_0}^{x_1} p_1(x) dx = \int_0^1 p_1(x_0+uh) h du$ . Note que  $p_1(x_0 + uh) =$

$f(x_0)L_0(x_0 + uh) + f(x_1)L_1(x_0 + uh) = f(x_0) \frac{(x_0+uh-x_1)}{-h} + f(x_1) \frac{(x_0+uh-x_0)}{h}$ . Assim,

$$\int_{x_0}^{x_1} p_1(x) dx = -f(x_0) \int_0^1 (u-1) h du + f(x_1) \int_0^1 uh du. \text{ Agora, observe que}$$

$$\int_0^1 (u-1) h du = \left[ h \frac{(u-1)^2}{2} \right]_0^1 = -h/2; \quad \int_0^1 uh du = \left[ h \frac{u^2}{2} \right]_0^1 = h/2.$$

Portanto,  $I_T = \int_{x_0}^{x_1} p_1(x) dx = \frac{h}{2} [f(x_0) + f(x_1)]$ .

Para se deduzir a fórmula do erro, note que  $N(x) = (x-x_0)(x-x_1) < 0$ ,  $\forall x \in (x_0, x_1)$ . Dessa forma,  $N(x)f^{(2)}(x_{\max}) \leq N(x)f^{(2)}(\xi(x)) \leq N(x)f^{(2)}(x_{\min})$ . Utilizando as propriedades das funções integráveis, segue-se que



$$\int_{x_0}^{x_1} N(x) f^{(2)}(x_{max}) dx \leq \int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx \leq \int_{x_0}^{x_1} N(x) f^{(2)}(x_{min}) dx.$$

Utilizando o Teorema do Valor Intermediário, pode-se mostrar que existe  $c \in (x_0, x_1)$  tal que  $\int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx = f^{(2)}(c) \int_{x_0}^{x_1} N(x) dx$ . Veja as observações a seguir.

**Observação:** O enunciado da versão do Teorema do Valor Intermediário que será utilizada nesta seção é a seguinte: “Seja  $g: [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  uma função contínua. Se  $g(a) < \lambda < g(b)$ , ou  $g(b) < \lambda < g(a)$ , então existirá  $c \in (a, b)$  tal que  $g(c) = \lambda$ .” Para obter a versão apresentada no módulo 1, basta considerar  $F(x) = g(x) - \lambda$ ; logo  $F$  será contínua e  $F(a)F(b) < 0$ .

**Observação:** Se  $\int_{x_0}^{x_1} N(x) dx = 0$  então, da última desigualdade anterior, segue que  $\int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx = 0$ . Portanto, para qualquer  $c \in (x_0, x_1)$ , tem-se que  $\int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx = f^{(2)}(c) \int_{x_0}^{x_1} N(x) dx$ .

**Observação:**  $\int_{x_0}^{x_1} N(x) dx < 0 \rightarrow f^{(2)}(x_{min}) \leq \frac{\int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx}{\int_{x_0}^{x_1} N(x) dx} \leq f^{(2)}(x_{max})$ .

Portanto, pelo Teorema do Valor Intermediário, existe  $c \in (x_0, x_1)$  tal que

$$\lambda = \frac{\int_{x_0}^{x_1} N(x) f^{(2)}(\xi(x)) dx}{\int_{x_0}^{x_1} N(x) dx} = f^{(2)}(c).$$

**Observação:** Suponha que  $f^{(2)}(x)$  não seja constante e que existam  $a_1$  e  $b_1$  em  $[a, b]$  tais que  $f^{(2)}(\xi_1) = f^{(2)}(x_{min})$  e  $f^{(2)}(\xi_2) = f^{(2)}(x_{max})$ , onde  $\xi_1 = \xi(a_1) \in (a, b)$  e  $\xi_2 = \xi(b_1) \in (a, b)$ . Sabe-se que  $\xi(x) \in (a, b)$ ,  $\forall x \in [a, b]$ , e  $f^{(2)}(x_{min}) \leq f^{(2)}(\xi(x)) \leq f^{(2)}(x_{max})$ . Então, da observação anterior, segue-se que  $f^{(2)}(\xi_1) \leq \lambda \leq f^{(2)}(\xi_2)$ . Do Teorema do Valor intermediário, existe  $c \in [\xi_1, \xi_2]$  ou  $c \in [\xi_2, \xi_1]$ , logo  $c \in (a, b)$ , tal que  $f^{(2)}(c) = \lambda$ . As demais desigualdades:  $f^{(2)}(x_{min}) < \lambda < f^{(2)}(x_{max})$ ;  $f^{(2)}(\xi_1) \leq \lambda < f^{(2)}(x_{max})$  e  $f^{(2)}(x_{min}) < \lambda \leq f^{(2)}(\xi_2)$  conduzirão ao mesmo resultado.

**Observação:** Se  $f^{(2)}(x)$  for constante então  $f^{(2)}(c) = \lambda$ , qualquer que seja  $c \in (a, b)$ .

**Observação:** Seguindo o mesmo procedimento adotado nas demonstrações anteriores e utilizando a versão do Teorema do Valor Intermediário exibido na primeira observação da Seção 5.1, pode-se demonstrar o seguinte resultado mais geral:

$\int_a^b \Omega(x)g(x) dx = g(c) \int_a^b \Omega(x) dx$ , onde  $g: [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  é uma função contínua e  $\Omega(x)$  é uma função integrável com  $\Omega(x) \geq 0$  (ou  $\Omega(x) \leq 0$ ) para todo  $x \in [a, b]$ .

Voltando à expressão do erro de integração da Regra do Trapézio, tem-se que

$$E_T = \int_{x_0}^{x_1} \frac{(x-x_0)(x-x_1)f^{(2)}(\xi(x))}{2!} dx = \frac{f^{(2)}(c)}{2} \int_{x_0}^{x_1} (x-x_0)(x-x_1) dx.$$

Utilizando, novamente, a troca de variáveis:  $x = x_0 + uh$ , obtém-se que

$$E_T = \frac{f^{(2)}(c)}{2} \int_0^1 (uh)(u-1)h hdu = \frac{h^3 f^{(2)}(c)}{2} \left[ \frac{u^3}{3} - \frac{u^2}{2} \right]_0^1 = -\frac{h^3 f^{(2)}(c)}{12},$$

onde  $c \in (a, b)$ .

**Observação:** Da fórmula do erro de integração anterior, segue-se que a Regra do Trapézio é exata (possui erro nulo) para funções constantes e para polinômios de grau um; porque, nesses casos, a derivada segunda de  $f$  é igual a zero.

## A Regra Repetida do Trapézio

A Regra Repetida do Trapézio consiste em dividir o intervalo  $[a, b]$  em  $m$  subintervalos de mesmo comprimento  $h = (b-a)/m$  e aproximar a função  $f$  por um polinômio de Lagrange de grau 1,  $p_1^{(i)}(x)$ , em cada subintervalo  $[x_{i-1}, x_i]$ , onde  $x_i = x_0 + ih$ ,  $1 \leq i \leq m$ ;  $x_0 = a$  e  $x_m = b$ .

Dessa forma,

$$\int_a^b f(x) dx = \sum_{i=1}^m \int_{x_{i-1}}^{x_i} f(x) dx \approx \sum_{i=1}^m \int_{x_{i-1}}^{x_i} p_1^{(i)}(x) dx = \sum_{i=1}^m \frac{h}{2} [f(x_{i-1}) + f(x_i)].$$

**Observação:** a integral do polinômio  $p_1^{(i)}(x)$  é deduzida de forma análoga à dedução feita para a Regra Simples do Trapézio. Basta trocar  $x_0$  por  $x_{i-1}$  e  $x_1$  por  $x_i$ .

A Regra Repetida do Trapézio pode ser reescrita como:

$$I_{TR} = (h/2) \{f(x_0) + 2 \sum_{i=1}^{m-1} f(x_i) + f(x_m)\}.$$

De acordo com a fórmula do erro para a Regra Simples do Trapézio, o erro de integração, em cada intervalo  $[x_{i-1}, x_i]$ , será dado por  $-\frac{h^3 f^{(2)}(c_i)}{12}$ , onde  $c_i \in (x_{i-1}, x_i)$ ,  $1 \leq i \leq m$ . Portanto, o erro da Regra Repetida do Trapézio é dado por

$$E_{TR} = \sum_{i=1}^m -\frac{h^3 f^{(2)}(c_i)}{12} = -\frac{h^3}{12} \sum_{i=1}^m f^{(2)}(c_i) = -\frac{mh^3 f^{(2)}(\tilde{c})}{12}, \text{ onde } \tilde{c} \in (a, b).$$

**Observação:** A última igualdade na fórmula do erro da Regra Repetida do Trapézio é consequência do Teorema do Valor Intermediário. De fato,

$$mf^{(2)}(x_{\min}) \leq \sum_{i=1}^m f^{(2)}(c_i) \leq mf^{(2)}(x_{\max}) \rightarrow f^{(2)}(x_{\min}) \leq \frac{1}{m} \sum_{i=1}^m f^{(2)}(c_i) \leq f^{(2)}(x_{\max}).$$

## Estimativa do erro na Regra dos Trapézios

$$\text{Note que } |E_{TR}| = \frac{mh^3 |f^{(2)}(\tilde{c})|}{12} \leq \frac{mh^3}{12} \max \{|f^{(2)}(x)|; x \in [a, b]\}.$$

Esta estimativa do erro será utilizada em um problema apresentado a seguir. A história deste problema começa em 2004, quando eu era o tutor do Programa de Educação Tutorial (PET) da Matemática – Universidade Federal de Uberlândia (UFU).

Naquela ocasião, eu recomendei aos integrantes do PET-Matemática que submetessem trabalhos para a Semana Acadêmica da UFU. Aconselhei aos alunos que, quando fossem elaborar os seus resumos, priorizassem temas relacionados a aplicações da matemática e à disciplina de Cálculo Numérico. Estávamos dando continuidade ao projeto “A Matemática é Boa Temática”, que fora iniciado em 2003 com a realização de um concurso (com o mesmo título anterior) que foi vencido pela aluna da Faculdade de Matemática (FAMAT) da UFU, Giovana Trindade, com o projeto interdisciplinar: “A Biomatemática aplicada à herança da cor da pele no homem”. A premiação fora feita no dia 27 de novembro de 2003, no anfiteatro do bloco B, na UFU, durante o coquetel de encerramento da Terceira Semana de Matemática da FAMAT.

Os trabalhos produzidos, em 2004, pelos alunos do PET, Carolina Fernandes Molina Sanches, Éliton Meireles de Moura, Fabiana Alves Calazans, Flaviano Bahia Paulinelli Vieira, Gisliane Alves Pereira, Jairo Menezes e Sousa, Laís Bassame Rodrigues, Leandro Cruvinel Lemes, Maksuel Andrade Costa, Rafael Peixoto, Sandreane Poliana Silva e Wagner Frassetto, ainda estão guardados juntamente com outros documentos da

época em que fui tutor do PET (de agosto de 2001 a dezembro de 2005). Ótimas recordações!

Por envolver integração numérica, o problema apresentado pela aluna Gisliane Alves Pereira, em 2004, na Semana Acadêmica da UFU, foi selecionado para compor este texto (infelizmente, ao abrir o arquivo com a apresentação do problema, verifiquei que a aluna se esqueceu de incluir as referências bibliográficas!).

**Problema 2. (Regra dos Trapézios)** O Serviço de Proteção ao Consumidor (SPC) tem recebido muitas reclamações quanto ao peso real do pacote de 5 Kg de açúcar vendido nos supermercados. Para verificar a validade das reclamações, o SPC contratou uma firma especializada em estatística, que se dispôs a fazer uma estimativa da quantidade de pacotes que realmente continham menos de 5 Kg. Como é inviável a repesagem de todos os pacotes postos à venda, a firma responsável pesou uma amostra de 100 pacotes

e concluiu que o peso médio dos pacotes de açúcar era igual a  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 4.991 \text{ Kg}$

e o desvio padrão era  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2} = 0.005 \text{ Kg}$  (Os pesos  $x_i$  dos pacotes da

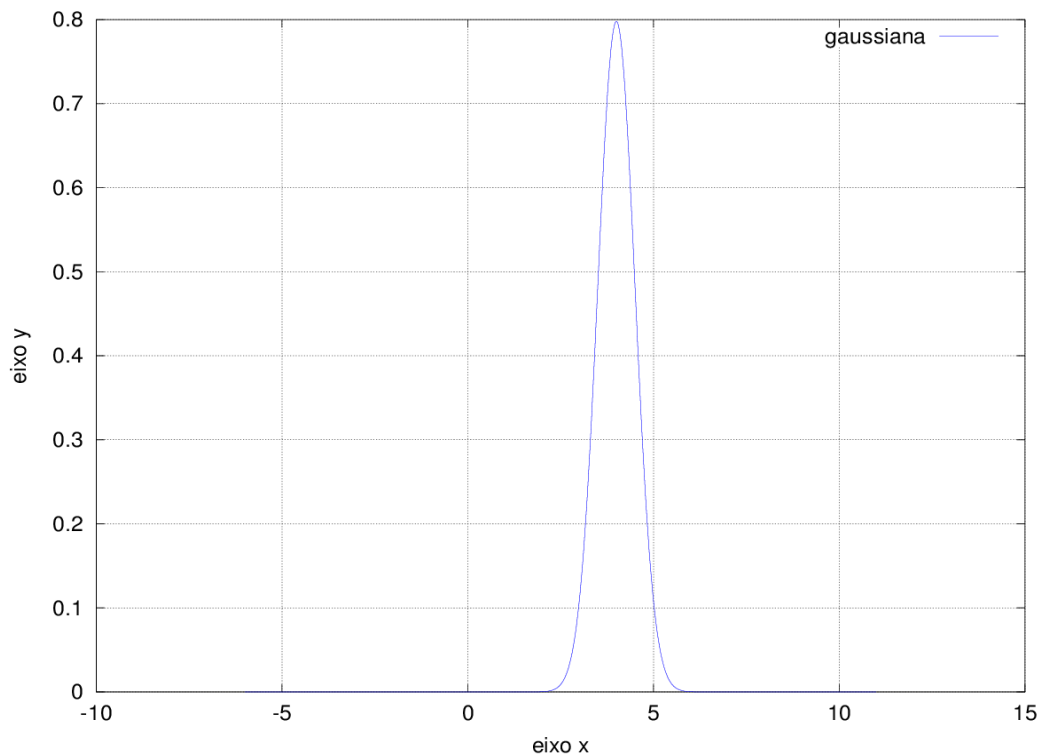
amostra não foram disponibilizados). Considere que a variável peso tenha distribuição normal e determine qual a probabilidade de se achar, no mercado, um pacote de açúcar com peso abaixo do peso nominal (com peso inferior a 5 Kg). Observação: A integração numérica deve ter erro inferior a  $10^{-6}$ .

**Solução do Problema 2:** Lembre-se de que uma distribuição normal tem expressão

dada por  $f(x) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{S}\right)^2}$ . O gráfico de  $f(x)$  está exibido na **Figura 4**. Note que

este gráfico é simétrico em relação à média ( $\bar{x}$ ). Como  $\int_{-\infty}^{\infty} f(x) dx = 1$  então

$$\int_{-\infty}^{\bar{x}} f(x) dx = \int_{\bar{x}}^{\infty} f(x) dx = 0,5.$$



**Figura 4 – Módulo 3: Gráfico da função Gaussiana  $f(x)$**

A integral de  $f(x)$  dá a frequência acumulada, ou seja,  $F(x_0) = \int_{-\infty}^{x_0} f(x) dx$  é a probabilidade de que  $x$  assumo um valor menor do que ou igual a  $x_0$ . No problema em questão, o que se deseja é determinar  $F(5) = \int_{-\infty}^5 \frac{1}{0,005\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-4,991}{0,005}\right)^2} dx$ .

Tendo em vista o que foi exposto anteriormente, tem-se que  $F(5) = 0,5 + \int_{\bar{x}}^5 \frac{1}{0,005\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-4,991}{0,005}\right)^2} dx$ . Fazendo a mudança de variável:  $z = \frac{x - \bar{x}}{S} = \frac{x - 4,991}{0,005}$  tem-se que  $F(5) = 0,5 + \frac{1}{\sqrt{2\pi}} \int_0^{1,8} e^{\frac{-1}{2}z^2} dz$ .

Para que o valor da integral anterior tenha erro inferior a  $10^{-6}$ , é necessário calcular o número de subintervalos,  $m$ , do intervalo  $[a, b] = [0, 1.8]$  de modo que

$$|E_{TR}| = \frac{mh^3 |f^{(2)}(\tilde{c})|}{12} \leq \frac{mh^3}{12} \max \{|f^{(2)}(x)|; x \in [a, b]\} < 10^{-6}.$$

Observe que

$$f(z) = \exp(-0.5z^2) \rightarrow f'(z) = -z \cdot \exp(-0.5z^2) \rightarrow f^{(2)}(z) = (z^2 - 1) \cdot \exp(-0.5z^2) \rightarrow$$

$\rightarrow f^{(3)}(z) = (3z - z^3) \cdot \exp(-0.5z^2)$ . Portanto,  $f^{(3)}(z) = 0$  se, e somente se,  $z = 0$  ou  $z = 3^{1/2}$  ou  $z = -3^{1/2}$ . Assim,

$$\max\{|f^{(2)}(z)|; z \in [0, 1.8]\} = \max\{|f^{(2)}(0)|; |f^{(2)}(1.8)|; |f^{(2)}(-3^{1/2})|; |f^{(2)}(3^{1/2})|\} = 1.$$

Como  $h = (1.8 - 0)/m$ , segue-se que

$$\frac{mh^3}{12} \max\{|f^{(2)}(x)|; x \in [0, 1.8]\} = m^{-2} (1.8)^3 (1/12) < 10^{-6} \rightarrow m^2 > 0.486 \times 10^6 \rightarrow$$

$$m > 697.1370023 \rightarrow m = 698.$$

O valor aproximado da integral, utilizando a Regra Repetida do Trapézio com 698 subintervalos, é dado por 1.16324999. Dessa forma,  $F(5) = 0.964069603$ .

Através do resultado obtido anteriormente, pode-se concluir que existe uma probabilidade de 0,9641 ou 96,41% de se achar um pacote de açúcar com menos de 5 Kg, ou seja, 96,41% dos pacotes vendidos nos mercados estão com peso abaixo do peso nominal. ■

**Observação:** O código computacional com a Regra dos Trapézios, desenvolvido em Octave, será exibido a seguir.

%Regra dos Trapézios

clear

%a=input('de o extremo inferior do intervalo de integracao: a=');

%b=input('de o extremo superior do intervalo de integracao: b=');

a = 0;

b = 1.8;

%m=input('de o numero de subintervalos do intervalo [a,b]: m=');

m = 698;

h=(b-a)/m; %espaçamento entre os pontos

for j = 1:m+1

    x(j) = a + (j-1)\*h;

    y(j) = f\_trap(x(j)); %vetor que armazena os pontos relativos à função integrada

end

```
%Regra dos Trapezios
```

```
soma = 0;
```

```
for j = 2:1:n
```

```
    soma = soma + y(j);  
end
```

```
I_TR = (h/2)*(y(1)+ 2*soma + y(m+1));
```

```
fprintf('O valor da integral, obtido pela Regra dos Trapezios, eh dado por %12.8f\n',  
I_TR);
```

```
%Fim da rotina
```

A função `f_trap(z)` tem que ser definida no arquivo `f_trap.m`. O conteúdo deste arquivo é o seguinte:

```
function g = f_trap(z)  
    g = exp(-0.5*z*z);  
%Fim da rotina.
```

## 5.2 Regra de Simpson: Simples e Repetida

Na Regra de Simpson simples, a função  $f: [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$  será aproximada pelo polinômio de Lagrange de grau 2, sobre os pontos de interpolação igualmente espaçados  $x_0 = a$ ,  $x_1 = (a + b)/2$  e  $x_2 = b$ . Observe que o espaçamento entre os pontos é dado por  $h = (b - a)/2$ . Para a dedução da fórmula do erro na Regra de Simpson, a função  $f$  deverá possuir derivada quarta contínua no intervalo fechado  $[a, b]$ .

A partir das considerações anteriores, tem-se que

$$\int_a^b f(x) dx = \int_{x_0}^{x_2} p_2(x) dx + \int_{x_0}^{x_2} \frac{(x-x_0)(x-x_1)(x-x_2)f^{(3)}(\xi(x))}{3!} dx,$$

onde  $\xi(x) \in (a, b)$ ,  $\forall x \in [a, b]$ . A Regra de Simpson,  $I_S$ , será obtida da integral de  $p_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x)$ , onde

$$L_0(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}; \quad L_1(x) = \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \quad \text{e} \quad L_2(x) = \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}.$$

Utilizando a mesma troca de variáveis exibida no início da Seção 5.1,  $x = x_0 + uh$ ,

$$\text{obtem-se que} \quad \int_{x_0}^{x_2} p_2(x) dx = \int_0^2 p_2(x_0+uh) h du.$$

Note que

$$p_2(x_0 + uh) = f(x_0)L_0(x_0 + uh) + f(x_1)L_1(x_0 + uh) + f(x_2)L_2(x_0 + uh) =$$

$$f(x_0) \frac{(u-1)h(u-2)h}{-h(-2h)} + f(x_1) \frac{uh(u-2)h}{h(-h)} + f(x_2) \frac{uh(u-1)h}{2h(h)}.$$

Assim,

$$\int_{x_0}^{x_2} p_2(x) dx = \frac{f(x_0)}{2} \int_0^2 (u-1)(u-2)h du - f(x_1) \int_0^2 u(u-2)h du$$

$$+ \frac{f(x_2)}{2} \int_0^2 u(u-1)h du.$$

Agora, observe que

$$\int_0^2 (u-1)(u-2)h du = \left[ h\left(\frac{u^3}{3} - \frac{3u^2}{2} + 2u\right) \right]_0^2 = 2h/3;$$

$$\int_0^2 u(u-2)h du = \left[ h\left(\frac{u^3}{3} - u^2\right) \right]_0^2 = -4h/3 \text{ e } \int_0^2 u(u-1)h du = \left[ h\left(\frac{u^3}{3} - \frac{u^2}{2}\right) \right]_0^2 = 2h/3.$$

$$\text{Portanto, } I_S = \int_{x_0}^{x_2} p_2(x) dx = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)].$$

Para se demonstrar a fórmula do erro na Regra de Simpson, vamos utilizar a seguinte igualdade que foi deduzida na Seção 4.4 (3ª estimativa para o erro de interpolação):

$$\frac{f^{(2+1)}(\xi(x))}{(2+1)!} = f[x_0, x_1, x_2, x]. \text{ Dessa forma,}$$

$$E_S = \int_{x_0}^{x_2} \frac{(x-x_0)(x-x_1)(x-x_2)f^{(3)}(\xi(x))}{3!} dx = \int_{x_0}^{x_2} N(x)f[x_0, x_1, x_2, x] dx.$$

Utilizando integração por partes com  $u = f[x_0, x_1, x_2, x]$  e  $dv = N(x) dx$ , obtém-se  $v =$

$$\Omega(x) = \int_{x_0}^x N(t) dt. \text{ Assim, } E_S = f[x_0, x_1, x_2, x] \Omega(x) \Big|_{x_0}^{x_2} - \int_{x_0}^{x_2} f'[x_0, x_1, x_2, x] \Omega(x) dx.$$

Pode-se mostrar que (i)  $N(x_1 - t) = -N(x_1 + t)$ , para qualquer número real  $t$ ; (ii)  $\Omega(x) \geq 0$ , para todo  $x \in [x_0, x_2]$ ; (iii)  $\Omega(x_0) = \Omega(x_2) = 0$ ; (iv)  $f'[x_0, x_1, x_2, x] = \frac{f^{(4)}(\xi(x))}{(4)!}$ , com  $\xi(x) \in (x_0, x_2)$ .

**Observação:** Os itens anteriores serão demonstrados logo após a dedução da fórmula do erro da Regra de Simpson.



De acordo com os itens anteriores e utilizando-se o resultado que foi apresentado na 6ª **Observação** da Seção 5.1, segue-se que

$$E_S = - \frac{f^{(4)}(\xi(c))}{(4)!} \int_{x_0}^{x_2} \Omega(x) dx = \frac{f^{(4)}(\xi(c))}{(4)!} \int_{x_0}^{x_2} (x - x_0)N(x)dx.$$

Na última igualdade foi utilizada integração por partes com  $\mathbf{u} = \Omega(x)$ ;  $d\mathbf{v} = 1dx$ ;  $d\mathbf{u} = \Omega'(x) = N(x)$  e  $\mathbf{v} = (x - x_0)$ . Fazendo a troca de variáveis  $x = x_0 + uh$ , obtém-se que

$$\begin{aligned} E_S &= \frac{f^{(4)}(C)}{(4)!} \int_0^2 uh h^3 u(u-1)(u-2) h du = \frac{h^5 f^{(4)}(C)}{24} \int_0^2 u^2(u^2 - 3u + 2) du \rightarrow \\ &\rightarrow E_S = \frac{h^5 f^{(4)}(C)}{24} \left[ \frac{u^5}{5} - \frac{3u^4}{4} + \frac{2u^3}{3} \right]_0^2 = \frac{-h^5 f^{(4)}(C)}{90}, \text{ com } C \in (x_0, x_2). \end{aligned}$$

**Observação:** De acordo com a fórmula do erro anterior, a Regra Simples de Simpson é exata ( $E_S = 0$ ) para polinômios de grau até 3 porque, nesses casos,  $f^{(4)}(x) = 0$  para todo número real  $x$ .

## Propriedades importantes e suas demonstrações

**Propriedade 1:**  $N(x) = (x - x_0)(x - x_1)(x - x_2) \rightarrow N(x_1 - t) = -N(x_1 + t), \forall t \in \mathbb{R}$ , se os pontos forem igualmente espaçados, ou seja,  $x_i = x_0 + ih, h = (b - a)/2, 1 \leq i \leq 2$ .

**Demonstração:**  $N(x_1 - t) = (x_1 - t - x_0)(x_1 - t - x_1)(x_1 - t - x_2) = (h - t)(-t)(-h - t) = -(-h + t)(t)(h + t)$  e  $N(x_1 + t) = (x_1 + t - x_0)(x_1 + t - x_1)(x_1 + t - x_2) = (h + t)(t)(-h + t)$ . Portanto,  $N(x_1 - t) = -N(x_1 + t)$ . ■

**Propriedade 2:**  $\Omega(x) = \int_{x_0}^x N(t) dt \geq 0, \forall x \in (x_0, x_2)$ .

**Demonstração:** Se  $x_0 \leq x \leq x_1$  então  $x_0 \leq t \leq x \leq x_1$ . Assim,  $N(t) = (t - x_0)(t - x_1)(t - x_2) \geq 0 \rightarrow \Omega(x) \geq 0$ . Agora, considere  $0 < s = x - x_1$  e  $x_1 < x \leq x_2$ . Assim, utilizando trocas de variáveis convenientes e a propriedade 1, obtemos que

$$\Omega(x) = \int_{x_0}^{x_1-s} N(t) dt + \int_{x_1-s}^{x_1} N(t) dt + \int_{x_1}^x N(t) dt = \int_{x_0}^{x_1-s} N(t) dt - \int_s^0 N(x_1 - T) dT +$$

$$\int_0^s N(x_1 + T) dT = \int_{x_0}^{x_1-s} N(t) dt + \int_0^s N(x_1 - T) dT - \int_0^s N(x_1 - T) dT = \int_{x_0}^{x_1-s} N(t) dt \geq 0,$$

pois  $x_0 \leq x_1 - s < x_1$ . □

**Propriedade 3:**  $\Omega(x_0) = \Omega(x_2) = 0$ .

**Demonstração:**  $\Omega(x_0) = \int_{x_0}^{x_0} N(t) dt = 0$ ;  $\Omega(x_2) = \int_{x_0}^{x_1} N(t) dt + \int_{x_1}^{x_2} N(t) dt$ . Fazendo a troca de variáveis  $t = x_1 - T$ , na primeira integral, e  $t = x_1 + T$ , na segunda integral, obtém-se que

$$\Omega(x_2) = - \int_h^0 N(x_1 - T) dT + \int_0^h N(x_1 + T) dT = \int_0^h N(x_1 - T) dT + \int_0^h N(x_1 + T) dT = 0,$$

pela propriedade 1. □

**Propriedade 4:** Suponha que  $f$  tenha derivada quarta contínua em  $[a, b]$  e que os pontos distintos dois a dois  $x, x_0, x_1$  e  $x_2$  pertençam ao intervalo  $[a, b]$ . Então,

$$f'[x_0, x_1, x_2, x] = \frac{f^{(4)}(\xi(x))}{(4)!}, \text{ com } \xi(x) \in (a, b).$$

Para facilitar a demonstração da propriedade 4, serão apresentados e demonstrados alguns lemas sobre a forma integral da diferença dividida.

Como motivação, suponha que  $f$  satisfaça as condições do Teorema do Valor Médio (veja as referências na Seção 3 do Módulo 1). Então,  $f[x_0, x_1] = [f(x_1) - f(x_0)]/(x_1 - x_0) = f'(c)$ , com  $c$  entre  $x_0$  e  $x_1$ . Logo,  $c = x_0 + t(x_1 - x_0)$ , onde  $0 < t < 1$ . Note que

$$\int_0^1 f'(x_0 + t(x_1 - x_0)) dt = \frac{1}{(x_1 - x_0)} [f(x_0 + t(x_1 - x_0))]_0^1 = f[x_0, x_1].$$

Voltando à Seção 4.3 deste módulo, depois da primeira observação daquela seção, note que  $d_2 = f[x_0, x_1, x_2] = \frac{1}{x_2 - x_1} \left\{ \frac{f(x_2) - f(x_0)}{x_2 - x_0} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right\}$ . Através de um raciocínio análogo ao do parágrafo anterior, obtém-se que

$$d_2 = f[x_0, x_1, x_2] = \frac{1}{(x_2 - x_1)} \int_0^1 \{f'(x_0 + t_1(x_2 - x_0)) - f'(x_0 + t_1(x_1 - x_0))\} dt_1.$$

Observe que

$$(i) \quad f'(x_0 + t_1(x_2 - x_0)) = f'(x_0 + t_1(x_2 - x_1 + x_1 - x_0)) = f'(x_0 + t_1(x_1 - x_0) + t_1(x_2 - x_1));$$

$$(ii) \quad f'(x_0 + t_1(x_1 - x_0)) = f'(x_0 + t_1(x_1 - x_0) + 0(x_2 - x_1)).$$

Assim,  $f[x_0, x_1, x_2] = \int_0^1 \int_0^{t_1} f''(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1)) dt_2 dt_1$ . Em geral, tem-se o

**Lema 1.** Suponha que  $f$  seja derivável até ordem  $n$  em  $[a, b]$  e que  $x_i \in [a, b]$ ,  $0 \leq i \leq n$ , sejam pontos distintos dois a dois. Então

$$f[x_0, x_1, x_2, \dots, x_n] = \int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-1}} f^{(n)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + \dots + t_n(x_n - x_{n-1})) dt_n \dots dt_2 dt_1.$$

**Demonstração:** Primeiro observe que  $a \leq x_0 + \sum_{i=1}^n t_i(x_i - x_{i-1}) \leq b$ , pois  $0 \leq t_n \leq t_{n-1} \leq$

$\dots \leq t_2 \leq t_1 \leq 1$  e  $x_0 + \sum_{i=1}^n t_i(x_i - x_{i-1}) = \sum_{i=0}^n \alpha_i x_i$ , onde  $0 \leq \alpha_0 = 1 - t_1$ ;  $0 \leq \alpha_i = t_i - t_{i+1}$ ,

$1 \leq i \leq n-1$  e  $0 \leq \alpha_n = t_n$ ; logo  $\sum_{i=0}^n \alpha_i = 1$ . Além disso, pelo **Teorema 2** da Seção 4.3,

tem-se que  $f[x_0, x_1, x_2, \dots, x_{n-2}, x_{n-1}] = f[x_{n-1}, x_0, x_1, x_2, \dots, x_{n-2}]$  e  $f[x_{n-1}, x_0, x_1, x_2, \dots, x_{n-2}, x_n] = f[x_0, x_1, x_2, \dots, x_{n-2}, x_{n-1}, x_n]$ .

Observe, também, que utilizando a Regra da Cadeia para calcular a derivada de  $f^{(n-1)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + \dots + t_n(x_n - x_{n-1}))$  com respeito à variável  $t_n$ , e utilizando o Teorema Fundamental do Cálculo, obtém-se que

$$\int_0^{t_{n-1}} f^{(n)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + \dots + t_n(x_n - x_{n-1})) dt_n = \frac{1}{x_n - x_{n-1}} \left[ f^{(n-1)}(x_0 + \sum_{i=1}^{n-1} t_i(x_i - x_{i-1}) + t_n(x_n - x_{n-1})) \right]_0^{t_{n-1}}.$$

Agora, como hipótese de indução, suponha que as diferenças divididas de ordem  $n-1$  satisfaçam a fórmula integral (antes do enunciado do lema, foi mostrado que a fórmula integral é válida para diferenças divididas de ordens 1 e 2). Assim,

$$\int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-1}} f^{(n)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + \dots + t_n(x_n - x_{n-1})) dt_n \dots dt_2 dt_1 =$$

$$\int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-2}} \frac{1}{x_n - x_{n-1}} \left[ f^{(n-1)}(x_0 + \sum_{i=1}^{n-1} t_i(x_i - x_{i-1}) + t_n(x_n - x_{n-1})) \right]_0^{t_{n-1}} dt_{n-1} \dots dt_2 dt_1 =$$

$$\int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-2}} \frac{1}{x_n - x_{n-1}} f^{(n-1)}(x_0 + \sum_{i=1}^{n-2} t_i(x_i - x_{i-1}) + t_{n-1}(x_n - x_{n-1} + x_{n-1} - x_{n-2})) dt_{n-1} \dots dt_2 dt_1$$

$$\begin{aligned}
& - \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-2}} \frac{1}{x_n - x_{n-1}} f^{(n-1)} \left( x_0 + \sum_{i=1}^{n-2} t_i (x_i - x_{i-1}) + t_{n-1} (x_{n-1} - x_{n-2}) \right) dt_{n-1} \dots dt_2 dt_1 = \\
& \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-2}} \frac{1}{x_n - x_{n-1}} f^{(n-1)} \left( x_0 + \sum_{i=1}^{n-2} t_i (x_i - x_{i-1}) + t_{n-1} (x_n - x_{n-2}) \right) dt_{n-1} \dots dt_2 dt_1 \\
& - \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-2}} \frac{1}{x_n - x_{n-1}} f^{(n-1)} \left( x_0 + \sum_{i=1}^{n-2} t_i (x_i - x_{i-1}) + t_{n-1} (x_{n-1} - x_{n-2}) \right) dt_{n-1} \dots dt_2 dt_1 = \\
& = (f[x_0, x_1, x_2, \dots, x_{n-2}, x_n] - f[x_0, x_1, x_2, \dots, x_{n-2}, x_{n-1}]) / (x_n - x_{n-1}) = f[x_{n-1}, x_0, x_1, x_2, \dots, x_{n-2}, x_n] \\
& = f[x_0, x_1, x_2, \dots, x_{n-2}, x_{n-1}, x_n]. \quad \square
\end{aligned}$$

**Lema 2.** Suponha que  $f$  seja uma função real derivável até ordem  $n + 1$  e que satisfaça as seguintes condições: (i)  $f^{(k)}(0) = 0$ , para todo  $k$  tal que  $0 \leq k \leq n$ ; (ii)  $f(1) = 1$ . Então,

$$\int_0^1 \int_0^{t_1} \cdots \int_0^{t_n} f^{(n+1)}(t_{n+1}) dt_{n+1} dt_n \dots dt_2 dt_1 = 1.$$

**Demonstração (por indução finita):** Se  $n = 1$  então:  $f(0) = 0$ ;  $f'(0) = 0$  e  $f(1) = 1$ . Assim,

$$\int_0^1 \int_0^{t_1} f^{(2)}(t_2) dt_2 dt_1 = \int_0^1 [f'(t_1) - f'(0)] dt_1 = f(1) - f(0) = 1.$$

Considere a seguinte Hipótese de Indução:  $\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} f^{(n)}(t_n) dt_n dt_{n-1} \dots dt_2 dt_1 = 1$ , se (i)  $f^{(k)}(0) = 0$ , para todo  $k$  tal que  $0 \leq k \leq n - 1$ ; (ii)  $f(1) = 1$ . Dessa forma, se  $f^{(n)}(0) = 0$ ,

$$\begin{aligned}
& \int_0^1 \int_0^{t_1} \cdots \int_0^{t_n} f^{(n+1)}(t_{n+1}) dt_{n+1} dt_n \dots dt_2 dt_1 = \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} [f^{(n)}(t_{n+1})]_0^{t_n} dt_n \dots dt_2 dt_1 = \\
& \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} [f^{(n)}(t_n) - f^{(n)}(0)] dt_n \dots dt_2 dt_1 = \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} f^{(n)}(t_n) dt_n \dots dt_2 dt_1 = 1. \quad \square
\end{aligned}$$

**Lema 3.**  $\int_0^1 \int_0^{t_1} \cdots \int_0^{t_n} dt_{n+1} dt_n \dots dt_2 dt_1 = 1/(n + 1)!.$

**Demonstração:** Seja  $f(t) = t^{n+1}$  então  $f^{(k)}(t) = (n+1)(n)...(n+2-k) t^{(n+1-k)}$  e  $(n+1)! = f^{(n+1)}(t)$ . Dessa forma: (i)  $f^{(k)}(0) = 0$ , para todo  $k$  tal que  $0 \leq k \leq n$ ; (ii)  $f(1) = 1$ . Portanto, pelo **Lema 2**,

$$\int_0^1 \int_0^{t_1} \dots \int_0^{t_n} f^{(n+1)}(t_{n+1}) dt_{n+1} dt_n \dots dt_2 dt_1 = 1 \rightarrow \int_0^1 \int_0^{t_1} \dots \int_0^{t_n} dt_{n+1} dt_n \dots dt_2 dt_1 = 1/(n+1)! \quad \square$$

Com auxílio dos lemas anteriores, já é possível demonstrar a **Propriedade 4**.

**Demonstração da Propriedade 4:** De acordo com o **Lema 1**, tem-se que

$$f[x_0, x_1, x_2, x] = \int_0^1 \int_0^{t_1} \int_0^{t_2} f^{(3)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + t_3(x - x_2)) dt_3 dt_2 dt_1.$$

Como  $f$  tem derivada quarta contínua, então é possível derivar a diferença dividida com respeito à variável  $x$ . Dessa forma,

$$f'[x_0, x_1, x_2, x] = \int_0^1 \int_0^{t_1} \int_0^{t_2} f^{(4)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + t_3(x - x_2)) t_3 dt_3 dt_2 dt_1.$$

Sabendo que  $t_3 \geq 0$  e que  $f^{(4)}$  é contínua em  $[a, b]$ , o Teorema do Valor Intermediário pode ser aplicado (veja a 6ª **Observação** da Seção 5.1) sucessivamente para se obter:

$$\begin{aligned} f'[x_0, x_1, x_2, x] &= \int_0^1 \int_0^{t_1} \int_0^{t_2} f^{(4)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + t_3(x - x_2)) \int_0^{t_2} t_3 dt_3 dt_2 dt_1 = \\ &= \int_0^1 f^{(4)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + t_3(x - x_2)) \int_0^{t_1} F(t_2) dt_2 dt_1 = \\ &= f^{(4)}(x_0 + c_1(x_1 - x_0) + c_2(x_2 - x_1) + c_3(x - x_2)) \int_0^1 F(t_1) dt_1 = \\ &= f^{(4)}(\xi(x)) \int_0^1 \int_0^{t_1} \int_0^{t_2} t_3 dt_3 dt_2 dt_1 = f^{(4)}(\xi(x)) \int_0^1 \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} dt_4 dt_3 dt_2 dt_1 = \frac{f^{(4)}(\xi(x))}{(4)!}, \end{aligned}$$

de acordo com o **Lema 3**. □

**Observações:** Na demonstração da **Propriedade 4**, tem-se que  $F(t_2) = \int_0^{t_2} t_3 dt_3 > 0$  e

$$F(t_1) = \int_0^{t_1} F(t_2) dt_2 > 0; \quad c_3 \in (0, t_2), \quad c_2 \in (0, t_1), \quad c_1 \in (0, 1) \quad \text{e} \quad \xi(x) = x_0 + c_1(x_1 - x_0) + c_2(x_2 - x_1) + c_3(x - x_2) \rightarrow \xi(x) \in (a, b).$$

## A Regra Repetida de Simpson

A Regra Repetida de Simpson consiste em dividir o intervalo  $[a, b]$  em um número par,  $m$ , de subintervalos de mesmo comprimento  $h = (b - a)/m$ , construir os pontos igualmente espaçados  $x_i = x_0 + ih$ ,  $1 \leq i \leq m$ , com  $x_0 = a$  e  $x_m = b$ , e aproximar a função  $f$  por um polinômio de Lagrange de grau 2,  $p_2^{(i)}(x)$ , em cada subintervalo  $[x_{2i-2}, x_{2i}]$ , onde  $1 \leq i \leq m/2$ .

$$\begin{aligned} \text{Dessa forma, } \int_a^b f(x) dx &= \sum_{i=1}^{m/2} \int_{x_{2i-2}}^{x_{2i}} f(x) dx \approx \\ &\sum_{i=1}^{m/2} \int_{x_{2i-2}}^{x_{2i}} p_2^{(i)}(x) dx = \sum_{i=1}^{m/2} \frac{h}{3} [f(x_{2i-2}) + 4f(x_{2i-1}) + f(x_{2i})]. \end{aligned}$$

**Observação:** a integral do polinômio  $p_2^{(i)}(x)$  é deduzida de forma análoga à dedução feita para a Regra Simples de Simpson. Basta trocar  $x_0$  por  $x_{2i-2}$ ,  $x_1$  por  $x_{2i-1}$  e  $x_2$  por  $x_{2i}$ .

A Regra Repetida de Simpson pode ser reescrita como:

$$I_{SR} = (h/3) \{ f(x_0) + 4 \sum_{i=1}^{m/2} f(x_{2i-1}) + 2 \sum_{i=1}^{(m/2)-1} f(x_{2i}) + f(x_m) \}.$$

De acordo com a fórmula do erro para a Regra Simples de Simpson, o erro de integração, em cada intervalo  $[x_{2i-2}, x_{2i}]$ , será dado por  $-\frac{h^5 f^{(4)}(C_i)}{90}$ , onde  $C_i \in (x_{2i-2}, x_{2i})$ ,  $1 \leq i \leq m/2$ . Portanto, o erro da Regra Repetida de Simpson é dado por

$$E_{SR} = \sum_{i=1}^{m/2} -\frac{h^5 f^{(4)}(C_i)}{90} = -\frac{h^5}{90} \sum_{i=1}^{m/2} f^{(4)}(C_i) = -\frac{mh^5 f^{(4)}(\tilde{C})}{180}, \text{ onde } \tilde{C} \in (a, b).$$

**Observação:** A última igualdade na fórmula do erro da Regra Repetida de Simpson é consequência do Teorema do Valor Intermediário. De fato,

$$\begin{aligned} (m/2) f^{(4)}(x_{\min}) &\leq \sum_{i=1}^{m/2} f^{(4)}(C_i) \leq (m/2) f^{(4)}(x_{\max}) \rightarrow \\ \rightarrow f^{(4)}(x_{\min}) &\leq \frac{2}{m} \sum_{i=1}^{m/2} f^{(4)}(C_i) \leq f^{(4)}(x_{\max}). \end{aligned}$$

## Estimativa do erro na Regra Repetida de Simpson

$$\text{Note que } |E_{SR}| = \frac{mh^5 |f^{(4)}(\tilde{C})|}{180} \leq \frac{mh^5}{180} \max\{|f^{(4)}(x)|; x \in [a, b]\}.$$

**Exemplo 11. (Estimativa de Erro na Regra de Simpson)** Utilizando a função de integração do **Problema 2** da Seção 5.1, será determinado o número de subintervalos para se obter erro de integração inferior a  $10^{-6}$ , com a Regra de Simpson.

Naquele problema, o objetivo era calcular o valor de  $F(5) = 0,5 + \frac{1}{\sqrt{2\pi}} \int_0^{1.8} e^{-\frac{1}{2}z^2} dz$ .

Para que o valor aproximado da integral tenha erro inferior a  $10^{-6}$ , é necessário calcular o número de subintervalos,  $m$ , do intervalo  $[a, b] = [0, 1.8]$  de modo que

$$|E_{SR}| = \frac{mh^5 |f^{(4)}(\tilde{C})|}{180} \leq \frac{mh^5}{180} \max\{|f^{(4)}(x)|; x \in [a, b]\} < 10^{-6}.$$

Observe que

$f(z) = \exp(-0.5z^2) \rightarrow f'(z) = -z \cdot \exp(-0.5z^2) \rightarrow f^{(2)}(z) = (z^2 - 1) \cdot \exp(-0.5z^2) \rightarrow$   
 $\rightarrow f^{(3)}(z) = (3z - z^3) \cdot \exp(-0.5z^2) \rightarrow f^{(4)}(z) = (3 - 6z^2 + z^4) \cdot \exp(-0.5z^2) \rightarrow f^{(5)}(z)$   
 $= (-15z + 10z^3 - z^5) \cdot \exp(-0.5z^2)$ . Portanto,  $f^{(5)}(z) = 0$  se, e somente se,  $z = 0$  (as outras raízes são números complexos). Assim,  $\max\{|f^{(4)}(z)|; z \in [0, 1.8]\} = \max\{|f^{(4)}(0)|; |f^{(4)}(1.8)|\} = 3$ . Como  $h = (1.8 - 0)/m$ , segue-se que

$$\frac{mh^5}{180} \max\{|f^{(4)}(x)|; x \in [0, 1.8]\} = m^{-4}(1.8)^5(1/180) < 10^{-6} \rightarrow m^4 > (1.8)^4 \times 10^4 \rightarrow$$

$m > 18 \rightarrow m = 20$ , pois  $m$  deve ser um número par.

Comparando este valor de  $m$  (número de subintervalos) com o obtido pela Regra dos Trapézios (698 subintervalos) fica evidente que a Regra de Simpson é mais eficiente do que a Regra dos Trapézios. O valor aproximado da integral, utilizando a Regra Repetida de Simpson com 20 subintervalos, é dado por 1.16325015147918. Se fossem utilizados 698 subintervalos, o valor obtido seria 1.16325018351050. O valor aproximado da integral, utilizando a Regra dos Trapézios com 698 subintervalos, é igual a 1.16324998610089.

**Observação:** O código computacional com a Regra Repetida de Simpson, desenvolvido em Octave, será exibido a seguir.

```

%Regra de Simpson (1/3)

clear

a=input('entre com o valor do extremo inferior da integral: a=');
b=input('entre com o valor do extremo superior da integral: b=');
n=input('entre com o numero de subintervalos do intervalo [a,b]: n(par)=');

h=(b-a)/n; %espacamento entre os pontos

for j = 1:n+1
    x(j) = a + (j-1)*h;
    y(j) = f_simp(x(j));
end

%Regra de Simpson

soma_p = 0;
soma_imp = 0;

for j = 2:2:(n-2)
    soma_p = soma_p + y(j);
    soma_imp = soma_imp + y(j+1);
end

IS = (h/3)*(y(1)+4*(soma_p+y(n))+2*(soma_imp)+y(n+1));

fprintf('O valor da integral, obtido por Simpson, eh dado por %12.8fn', IS);
%Fim da rotina

A função f_simp(z) tem que ser definida no arquivo f_simp.m. O conteúdo deste
arquivo é o seguinte:
        function g = f_simp(z)
            g = exp(-0.5*z*z);
        %Fim da rotina.

```



## 6. Atividades do Módulo 3

### Atividade 1 – Adquirindo e testando conhecimentos através da resolução da lista de exercícios

#### Enunciado da atividade

Em praticamente todos os exercícios desta lista você precisará lidar com polinômios; portanto, é fundamental que você recorde a teoria vista sobre este assunto na disciplina Fundamentos da Matemática Elementar I.

Esta lista contém exercícios referentes ao conteúdo do Módulo 3: Ajuste de Curvas, existência e unicidade do polinômio interpolador; Polinômio de Lagrange; Polinômio de Newton com Diferenças Divididas; Regra dos Trapézios e Regra de Simpson. Você tem que se esforçar para tentar fazer todos os exercícios.

#### Terceira Lista de Exercícios

1) Uma função da forma  $\varphi(x) = y^*/(b \cdot e^{(-ax)} + 1)$  ( $y^*$ : constante dada;  $a$  e  $b$ : parâmetros a serem determinados) é utilizada para ajustar os pontos  $(x_i, y_i)$  de uma tabela;  $\varphi(x_i) \approx y_i$ . Este é um modelo não linear de ajuste de curva. Para simplificar o ajuste, será utilizado o método dos quadrados mínimos (MQM) em um problema linearizado, ou seja, será preciso modificar os pontos da tabela, levando-se em consideração os pontos  $z_i = T(y_i)$ .

(i) Obtenha a linearização,  $T$ , adequada para este problema.

(ii) O sistema linear resultante terá quantas equações? Justifique .

2) Considere os pontos a seguir.

$x$	-1.0	0.0	1.0	2.0
$f(x)$	1.7321	1.0	1.7321	3.0

Os pontos serão ajustados por uma função do tipo  $Q(x) = Q(x; a, b) = a(1 + bx^2)^{1/2}$ . Observe que  $Q(x)$  não é linear nos parâmetros  $a$  e  $b$ . Para que o Método dos Quadrados Mínimos – modelo linear – seja utilizado, deve-se linearizar o problema.

i) Justifique por que a transformação  $T(y) = y^2$  é adequada para linearizar o problema. Quais serão os novos parâmetros?

ii) Calcule somente o vetor de termos independentes do sistema linear do problema linearizado (se  $Ax = b$ , então  $b$  é o vetor de termos independentes).

iii) Estime o valor de  $f(1.5)$  sabendo que  $p_1 = 1.000113$  e  $p_2 = 1.99998$  são os parâmetros do problema linearizado associados às funções  $g_1(x) = 1$  e  $g_2(x) = x^2$ , respectivamente.

3) Uma função da forma  $\varphi(x) = 20 + [1/(a + bx)]$  é utilizada para ajustar os pontos  $(x_i, y_i)$  de uma tabela;  $\varphi(x_i) \approx y_i$ . Este é um modelo não linear de ajuste de curva. Para simplificar o ajuste, será utilizado o método dos quadrados mínimos (MQM) em um problema linearizado. Será preciso modificar os pontos da tabela, levando-se em consideração os pontos  $z_i = T(y_i)$ .

(i) Obtenha a linearização,  $T$ , adequada para este problema.

(ii) O sistema linear resultante terá quantas equações? Justifique.

4) Considere os pontos tabelados a seguir.

$x$	0	1	2	3	4	5	6
$y$	32	47	65	92	132	190	275

O ajuste dos pontos será feito por uma função do tipo  $\varphi(x) = ab^x$ . Observe que  $\varphi$  não é linear nos parâmetros  $a$  e  $b$ . Para se utilizar o modelo linear do método dos quadrados mínimos, será preciso obter uma linearização. Nesse caso, a linearização adequada é a função  $\ln(y)$ .

(i) Calcule os termos independentes associados ao problema linearizado.

(ii) Resolva o sistema linear e obtenha os novos parâmetros  $\beta_1 = 3.4704$  e  $\beta_2 = 0.3555$ , associados às funções  $g_1(x) = 1$  e  $g_2(x) = x$ , respectivamente. Em seguida, obtenha uma estimativa para  $y(7)$ .

(iii) Utilizando uma função exponencial do tipo  $\psi(x) = ax^b$  para ajustar os pontos da tabela anterior, obtemos  $a = 38.83871$  e  $b = 2.61978$ , pelo método dos quadrados mínimos não linear. Explique por que esta função não é adequada para ajustar os pontos tabelados.

5) Considere os pontos a seguir.

$x$	-1.0	0.0	1.0	2.0
$f(x)$	20.5	21.0	20.5	20.2774

Os pontos serão ajustados por uma função do tipo  $Q(x) = Q(x; a, b) = 20 + \frac{1}{a\sqrt{1+bx^2}}$ . Observe que  $Q(x)$  não é linear nos parâmetros  $a$  e  $b$ . Para que o modelo linear do Método dos Mínimos Quadrados seja utilizado, deve-se considerar uma transformação adequada.

(i) Use a transformação  $T(y) = 1/(y-20)^2$  para linearizar o problema. Quais serão os novos parâmetros associados à família de curvas do tipo  $T(Q(x))$ ?

(ii) Calcule somente os termos independentes do sistema linear do problema linearizado.

(iii) Estime o valor de  $f(3.0)$ , sabendo que  $p_1 = 1.00077946$  e  $p_2 = 2.9987009$  são os valores dos parâmetros do problema linearizado, associados às funções  $g_1(x) = 1$  e  $g_2(x) = x^2$ , respectivamente.

6) Dada a tabela

$x$	0.00	0.10	0.50	1.00	1.50
$f(x)$	2.00	2.22	3.72	8.39	21.08

suponha que o ajuste dos pontos seja feito pelo Método dos Quadrados Mínimos (MQM) com uma função do tipo  $Q(x) = Q(x; a_1, a_2) = a_1g_1(x) + a_2g_2(x)$ . Faça o diagrama de dispersão e indique qual das funções abaixo fornecerá o melhor ajuste: a)  $g_1(x) = 1$  e  $g_2(x) = e^x$ ; b)  $g_1(x) = 1$  e  $g_2(x) = 1/x$ ; c)  $g_1(x) = 1$  e  $g_2(x) = \sin(x)$ . Justifique a sua resposta.

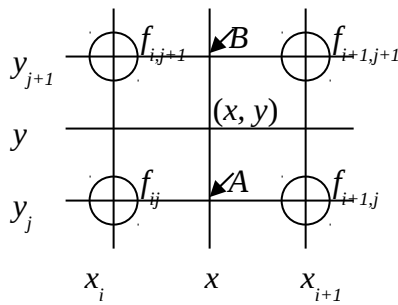
7) Considere a seguinte função  $f(k) = \sum_{i=1}^k i$ . Note que  $f(1) = 1$  e  $f(2) = 1 + 2 = 3$ .

(i) Atribua valores para  $k$  e construa o polinômio de Lagrange de grau dois que interpola a função  $f$ . (ii) Sabendo-se que  $f(k) = \frac{1}{2} k(k+1)$ , obteríamos resultados melhores se interpolássemos  $f$  utilizando um polinômio de Lagrange de grau 3?

8) O polinômio de Lagrange, de grau  $n$ , que interpola uma função real  $f: \mathbb{R} \rightarrow \mathbb{R}$  nos pontos  $x_0, x_1, \dots, x_n$  é dado por  $p_n(x) = \sum_{i=0}^n f(x_i) L_i(x)$ , onde  $L_i(x) = N_i(x)/N_i(x_i)$  e  $N_i(x)$  é obtido do polinômio  $N(x) = (x - x_0)(x - x_1) \dots (x - x_i) \dots (x - x_n)$  retirando-lhe o fator

$(x - x_i)$ . Se  $f$  for um polinômio constante, pode-se mostrar que  $\sum_{i=0}^n L_i(x) = 1, \forall x \in \mathbb{R}$ . Justifique este fato. Lembre-se de que o polinômio interpolador é único.

9) Seja  $f(x,y)$  uma função definida sobre os pares  $(x,y)$ , com  $x_i \leq x \leq x_{i+1}$  e  $y_j \leq y \leq y_{j+1}$ .



Esquema da aproximação

y	x			
	75	100	125	150
120.5	13	45	51	61
100.0	35	55	64	70
81.5	54	68	72	79
65.0	72	78	82	87
42.5	89	90	91	93

Tabela de valores de  $f(x, y)$

A função  $f(x,y)$  pode ser aproximada da seguinte forma (veja Fig. 1): primeiro faça a interpolação linear através de  $f_{ij} = f(x_i, y_j)$  e  $f_{i+1,j} = f(x_{i+1}, y_j)$  e obtenha a aproximação  $f_A$ ; em seguida, através de  $f_{ij+1} = f(x_i, y_{j+1})$  e  $f_{i+1,j+1} = f(x_{i+1}, y_{j+1})$  obtenha a aproximação  $f_B$ . Interpole linearmente através de  $f_A$  e  $f_B$  para obter a aproximação desejada.

Sejam  $\alpha = (x - x_i)/(x_{i+1} - x_i)$  e  $\beta = (y - y_j)/(y_{j+1} - y_j)$ . Mostre que

$$f(x, y) \sim (1 - \alpha)(1 - \beta) f_{ij} + \alpha(1 - \beta) f_{i+1,j} + (1 - \alpha)\beta f_{i,j+1} + \alpha\beta f_{i+1,j+1} = p(x,y) = f_{ij}.L_{ij}(x, y) + f_{i+1,j}.L_{i+1,j}(x, y) + f_{i,j+1}.L_{i,j+1}(x, y) + f_{i+1,j+1}.L_{i+1,j+1}(x, y); \text{ tal que } p(x_i, y_j) = f(x_i, y_j).$$

Note que  $L_{ij}(x, y) = P_i(x)Q_j(y)$ , em que  $P_i$  é um polinômio na variável  $x$  e  $Q_j$  é um polinômio na variável  $y$ .

Utilizando a fórmula anterior, juntamente com os pontos da tabela acima, obtenha uma aproximação para  $f(x, y) = f(110, 98)$ .

10) Considere  $x_0, x_1, x_2$  e  $x_3$  pontos distintos dois a dois. O polinômio interpolador com diferenças divididas (polinômio de Newton) de grau três é dado por

$$p_3(x) = d_0 + d_1(x-x_0) + d_2(x-x_0)(x-x_1) + d_3(x-x_0)(x-x_1)(x-x_2).$$

Para o cálculo de valores numéricos é conveniente reescrever o polinômio como segue.

$$p_3(x) = (d_0 + (x-x_0)(d_1 + (x-x_1)(d_2 + (x-x_2)d_3))).$$

- i) Compare o número de adições (somas e subtrações) e multiplicações efetuadas em cada uma das expressões de  $p_3(x)$ .
- ii) Conhecido os valores das diferenças divididas  $d_0, d_1, d_2$  e  $d_3$ , a melhor maneira de se calcular  $p_3(c)$  é através da regra dos parênteses encaixados:

$$\begin{aligned} b &= d_3; \\ b &= d_2 + (c-x_2)b; \\ b &= d_1 + (c-x_1)b; \\ b &= d_0 + (c-x_0)b. \end{aligned}$$

O último valor de  $b$  é igual a  $p_3(c)$ . Agora, escreva um algoritmo para o cálculo do valor numérico do polinômio de Newton de grau  $n$ .

11) A tabela abaixo pode ser encontrada em diversos livros didáticos de Geografia Geral e do Brasil. Ela registra a evolução da dívida externa de alguns países Latino-Americanos. Alguns dados (1984) foram extraídos do RELATÓRIO do desenvolvimento humano 2000. Nova York: PNUD, Lisboa: Trinova, 2000.

Evolução da dívida externa – Valores Absolutos  
em bilhões de dólares.

País/ano	1977	1987	1998
Argentina	8.1	53.9	144.0
Brasil	28.3	109.4	232.0
Chile	4.9	18.7	36.3
Honduras	0.6	3.1	5.0
Jamaica	1.1	4.3	4.0
México	26.6	93.7	159.9
Venezuela	9.8	29.0	37.0

KENNEDY, Paul. *Preparando para o século XXI*, Rio de Janeiro: Campus, 1993; p. 208

Um estudante de Geografia precisava comparar as dívidas externas, em 1990, dos países que constam na tabela; pedindo ajuda a um professor de Cálculo Numérico, ele conseguiu as estimativas desejadas. O professor utilizou polinômios interpoladores de grau 2. Neste problema particular, qual é o polinômio interpolador mais adequado do ponto de vista operacional: o de Newton ou o de Lagrange? Justifique.

12) Suspeita-se que a tabela

x	-3.0	-2.0	-1.0	0.0	1.0	2.0
f(x)	-9.0	0.0	1.0	0.0	3.0	16.0

represente um polinômio de grau três. Como testar este fato? Justifique.

13) Quer-se interpolar a função  $\sin(x)$ , sobre o intervalo  $I = [a, b] = [0, \pi/4]$ , usando um polinômio de grau 2, com pontos igualmente espaçados. Qual deve ser o menor número,  $m$ , de subintervalos de  $I$  para se garantir um erro menor do que  $10^{-4}$ . Nesse caso, a fórmula a seguir é válida:

$$|f(x) - p_n(x)| < [h^{n+1} \cdot \max\{|f^{(n+1)}(x)|, x \in I\}]/(4(n+1)), \text{ onde } h = (b-a)/m.$$

14) O erro na interpolação linear é dado por  $E^{(2)} = (1/2) f''(c(x))N(x)$ , onde  $N(x) = (x-x_0)(x-x_1)$ . É fácil mostrar que  $x \in [x_0, x_1]$  se, e somente se, existe  $u \in [0, 1]$  tal que  $x = x_0 + uh$ , com  $h = x_1 - x_0$ .

Nesse caso, pode-se definir  $M(u) = |N(x_0 + uh)| = h^2 |(u)(u-1)|$ . Dessa forma,  $u_{\max}$  é o ponto de máximo de  $M$  se, e somente se,  $x_{\max} = x_0 + u_{\max}h$  é ponto de máximo de  $|N(x)|$ . Portanto,  $|E^{(2)}| \leq (h^2/8) \max\{|f''(x)|; x \in [x_0, x_1]\}$ .

Qual deve ser o menor número de pontos igualmente espaçados de uma tabela contendo valores de  $\cos(x)$ ,  $x \in [1, 2]$ , de modo que, ao se usar interpolação linear para aproximar  $\cos(x^*)$ , obtenha-se erro inferior a  $10^{-6}$ , qualquer que seja  $x^* \in [1, 2]$ ?

15) Dados  $n + 1$  pontos distintos dois a dois,  $x_0, x_1, x_2, x_3, \dots, x_n$ , pode-se provar, por indução finita, que uma base para o espaço dos polinômios de grau até  $n$ ,  $\mathcal{P}_n$ , é dada por  $\mathcal{B} = \{1, (x-x_0), (x-x_0)(x-x_1), (x-x_0)(x-x_1)(x-x_2), \dots, (x-x_0)(x-x_1)\dots(x-x_{n-1})\}$ . O polinômio interpolador com diferenças divididas de grau  $n$  é dado por  $p_n(x) = d_0 + d_1(x-x_0) + d_2(x-x_0)(x-x_1) + \dots + d_n(x-x_0)(x-x_1)(x-x_2)\dots(x-x_{n-1})$ . Prove que se  $\mathcal{B}' = \{1, (x-x_0), (x-x_0)(x-x_1), (x-x_0)(x-x_1)(x-x_2), \dots, (x-x_0)(x-x_1)\dots(x-x_{n-1}), (x-x_0)(x-x_1)\dots(x-x_n)\}$  for a base do espaço  $\mathcal{P}_{n+1}$ , então

$$p_{n+1}(x) = D_0 + D_1(x-x_0) + D_2(x-x_0)(x-x_1) + \dots + D_{n+1}(x-x_0)(x-x_1)\dots(x-x_n),$$

onde  $D_i = d_i$ ,  $0 \leq i \leq n$ .

16) Considere os pontos  $x_0, x_1, x_2$  e  $x_3$  distintos dois a dois e as seguintes bases de  $\mathcal{P}_3$ :

$$\begin{aligned} \mathcal{B} &= \{1, (x-x_0), (x-x_0)(x-x_1), (x-x_0)(x-x_1)(x-x_2)\}; \text{ (polinômio interpolador } p_3(x)) \\ &\quad \text{e} \\ \mathcal{B}_\sigma &= \{1, (x-x_{\sigma(0)}), (x-x_{\sigma(0)})(x-x_{\sigma(1)}), (x-x_{\sigma(0)})(x-x_{\sigma(1)})(x-x_{\sigma(2)})\}; \text{ (p. interp. } p_{3,\sigma}(x)), \end{aligned}$$

onde  $\sigma(k) = n - k$ , com  $n = 3$  e  $0 \leq k \leq n$ . Mostre que a diferença dividida de ordem três é dada por  $d_3 = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$ , onde  $f[x_1, x_2, x_3]$  e  $f[x_0, x_1, x_2]$  são diferenças divididas de ordem 2. Sugestão:  $h(x) = p_{3,\sigma}(x) - p_3(x) = 0$ , pois o polinômio interpolador é único; assim, o coeficiente de  $x^2$  em  $h(x)$  é igual a zero. Tente generalizar este resultado.

17) Dada a tabela

$x$	1	2	3	4	5	6
$f(x)$	0.841	0.909	0.141	-0.757	-0.959	-0.279

obtenha  $x^*$  tal que  $f(x^*) = 0$ . Use um polinômio de grau 2 e apresente uma estimativa do erro. A estimativa com um polinômio de grau 3 seria melhor? Justifique.

18) Dada a tabela

$x$	-0.5	-0.25	0	0.25	0.58
$f(x)$	0	0.5	1	0.75	0.42

considere o seguinte problema: “Encontrar  $x^*$  tal que  $f(x^*) = 0.6$ ”. Justifique as afirmativas abaixo. Não há necessidade de fazer contas.

i) Utilizando, por exemplo, os três primeiros pontos da tabela na construção do polinômio de grau 2, que dará a aproximação de  $x^*$ , a estimativa do erro, levando-se em conta as diferenças divididas de ordem 3, não será boa (mesmo estando correto o valor encontrado pela interpolação).

ii) A função  $f$  é linear por partes ( $f$  é uma reta em cada um dos seguintes intervalos:  $[-0.5, 0]$  e  $[0, 0.58]$ ), assim, o erro de interpolação deve ser igual a zero. Lembre-se de que se  $f$  é crescente (ou decrescente), então  $f^{-1}$  é também crescente (ou decrescente).

19) Dada a tabela

$x$	$\pi/8$	$\pi/4$	$(3\pi)/8$	$\pi/2$	$(3\pi)/4$	$(5\pi)/4$
$f(x)$	0.3827	0.7071	0.9239	1.0000	0.7071	-0.7071

(i) Por que os pontos entre  $f(\pi/4)$  e  $f((5\pi)/4)$  não são adequados para se calcular  $x^*$  tal que  $f(x^*) = 0.8$ ?

(ii) Obtenha o valor de  $x^*$  utilizando um polinômio de grau 2, de modo que o erro estimado seja mínimo.

(iii) Calcule uma estimativa para o erro.

20) Sabe-se que  $\int_0^1 \sin(x) dx = 0.459698$ .

(i) Mostre que o resultado numérico obtido pela Regra de 1/3 de Simpson, com  $h = 0.5$ , é uma aproximação muito boa desta integral. Justifique isso pela fórmula do erro:  $E_{(1/3)S} = -[h^4(b-a)f^{(iv)}(c)]/180$ . Sugestão: use uma estimativa do erro.

(ii) No item anterior, qual o valor de  $n$  deve ser usado para se obter uma aproximação com erro menor do que  $10^{-6}$ ?

21) Sabe-se que  $\int_{0.1}^1 \frac{1}{x} dx = \ln(10)$ .

(i) Verifique que o resultado numérico obtido pela regra de 1/3 de Simpson, com  $n = 2$ , não é uma boa aproximação para essa integral. Faça o cálculo pela regra dos trapézios também. Qual apresentará melhor aproximação? Justifique.

(ii) Qual o valor de  $n$  deve ser usado, na regra repetida de Simpson, para se obter uma aproximação com erro menor do que  $10^{-3}$ ? Calcule, também, o valor de  $n$  para a regra repetida do Trapézio.

22) Defina  $f$  como segue:  $f(x) = (1+x)$ , se  $0 \leq x \leq 1$ ;  $f(x) = (x+2)^3$ , se  $1 < x \leq 2$ . Observe que  $f$  é integrável no intervalo  $[0, 2]$ . Quais os procedimentos numéricos, para se obter a integral de  $f$  no intervalo dado, que utilizam esforço computacional mínimo com precisão máxima? Justifique a sua resposta utilizando os conhecimentos a respeito das regras de Simpson e do Trapézio. Exiba os pontos utilizados e os respectivos valores da função nestes pontos. Por que devemos considerar  $f(1) = 27$ , ao se integrar  $f$  no intervalo  $[1, 2]$ ?

23) Defina  $f$  como segue:  $f(x) = x^5 \ln(x)$ , se  $0.1 \leq x \leq 0.36$ ;  $f(x) = (x+1)$ , se  $0.36 < x \leq 1$ . Quais os procedimentos numéricos, para se obter uma aproximação da integral de  $f$  no intervalo dado, que utilizam esforço computacional mínimo com estimativa de erro menor do que  $10^{-9}$ ? Justifique a sua resposta utilizando os conhecimentos a respeito das regras de Simpson e do Trapézio. Sugestão: Se  $g(x) = x^5 \ln(x)$  então  $g^{(iv)}(x) = x[120 \ln(x) + 154]$ .

24) Quer-se interpolar e integrar a função  $\sin(x)$ , sobre o intervalo  $I = [a, b] = [0, \pi/4]$ , usando um polinômio de grau  $n = 2$ , com pontos igualmente espaçados. Qual deve ser o



menor número,  $m$ , de subintervalos de  $I$  para se garantir um erro menor do que  $10^{-4}$  tanto na integração quanto na interpolação? A fórmula a seguir é válida:

$$|f(x) - p_n(x)| \leq \frac{h^{n+1}}{4(n+1)} \max\{|f^{(n+1)}(x)|, x \in I\}, \text{ onde } h = (b-a)/m.$$

25) A fórmula do erro de integração relativa à regra de  $1/3$  de Simpson é obtida resolvendo-se a seguinte integral:  $I = \int_a^b f[x_0, x_1, x_2, x] N(x) dx$ . Lembre-se de que  $f^{(n+1)}(\xi(x))/(n+1)! = f[x_0, x_1, x_2, \dots, x_n, x]$ ;  $N(x) = (x-x_0)(x-x_1)(x-x_2)$ ;  $x_i = x_0 + ih$ ,  $1 \leq i \leq 2$ ;  $x_0 = a$  e  $h = (b-a)/2$ .

Utilizando a regra de integração por partes com  $u = f[x_0, x_1, x_2, x]$  e  $dv = N(x) dx$ , obtém-se  $v = \Omega(x) = \int_a^x N(t) dt$ . Assim,  $I = f[x_0, x_1, x_2, x] \Omega(x) \Big|_a^b - \int_a^b f''[x_0, x_1, x_2, x] \Omega(x) dx$ .

i) Faça o gráfico de  $N(x)$  ou o gráfico de  $\tilde{N}(u) = N(x_0 + uh)$ ,  $0 \leq u \leq 2$ .

ii) Mostre que  $N(x_1 - t) = -N(x_1 + t)$ , para todo  $t$  real (no caso geral, em que  $n$  é o número de subintervalos de  $[a, b]$  e é par, tem-se que  $N(x_m - t) = -N(x_m + t)$ ,  $\forall t$  real e  $m = n/2$ ).

iii) Mostre que  $\Omega(a) = \Omega(b) = 0$ . Sugestão: integre em dois intervalos,  $[x_0, x_1]$  e  $[x_1, x_2]$ ; considere as seguintes trocas de variáveis:  $x = x_1 - t$  (no primeiro intervalo) e  $x = x_1 + t$  (no segundo).

26) Considerando o polinômio interpolador de grau  $n$ ,  $p_n(x)$ , de uma função  $f: [a, b] \rightarrow \mathbb{R}$ , escrito na base de Newton,  $\mathcal{B} = \{1, (x-x_0), (x-x_0)(x-x_1), \dots, (x-x_0)(x-x_1)\dots(x-x_{n-1})\}$ , sabe-se que os coeficientes da combinação linear dos elementos da base são chamados de diferenças divididas.

Por exemplo, o coeficiente  $d_2$  que multiplica o elemento  $(x-x_0)(x-x_1)$  é obtido da igualdade  $p_n(x_2) = f(x_2)$ . Assim,

$$d_2 \equiv f[x_0, x_1, x_2] = \frac{1}{x_2 - x_1} \left\{ \frac{f(x_2) - f(x_0)}{x_2 - x_0} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right\}.$$

Para o cálculo do erro de integração, é adequado obter uma expressão integral para a diferença dividida. Observe que

$$f[x_0, x_1] = [f(x_1) - f(x_0)]/(x_1 - x_0) = f'(c),$$

com  $c$  entre  $x_0$  e  $x_1$  (basta usar o Teorema do Valor Médio); logo,  $c = x_0 + t \cdot (x_1 - x_0)$ ,

$0 < t < 1$ . Note que

$\int_0^1 f'(x_0 + t(x_1 - x_0)) dt = \frac{1}{(x_1 - x_0)} f(x_0 + t(x_1 - x_0)) \Big|_0^1 = f[x_0, x_1]$ . Através de um raciocínio análogo, obtém-se que

$$d_2 = f[x_0, x_1, x_2] = \frac{1}{(x_2 - x_1)} \int_0^1 \{f'(x_0 + t_1(x_2 - x_0)) - f'(x_0 + t_1(x_1 - x_0))\} dt_1.$$

Observe que

$$a) \quad f'(x_0 + t_1(x_2 - x_0)) = f'(x_0 + t_1(x_2 + (x_1 - x_1) - x_0)) \\ = f'(x_0 + t_1(x_1 - x_0) + t_1(x_2 - x_1));$$

$$b) \quad f'(x_0 + t_1(x_1 - x_0)) = f'(x_0 + t_1(x_1 - x_0) + 0(x_2 - x_1)).$$

Assim,  $f[x_0, x_1, x_2] = \int_0^1 \int_0^{t_1} f''(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1)) dt_2 dt_1$ . No caso geral, mostra-se por indução finita que

$$f[x_0, x_1, x_2, \dots, x_n] = \int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-1}} f^{(n)}(x_0 + t_1(x_1 - x_0) + t_2(x_2 - x_1) + \dots + t_n(x_n - x_{n-1})) dt_n \dots dt_2 dt_1.$$

i) Exiba a expressão integral para  $f[x_0, x_1, x_2, x]$  e calcule a sua derivada.

ii) Usando o teorema do valor intermediário obtém-se que  $f'[x_0, x_1, x_2, x] = f^{(4)}(c(x))$

$$\int_0^1 \int_0^{t_1} \int_0^{t_2} t_3 dt_3 dt_2 dt_1, c(x) \in (a, b). \text{ Calcule a integral anterior.}$$

## Atividade 2 – Método dos Quadrados Mínimos

Este exercício, relacionado ao crescimento de uma árvore, foi retirado do livro do professor Rodney Bassanezi (BASSANEZI, 2002), *Ensino-aprendizagem com modelagem matemática*. Eu aproveitei a questão para motivar os alunos da turma de Cálculo 1 que estavam iniciando, no segundo semestre de 2014, o curso de Agronomia da UFU.

Para você que gosta de matemática e de cálculo numérico, acredito que a resolução do exercício será bastante motivadora, também, já que estudamos matemática com o propósito e com o prazer de resolver problemas. Se a solução desses problemas auxiliam os colegas de outras áreas, o prazer é ainda maior.

### Enunciado da atividade

A tabela abaixo fornece os dados do crescimento de uma árvore. A variável  $t$  indica o tempo em anos e a variável  $y$  representa a altura da árvore em decímetros (dm).

$i$	1	2	3	4	5	6	7	8	9	10	11	12	13
$t_i$	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9
$y_i$	21.7	22.5	23.3	24	24.7	25.4	26	26.6	27.1	27.6	28.1	28.5	28.9

(a) Faça o diagrama de dispersão (marque no plano cartesiano cada um dos pontos  $(t, y)$  da tabela). Use algum programa de computador que faça gráfico.

(b) A função que será utilizada para ajustar os pontos da tabela é uma curva logística do tipo  $Q(x) = \frac{a}{1 + b \cdot e^{-\lambda x}}$ . A constante  $a$  é o valor assintótico de  $y$ , ou seja  $\lim_{t \rightarrow \infty} y(t) = a$ . Através do método dos quadrados mínimos, você deverá encontrar, primeiramente, o valor da constante  $a$ , utilizando os últimos 6 pontos da tabela; depois você calculará os parâmetros  $b$  e  $\lambda$ . Para isso, siga os passos abaixo.

(b.1) Faça o diagrama de dispersão dos pontos  $(y_k, y_{k+1})$ , da tabela abaixo. Note que eles estão alinhados, ou seja,  $y_{k+1} = \alpha_1 + \alpha_2 y_k$ . Se  $x = y_k$  e  $y = y_{k+1}$ , então  $y = \alpha_1 + \alpha_2 x$ .

$k$	1	2	3	4	5	6
$y_k$	26	26.6	27.1	27.6	28.1	28.5
$y_{k+1}$	26.6	27.1	27.6	28.1	28.5	28.9

(b.2) Usando o Método dos Quadrados Mínimos para obter os parâmetros da equação da reta anterior, obteremos o seguinte sistema linear (verifique isso):

$A\alpha = B$ , onde  $A = \begin{pmatrix} 6 & 163.9 \\ 163.9 & 4481.59 \end{pmatrix}$ ;  $\alpha^t = (\alpha_1 \quad \alpha_2)$  e  $B = \begin{pmatrix} 166.8 \\ 4560.48 \end{pmatrix}$ . Resolva esse sistema utilizando o método de Eliminação de Gauss com Pivoteamento Parcial.

(b.3) Sabendo-se que  $y_{k+1} = \alpha_1 + \alpha_2 y_k$ , o valor assintótico de  $y$  pode ser calculado como segue:  $a = \frac{\alpha_1}{1 - \alpha_2}$ . Justifique a afirmação anterior. Nesse caso,  $a = 33.77664948$ .

(b.4) Manipule a expressão da curva logística e use as propriedades da função logarítmica para mostrar que  $\ln\left(\frac{Q}{a - Q}\right) = -\ln(b) + \lambda t$ .

(b.5) Defina  $z_i = \ln\left(\frac{y_i}{a - y_i}\right) \approx \beta_1 + \beta_2 t_i$ , onde  $\beta_1 = -\ln(b)$  e  $\beta_2 = \lambda$ . Adicione uma nova linha na tabela com os valores de  $z_i$  e utilize todos os pontos  $(t_i, z_i)$ ,  $1 \leq i \leq 13$ , para construir o sistema linear oriundo do Método dos Quadrados Mínimos que fornecerá os parâmetros  $\beta_1$  e  $\beta_2$ . Você deverá encontrar o seguinte sistema:

$M\beta = B$ , onde  $M = \begin{pmatrix} 13 & 78 \\ 78 & 513.5 \end{pmatrix}$ ;  $\beta^t = (\beta_1 \quad \beta_2)$  e  $B = \begin{pmatrix} 15.5649914 \\ 102.458452 \end{pmatrix}$ .

Observação: Os valores do vetor  $B$  são sensíveis a erros de arredondamento e, portanto, podem variar bastante dependendo do número de casas decimais que foram utilizadas nas contas.

(b.6) Conclua que a função logística terá a seguinte expressão:

$$Q(x) = \frac{33.7766}{1 + 0.998541 \cdot e^{-0.199308t}}.$$

### Atividade 3 – Polinômio Interpolador de Lagrange

**Prezado(a) aluno(a),** Você poderá encontrar vários problemas práticos relacionados à interpolação polinomial na página da Faculdade de Matemática da UFU:

[www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) (Laboratório - Unidade 4).

#### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre Polinômio Interpolador de Lagrange: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 4), ambos localizados em [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) A tabela abaixo permite que sejam calculadas as estimativas da dívida externa, em 1992, de qualquer um dos países mencionados. Obtenha a estimativa da dívida do Brasil. Use o polinômio de Lagrange de grau 2. Aproveite algumas das contas anteriores (polinômio do Brasil) para calcular a estimativa da dívida do México, usando um polinômio de grau 2, também.

Evolução da dívida externa – Valores Absolutos  
em bilhões de dólares.

País/ano	1977	1987	1998
Argentina	8.1	53.9	144.0
Brasil	28.3	109.4	232.0
Chile	4.9	18.7	36.3
Honduras	0.6	3.1	5.0
Jamaica	1.1	4.3	4.0
México	26.6	93.7	159.9
Venezuela	9.8	29.0	37.0

KENNEDY, Paul. *Preparando para o século XXI*, Rio de Janeiro: Campus, 1993; p. 208

#### Informações sobre a Atividade 3.

(1) Polinômios da base ( $x_0 = 1977$ ,  $x_1 = 1987$  e  $x_2 = 1998$ ):

$$L_0(x) = \frac{(x-1987)(x-1998)}{(1977-1987)(1977-1998)} ; \quad L_1(x) = \frac{(x-1977)(x-1998)}{(1987-1977)(1987-1998)} \text{ e}$$
$$L_2(x) = \frac{(x-1977)(x-1987)}{(1998-1977)(1998-1987)} .$$

(2) Polinômio Interpolador do Brasil:  $p_{2,B}(x) = f_{0,B} L_0(x) + f_{1,B} L_1(x) + f_{2,B} L_2(x)$ , onde  $f_{0,B} = 28,3$ ;  $f_{1,B} = 109,4$  e  $f_{2,B} = 232,0$ .

(3) Polinômio Interpolador do México:  $p_{2,M}(x) = f_{0,M} L_0(x) + f_{1,M} L_1(x) + f_{2,M} L_2(x)$ , onde  $f_{0,M} = 26,6$ ;  $f_{1,M} = 93,7$  e  $f_{2,M} = 159,9$ .

(4) Valor Numérico do polinômio do Brasil:  $p_{2,B}(1992) = f_{0,B} L_0(1992) + f_{1,B} L_1(1992) + f_{2,B} L_2(1992) = -0.142857142 f_{0,B} + 0.818181818 f_{1,B} + 0.324675324 f_{2,B} \approx 160.791$ .

(5) Valor Numérico do polinômio do México:  $p_{2,M}(1992) = f_{0,M} L_0(1992) + f_{1,M} L_1(1992) + f_{2,M} L_2(1992) = -0.142857142 f_{0,M} + 0.818181818 f_{1,M} + 0.324675324 f_{2,M} \approx 124.779$ .

#### Atividade 4 – Polinômio de Newton – Interpolação Inversa

##### Prezado(a) aluno(a),

Você já deve ter visto na disciplina de Cálculo que toda função real contínua e monótona (crescente ou decrescente) definida em um intervalo fechado  $[a, b]$ , possui inversa. Em determinados problemas, precisamos usar um polinômio interpolador para aproximar a inversa de uma dada função. Acontece que, em muitos casos, a função pode ser dada por meio de uma tabela. Assim, como saber se a função representada por uma tabela possui inversa? É simples. Basta supor que a função tabelada seja contínua e considerar apenas os pontos da tabela nos quais podemos afirmar que a função é monótona (crescente ou decrescente). A interpolação inversa só terá validade no intervalo onde a função é estritamente crescente ou estritamente decrescente.

##### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre Polinômio de Newton e Interpolação Inversa: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 4), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Resolva o seguinte exercício (professor Balthazar, Unesp Rio Claro): Na tabela abaixo são registradas, em cada instante de tempo, as velocidades de um elevador. Calcule os instantes em que a velocidade do elevador se anula (tenha cuidado ao escolher os pontos). Use polinômios de grau 2. Apresente os cálculos das estimativas dos erros.

Tempo	0	1	2	3	4	5	6	7	8	9	10
Veloc.	2.0	4.0	6.2	3.9	0.3	-0.67	-0.32	0.18	4.1	6.3	4.8

#### Informações sobre a Atividade 4.

(1) Escolha dos pontos da primeira interpolação inversa.

Como o problema exige que seja feita uma estimativa de erro de interpolação então o polinômio interpolador deve ser o de Newton. A análise do erro vai precisar das diferenças divididas de ordem 3, pois o polinômio interpolador é de grau 2. Assim, os pontos escolhidos para a velocidade são: 6.2; 3.9; 0.3 e -0.67, que correspondem a uma função decrescente.

(2) Construção do polinômio interpolador.

#### Tabela de diferenças divididas

$v_i$ (veloc.)	$t_i$ (tempo)	D.D.Ordem 1	D.D.Ordem 2	D.D.Ordem 3
6.2	2	-0.434782608	-0.026610988	-0.027862309
3.9	3	-0.277777778	0.164803075	
0.3	4	-1.030927835		
-0.67	5			

#### Polinômio de grau 2

$$p_2(v) = d_0 + d_1(v - v_0) + d_2(v - v_0)(v - v_1);$$

onde

$$v_0 = 3.9; v_1 = 0.3; v_2 = -0.67; d_0 = 3; d_1 = -0.277777778; d_2 = 0.164803075.$$

#### Valor Numérico em $v = 0$ .

$$p_2(0) = 4.276152932 = t^* \text{ (primeiro instante em que a velocidade se anula).}$$

(3) Estimativa do erro.

$$|E^{(2)}(0)| \approx |0 - v_0| |0 - v_1| |0 - v_2| \text{ Máx}\{|DDO3|\} = 0.021841264.$$

(4) Escolha dos pontos da segunda interpolação inversa.

Comentários análogos aos do item 1. A análise do erro vai precisar das diferenças divididas de ordem 3, pois o polinômio interpolador é de grau 2. Assim, os pontos escolhidos para a velocidade são: -0.67; -0.32; 0.18; 4.1 e 6.3, que correspondem a uma função crescente.

(5) Construção do polinômio interpolador.

Tabela de diferenças divididas

$v_i$ (veloc.)	$t_i$ (tempo)	D.D.Ordem 1	D.D.Ordem 2	D.D.Ordem 3
-0.67	5	2.857142857	-1.008403361	0.128643619
-0.32	6	2	-0.394773294	0.064556206
0.18	7	0.25510204	0.032588793	
4.1	8	0.454545455		
6.3	9			

Polinômio de grau 2

$$p_2(v) = d_0 + d_1(v - v_0) + d_2(v - v_0)(v - v_1).$$

Como o intuito é o de se obter o menor valor para o erro de interpolação então o polinômio interpolador deve envolver pontos  $v_0$ ,  $v_1$  e  $v_2$  tais que o valor de  $|N(0)| = |0 - v_0| |0 - v_1| |0 - v_2|$  seja o menor possível. Assim, a escolha adequada é a seguinte:

$$v_0 = -0.67; v_1 = -0.32; v_2 = 0.18; d_0 = 5; d_1 = 2.857142857 \text{ e } d_2 = -1.008403361.$$

Valor Numérico em  $v = 0$ .

$$p_2(0) = 6.698084034 = t^* \text{ (segundo instante em que a velocidade se anula).}$$

$$6) \text{ Estimativa do erro: } |E^{(2)}(0)| \approx |0 - v_0| \cdot |0 - v_1| \cdot |0 - v_2| \text{ Máx}\{|DDO3|\} \approx 0.00496.$$



## Atividade 5 – Regra dos Trapézios

### Prezado(a) aluno(a),

Este exercício é continuação da atividade anterior. A primeira vez que tive contato com ele foi em 1987, na aula de cálculo numérico, quando fiz a graduação (1986 - 1989) em Matemática na Unesp de Rio Claro. A primeira parte da disciplina foi ministrada pelo professor José Bezerra Leite (que também ministrou para a minha turma, no ano seguinte, a disciplina Análise Numérica) e a segunda parte foi ministrada pelo professor José Manoel Balthazar.

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre a Regra dos Trapézios: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 5), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Resolva o seguinte exercício: Na tabela abaixo são registradas, em cada instante de tempo, as velocidades de um elevador. Calcule o espaço total percorrido pelo elevador  $(\int_0^{10} |v| dt)$ . Use a Regra dos trapézios. Lembre-se de que  $|v|$  não possui derivada no instante  $t$  em que a velocidade do elevador se anula ( $t = 4.28$  e  $t = 6.70$ ). Nesses casos, você deverá utilizar a Regra Simples do Trapézio.

Tempo	0	1	2	3	4	5	6	7	8	9	10
Veloc.	2.0	4.0	6.2	3.9	0.3	-0.67	-0.32	0.18	4.1	6.3	4.8

### Informações sobre a Atividade 5.

Dada a tabela

Tempo	0	1	2	3	4	5	6	7	8	9	10
Veloc.	2.0	4.0	6.2	3.9	0.3	-0.67	-0.32	0.18	4.1	6.3	4.8

Note que

$$\int_0^{10} |v| dt = \int_0^4 |v| dt + \int_4^{4.28} |v| dt + \int_{4.28}^5 |v| dt + \int_5^6 |v| dt + \int_6^{6.70} |v| dt + \int_{6.70}^7 |v| dt + \int_7^{10} |v| dt = I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7.$$

Observe que os pontos tabelados,  $t_i$  (tempo), são igualmente espaçados, ou seja,  $t_i = ih$ ,  $0 \leq i \leq 10$ , com  $h = 1$ . Dessa forma, podemos usar a Regra dos Trapézios para calcular as integrais  $I_1$  e  $I_7$ :

$$I_1 \approx (h/2) \{ |v(0)| + 2 [ |v(1)| + |v(2)| + |v(3)| ] + |v(4)| \} = 0.5 \{ 2 + 2 [ 4 + 6.2 + 3.9 ] + 0.3 \} = 15.25.$$

$$I_7 \approx (h/2) \{ |v(7)| + 2 [ |v(8)| + |v(9)| ] + |v(10)| \} = 0.5 \{ 0.18 + 2 [ 4.1 + 6.3 ] + 4.8 \} = 12.89.$$

As demais integrais são calculadas com a regra simples do trapézio, contendo um único intervalo. Assim:

$$I_2 \approx [ (4.28 - 4) / 2 ] \{ |v(4)| + |v(4.28)| \} = 0.14 \{ 0.3 + 0 \} = 0.042.$$

$$I_3 \approx [ (5 - 4.28) / 2 ] \{ |v(4.28)| + |v(5)| \} = 0.36 \{ 0 + 0.67 \} = 0.2412.$$

$$I_4 \approx [ (6 - 5) / 2 ] \{ |v(5)| + |v(6)| \} = 0.50 \{ 0.67 + 0.32 \} = 0.495.$$

$$I_5 \approx [ (6.70 - 6) / 2 ] \{ |v(6)| + |v(6.70)| \} = 0.35 \{ 0.32 + 0 \} = 0.112.$$

$$I_6 \approx [ (7 - 6.70) / 2 ] \{ |v(6.70)| + |v(7)| \} = 0.15 \{ 0 + 0.18 \} = 0.027.$$

$$\text{Portanto, } \int_0^{10} |v| dt = I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 \approx 29.0572 \text{ m.}$$

## Atividade 6 – Regra de Simpson

### Prezado(a) aluno(a),

Este exercício é uma aplicação de integração numérica. A modelagem do problema vai utilizar a fórmula do comprimento de arco,  $L$ , de uma curva representada por uma função  $f: I = [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ . Em Cálculo 2 você aprendeu que  $L =$

$$\int_a^b \sqrt{1 + (f'(t))^2} dt .$$

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre a Regra de Simpson: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 5), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) Resolva o seguinte exercício: Uma telha ondulada tem a forma da função dada por  $f(t) = 0.05 \sin(50 t)$ ,  $t$  in  $[1, 2]$ . A parte não ondulada da telha mede 3 metros. Quarenta telhas serão utilizadas para cobrir uma garagem e serão pintadas com uma tinta especial impermeabilizante que, também, tem o efeito de reduzir o aquecimento do local coberto. Um litro desta tinta custa R\$2,80. O rendimento da tinta é o seguinte: 9 litros de tinta para cada  $5 \text{ m}^2$ . Qual o custo para se pintar as 40 telhas? Sugestão: a área de uma telha é  $A = 3 \times L$ . Use a Regra de Simpson com 8 subintervalos para calcular  $L$ . Qual seria a economia se o valor de  $L$  fosse calculado com 1000 subintervalos?

### Informações sobre a Atividade 6.

(1) Obtenção de  $h$ .

A integral  $\int_a^b \sqrt{1 + (f'(t))^2} dt$  vai ser resolvida pela Regra de Simpson com  $m = 8$  subintervalos;  $a = 1$  e  $b = 2$ . Dessa forma,  $h = (b - a)/m = (2 - 1)/8 = 0.125$ .

(2) Cálculo da função que será integrada:

$$g(t) = [1 + (f'(t))^2]^{1/2}, \text{ onde } f(t) = 0.05\sin(50 t).$$

Observe que  $g(t) = [1 + 6.25 \cos^2(50 t)]^{1/2}$ , pois  $f'(t) = 2.5 \cos(50 t)$ .

(3) Cálculo da Integral pela Regra de Simpson com  $m = 8$  subintervalos.

Observe que  $t_i = t_0 + ih = 1 + i 0.125$ . Assim,  $t_0 = 1$ ;  $t_1 = 1.125$ ;  $t_2 = 1.25$ ;  $t_3 = 1.375$ ;  $t_4 = 1.50$ ;  $t_5 = 1.625$ ;  $t_6 = 1.75$ ;  $t_7 = 1.875$  e  $t_8 = 2$ . Dessa forma.

$$IS = (h/3) \{g(t_0) + 4 \sum_{i=1}^{\frac{m}{2}} g(t_{2i-1}) + 2 \sum_{i=1}^{\frac{m}{2}-1} g(t_{2i}) + g(t_m)\} =$$

$$(0.125/3) \{ g(t_0) + 4 [ g(t_1) + g(t_3) + g(t_5) + g(t_7)] + 2 [g(t_2) + g(t_4) + g(t_6)] + g(t_8) \} .$$

Portanto,  $IS = (0.125/3) 60.14349999 = 2.505979166$ .

(4) Custo para pintar 40 telhas.

Primeiro precisamos calcular a área de uma telha:  $A = 3 \times L = 3 \times 2.505979166 \text{ m}^2 = 7.517937499 \text{ m}^2$ . Dessa forma, a área de 40 telhas é igual a  $AT = 300.7175 \text{ m}^2 \approx 301 \text{ m}^2$ . Como o rendimento da tinta é de  $5 \text{ m}^2$  para cada 9 litros, então para se pintar  $301 \text{ m}^2$  serão necessários 541.8 litros de tinta. Como 1 litro de tinta custa R\$2, 80, então o custo total para se pintar 40 telhas será igual a R\$ 1.517,04.

(5) Cálculo da Regra de Simpson com 1000 subintervalos.

Neste caso, devemos utilizar um código computacional para obter o valor aproximado da integral pela regra de Simpson. O valor da integral obtido pela regra de Simpson, com  $m = 1000$  subintervalos, é dado por  $IS = 1.9416$  (utilizando arredondamento simétrico na quarta casa decimal depois da vírgula).

(6) Cálculo do valor economizado.

Com o valor mais preciso para  $L$ , a área de uma telha é dada por  $A = 3 \times L = 3 \times 1.9416 \text{ m}^2 = 5.8248 \text{ m}^2$ . Dessa forma, a área de 40 telhas é igual a  $AT = 232.992 \text{ m}^2 \approx 233 \text{ m}^2$ . Como o rendimento da tinta é de  $5 \text{ m}^2$  para cada 9 litros, então, para se pintar  $233 \text{ m}^2$ , serão necessários 419.4 litros de tinta. Como 1 litro de tinta custa R\$ 2,80, então o custo total para se pintar 40 telhas será igual a R\$ 1.174,32. Em relação ao custo anterior (valor menos preciso de  $L$ ), isso daria uma economia de R\$ 342,72.

#### Algoritmo da Regra de Simpson

%Métodos de Integração Numérica ; %fsimp é a função que vai ser integrada

% a = extremo inferior do intervalo de integração;  
a = 1;

%b = extremo superior do intervalo de integração;  
b = 2;

%m = número de subintervalos do intervalo [a,b]; m deve ser par;  
m = 1000;

h=(b-a)/m; %espaçamento entre os pontos

```

for j = 1:m+1
    x(j) = a + (j-1)*h; %pontos igualmente espaçados; o 1º ponto é x(1) e não x(0).
    y(j) = fsimp(x(j)); %avaliação da função nos pontos da regra de integração
end

%Regra de Simpson (1/3)

soma_p = 0; %soma de índices pares
soma_imp = 0; %soma de índices ímpares

for j = 2:2:(m-2)
    soma_p = soma_p + y(j);
    soma_imp = soma_imp + y(j+1);
end

% Observação: os pontos de integração são: x(1); x(2); ..., x(m+1); m = 1000.
% Como o menor índice é 1 e não 0 então as somas da regra de Simpson
% precisam ser reajustadas. A regra está exibida abaixo.

IS = (h/3)*(y(1)+4*(soma_p+y(m))+2*(soma_imp)+y(m+1));
fprintf('O valor da integral, obtido por Simpson, eh dado por %12.24f\n', IS);

%Fim da rotina computacional
A função fsimp foi implementada em Octave e o arquivo recebeu o seguinte nome:
fsimp.m. A rotina computacional está descrita abaixo.

function g = fsimp(x)
    g = 1 + 6.25*cos(50.0*x)*cos(50.0*x);
    g = sqrt(g);
    %sqrt calcula a raiz quadrada de um número real positivo.
%Fim da rotina

```

## 7. Atividades suplementares



O problema proposto na atividade 1 abaixo foi apresentado, no primeiro semestre de 2005, por Clovis Antonio da Silva, quando era monitor de Cálculo Diferencial e Integral 2 do curso de Graduação em Matemática da Universidade Federal de Uberlândia. Entre 2003 e 2005, tive o prazer de trabalhar com o Clovis em dois projetos de Iniciação Científica. Depois disso, sugeri que ele fosse fazer o Mestrado em Nova Friburgo, no Instituto Politécnico da Universidade do Estado do Rio de Janeiro – IPRJ. Ele foi

orientado inicialmente pelo professor Luis Felipe Feres Pereira e por mim, no final do Mestrado, em 2008, depois que o professor Felipe Pereira foi contratado pela Universidade de Wyoming – USA.

Referências sugeridas pelo aluno Clovis Antonio da Silva: (1) Hamilton Luiz Guidorizzi. *Um Curso de Cálculo*, v.2, Rio de Janeiro, LTC Editora, 5a edição. (2) Heyder Diniz Silva (ex professor da FAMAT - UFU). Apostilas de Estatística e de Probabilidade.

### Atividade 1

A duração de um certo tipo de pneu, em quilômetros rodados, é uma variável que possui distribuição normal, com média igual a 60.000 km e desvio padrão igual a 10.000 km.

- (i) Qual a probabilidade de um pneu, aleatoriamente escolhido, durar mais de 75.000 km?
- (ii) Qual a probabilidade de um pneu, aleatoriamente escolhido, durar entre 50.000 km e 70.000 km?
- (iii) Qual a probabilidade de um pneu, aleatoriamente escolhido, durar entre 63.000 km e 70.000 km?

Sugestões:

(1) Reveja o **Problema 2** da Seção 5.1. Utilize a função de probabilidade que foi dada naquele problema juntamente com as suas propriedades.

(2) Utilize a mesma troca de variáveis sugerida no **Problema 2**. Dessa forma:

$$(i) P(X \geq 75.000) = P(Z \geq 1,5) = 0,5 - 0,4332 = 0,0668.$$

$$(ii) P(50.000 \leq X \leq 70.000) = P(-1 \leq Z \leq 1) = 0,3413 + 0,3413 = 0,6826.$$

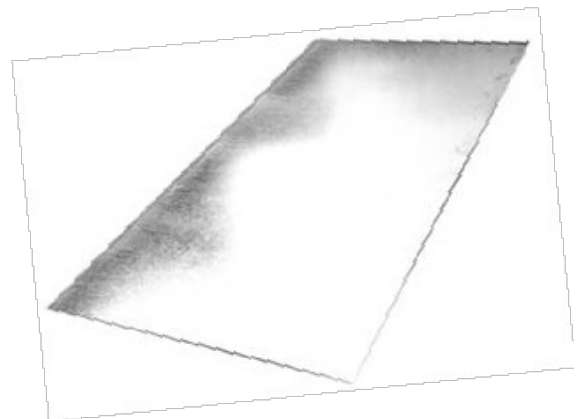
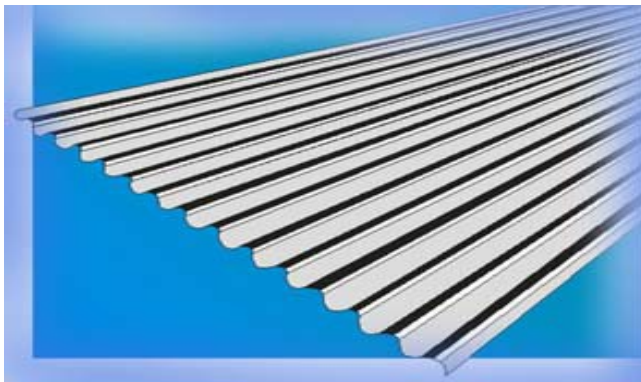
$$(iii) P(63.000 \leq X \leq 70.000) = P(0,3 \leq Z \leq 1) = 0,3413 - 0,1179 = 0,2234.$$

O problema proposto na atividade 2 abaixo foi apresentado pelos alunos Bruno Nunes de Souza, Maksuel Andrade Costa e Ricardo de Oliveira Hakime, quando cursavam, no primeiro semestre de 2005, a disciplina Cálculo Diferencial e Integral 2 na Faculdade de Matemática da UFU. Uma das tarefas dos alunos daquela turma de Cálculo 2 era a apresentação de problemas práticos que utilizassem a teoria vista em aula. Posteriormente, em 2007, eu solicitei ao monitor de Cálculo Numérico (Miller, do curso de Engenharia Civil da UFU) que aperfeiçoasse o problema com o intuito de motivar o estudo de técnicas de integração numérica. O trabalho dos alunos está apresentado abaixo e o trabalho do Miller foi proposto na Atividade 5, Seção 6, deste Módulo 3.

Referência sugerida pelos alunos: Thomas, George B. *Cálculo*. Volume 1, 10ª Edição. São Paulo, Pearson Education, 2002.

### Atividade 2

Uma empresa fabrica folhas de metal onduladas para telhados. As seções transversais dessas folhas têm a forma da curva:  $f(x) = \sin([3\pi x]/20)$ ,  $0 \leq x \leq 20$ . Se as telhas devem ser moldadas a partir de folhas planas, por um processo que não estique o material, qual deve ser o comprimento da folha original?



Sugestão: Utilize a forma de comprimento de arco (veja a atividade 5 da Seção 6 – Módulo 3). Obtenha uma aproximação para a integral através da Regra dos Trapézios ou através da Regra repetida de Simpson.

## Módulo 4

# Solução Numérica de Problema de Valor Inicial

### 1. Introdução

**Prezado estudante, seja bem vindo.**

Neste módulo você utilizará a série de Taylor de uma função (conteúdo que faz parte da disciplina Cálculo III) para desenvolver métodos numéricos que serão aplicados a problemas que envolvem equações diferenciais ordinárias (EDO) de primeira ordem (conteúdo que também faz parte da disciplina Cálculo III). Além disso, as técnicas de integração numérica também serão utilizadas para o desenvolvimento de métodos que resolverão numericamente um Problema de Valor Inicial (PVI) de primeira ordem.

Resolver um Problema de Valor Inicial (PVI) de primeira ordem consiste em resolver uma EDO de primeira ordem para a qual a função incógnita ( $y(t)$ ) é conhecida em um ponto ( $y(t_0) = y_0$ , valor inicial):

$$y'(t) = f(t, y(t)) \text{ e } y(t_0) = y_0 \in \mathbb{R},$$

onde  $f: I \times \mathbb{R} \rightarrow \mathbb{R}$  é uma função real dada;  $y: I \rightarrow \mathbb{R}$  é a função incógnita; o intervalo  $I$  é dado por  $I = (t_0 - d, t_0 + d)$ , onde  $d > 0$  é um número real escolhido adequadamente, de modo a garantir a existência de solução da EDO.

**Observação.** Não deixe de rever os tópicos relacionados à série de Taylor e a equações diferenciais de primeira ordem (leia, por exemplo, o material didático produzido para a disciplina Cálculo III).

O desenvolvimento de métodos numéricos baseados em série de Taylor, aplicados em PVI, contém as seguintes etapas:

- (i) Considere um Intervalo  $I$  da forma  $[t_0, T]$ .
- (ii) Construa uma partição de  $I$  com  $n$  pontos igualmente espaçados:  $t_i = t_0 + ih$ ,  $0 \leq i \leq n$ , onde  $h = (T - t_0)/n$ .
- (iii) Considere a fórmula de Taylor da função  $y$  em torno do ponto  $t_i = t_{i-1} + h$ :



$$y(t_i) = y(t_{i-1}) + hy'(t_{i-1}) + (1/2!)h^2y''(t_{i-1}) + \dots + (1/k!)h^ky^{(k)}(t_{i-1}) + \text{Erro},$$

onde  $\text{Erro} = [1/(k+1)!]h^{k+1}y^{(k+1)}(c_{i-1}) = O(h^{k+1})$ , onde  $c_{i-1} \in (t_{i-1}, t_i)$ .

(iv) Despreze o erro na expressão anterior; troque  $y(t_i)$  e  $y(t_{i-1})$  pelos valores aproximados  $y_i$  e  $y_{i-1}$ ; para cada  $k$ , troque a derivadas  $y^{(k)}(t_{i-1})$  por  $\frac{d^{k-1}f(t_{i-1}, y_{i-1})}{dt^{k-1}}$ .

Lembre-se de que  $y'(t) = f(t, y(t))$ .

**Observação:** Seja  $g$  uma função de uma variável real  $h$ . Diz-se que  $g(h) = O(h^p)$  quando  $\lim_{h \rightarrow 0} \frac{g(h)}{h^p}$  existe e é igual a uma constante.

O método mais simples obtido com o procedimento anterior é o método de Euler. Nesse caso, a fórmula de Taylor é dada por  $y(t_i) = y(t_{i-1}) + hy'(t_{i-1}) + \text{Erro}$ , onde  $\text{Erro} = [1/(2!)]h^2y''(c_{i-1}) = O(h^2)$  e  $c_{i-1} \in (t_{i-1}, t_i)$ . Seguindo o passo (iv) anterior, obtém-se o método de Euler:

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}), \quad 1 \leq i \leq n.$$

O erro local do método de Euler é o erro de truncamento da série de Taylor da função  $y$ . No cálculo do Erro Local, considera-se que  $y(t_{i-1}) = y_{i-1}$ . Como o Erro Local do método de Euler é  $O(h^2)$ , diz-se que este método tem ordem 1.

Um método de ordem 2 e passo 2 (precisará de 2 valores iniciais:  $y_0$  e  $y_1$ ) é facilmente obtido considerando-se o seguinte desenvolvimento de Taylor:

$$\begin{aligned} y(t_i) - y(t_{i-2}) &= y(t_{i-1} + h) - y(t_{i-1} - h) = y(t_{i-1}) + hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_1 - \{y(t_{i-1}) - \\ &hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_2\} = 2hy'(t_{i-1}) + \text{Erro}, \text{ onde } \text{Erro} = E_1 - E_2 = \frac{h^3}{3!} \{y'''(c_{i-1}) + \\ &y'''(d_{i-1})\} = O(h^3); \quad c_{i-1} \in (t_{i-1}, t_i) \text{ e } d_{i-1} \in (t_{i-2}, t_{i-1}). \end{aligned}$$

Seguindo o passo (iv) anterior, obtém-se o Método do Ponto Médio ou Regra do Ponto Médio (livro de Cálculo Numérico, da Neide Bertoldi Franco – referência [6] do Módulo 1):

$$y_i = y_{i-2} + 2h.f(t_{i-1}, y_{i-1}), \quad 2 \leq i \leq n.$$

O erro local do Método do Ponto Médio é o erro de truncamento do desenvolvimento de Taylor anterior. No cálculo do Erro Local, considera-se  $y(t_{i-1}) = y_{i-1}$  e  $y(t_{i-2}) = y_{i-2}$ . Como o Erro Local do Método do Ponto Médio é  $O(h^3)$ , diz-se que este método tem ordem 2.

Outros autores, como Chapra e Canale, Burden e Faires, atribuem o nome de Método do Ponto Médio ao seguinte esquema:

$$y_i = y_{i-1} + h \cdot f(t_{i-1} + h/2, y_{i-1} + (h/2) \cdot f(t_{i-1}, y_{i-1})), \quad 1 \leq i \leq n,$$

que é um método de Runge-Kutta de ordem 2, também conhecido como Método de Euler Modificado.



### Referências

- [1] BURDEN, R.L., FAIRES, J.D., *Análise Numérica* - tradução da 8ª edição norte-americana, São Paulo, Cengage Learning, 721 p., 2008.
- [2] CHAPRA, S.C., CANALE, R.P., *Métodos Numéricos para Engenharia*, São Paulo, McGraw-Hill, 809 p., 2008.

## 2. Objetivos e conteúdo do Módulo 4

Desenvolver métodos numéricos baseados em série de Taylor e em integração numérica para resolver numericamente Problemas de Valor Inicial. Para você ficar familiarizado com os métodos numéricos apresentados neste módulo, não deixe de resolver os exercícios propostos e não deixe de ler as referências bibliográficas indicadas neste texto e procure outras referências que você achar conveniente.

### Conteúdos básicos do Módulo 4

- Métodos Baseados em Série de Taylor para a resolução de Problema de Valor Inicial de 1ª ordem.
- Métodos Baseados em Integração Numérica para resolução de PVI de 1ª ordem.

## 3. Métodos Baseados em Série de Taylor

### 3.1 Método de Euler

Como foi dito na introdução deste módulo, o Método de Euler é o método mais simples que pode ser aplicado a um PVI. Nesse caso, de acordo com os passos (i – iv) descritos na Seção 1 anterior, a expansão de Taylor da função incógnita é dada por  $y(t_i) = y(t_{i-1}) + hy'(t_{i-1}) + [1/(2!)]h^2y''(c_{i-1})$ ,  $c_{i-1} \in (t_{i-1}, t_i)$  e o método de Euler é dado por

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}), \quad 1 \leq i \leq n.$$

### Exemplo 1. (Método de Euler)

Considere o seguinte PVI:

$$y'(t) = f(t, y(t)) = t + y; \quad y(1) = y(t_0) = y_0 = 2.$$

Seja  $h = 0.05$ . Obtenha uma aproximação da solução em  $T = 1.2$ .

A equação anterior é uma clássica EDO linear de primeira ordem cuja solução analítica é conhecida. A técnica do fator integrante (veja a observação (\*) adiante) é utilizada para se obter a solução de uma EDO deste tipo. No exemplo em questão, a solução analítica é dada por:  $y(t) = 4e^{(t-1)} - t - 1$ . A solução analítica vai nos auxiliar no quesito relacionado à precisão do método numérico.

Este Problema de Valor Inicial será resolvido, primeiramente, com o Método de Euler e irá nos acompanhar até o final deste módulo 3. Com o intuito de apresentar uma sequência evolutiva dos métodos numéricos aplicados a PVI, este primeiro exemplo será utilizado como modelo de comparação entre os diversos métodos exibidos aqui. A eficiência de cada método será caracterizada basicamente por dois fatores: o esforço computacional e a precisão da solução.

De acordo com o enunciado do problema, tem-se que

(i)  $I = [t_0, T] = [1, 1.2]$ ;

(ii)  $h = (T - t_0)/n \rightarrow n = (1.2 - 1)/0.05 = 20/5 = 4$ ; logo  $t_i = t_0 + ih$ ,  $0 \leq i \leq 4$ , ou seja,

$$t_0 = 1; \quad t_1 = 1.05; \quad t_2 = 1.1; \quad t_3 = 1.15; \quad t_4 = 1.2.$$

Sabendo que  $f(t, y) = t + y$ , o método de Euler,  $y_i = y_{i-1} + hf(t_{i-1}, y_{i-1})$ , dará origem aos seguintes valores:

$$y_1 = y_0 + hf(t_0, y_0) = 2 + 0.05(t_0 + y_0) = 2 + 0.05(1 + 2) = 2.15;$$

$$y_2 = y_1 + hf(t_1, y_1) = 2.15 + 0.05(t_1 + y_1) = 2.15 + 0.05(1.05 + 2.15) = 2.31;$$

$$y_3 = y_2 + hf(t_2, y_2) = 2.31 + 0.05(t_2 + y_2) = 2.31 + 0.05(1.1 + 2.31) = 2.4805;$$

$$y_4 = y_3 + hf(t_3, y_3) = 2.4805 + 0.05(t_3 + y_3) = 2.4805 + 0.05(1.15 + 2.4805) = 2.662025.$$

Portanto,  $y(1.2) = y(t_4) \approx y_4 = 2.662025$ . Note que a solução analítica,  $y(t) = 4e^{(t-1)} - t - 1$ , fornece o seguinte valor:  $y(1.2) = 2.685611033$ . ■



Saiba Mais

(\*) **Observação:**  $y'(t) = a(t) + b(t)y(t) \rightarrow y'(t) - b(t)y(t) = a(t)$ ;

$$\Phi(t) = e^{-\int_{t_0}^t b(x) dx} \rightarrow \frac{d(y(t)\Phi(t))}{dt} = y'(t)\Phi(t) - b(t)y(t)\Phi(t) = a(t)\Phi(t)$$

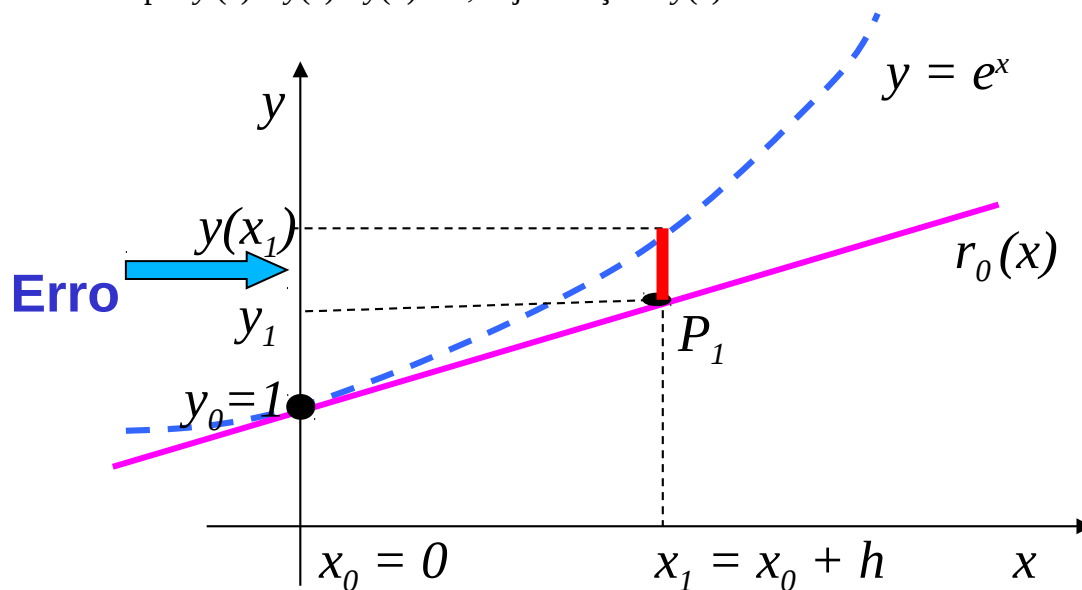
$$\rightarrow y(t) = e^{\int_{t_0}^t b(x) dx} \left\{ \int_{t_0}^t a(x)\Phi(x) dx + y_0 \right\}.$$

A interpretação geométrica exibida a seguir foi retirada do projeto “Aperfeiçoamento das Técnicas de Ensino-Aprendizagem da Disciplina Cálculo Numérico”, no âmbito do Programa Institucional de Bolsas de Graduação da UFU – PIBEG, realizado pela equipe composta por quatro professores da FAMAT: Sezimária de Fátima Pereira Saramago (coordenadora do projeto), Alessandro Alves Santana (colaborador); Célia Aparecida Zorzo Barcelos (orientadora) e César Guilherme de Almeida (orientador) e por dois alunos de graduação: Paulo Balduino Flabes Neto (Faculdade de Engenharia Mecânica) e Warlisson de Inácio Miranda (Faculdade de Matemática). O projeto foi realizado no período de 01 de outubro de 2007 a 31 de outubro de 2008.

Todo o material didático produzido pela equipe do projeto citado anteriormente está disponível no seguinte endereço eletrônico (na página da FAMAT): <http://www.portal.famat.ufu.br/node/278>.

## Interpretação Geométrica do Método de Euler

Considere a reta  $r_0(x)$  que passa por  $(x_0, y_0)$  e tem coeficiente angular  $y'(x_0) = f(x_0, y_0)$ , ou seja  $r_0(x) = y_0 + (x - x_0)f(x_0, y_0)$ . No Método de Euler,  $y(x_1) \approx r_0(x_1) = y_0 + (x_1 - x_0)f(x_0, y_0) = y_0 + hf(x_0, y_0) = y_1$ . A **Figura 1** abaixo exibe esta aproximação para o caso em que o PVI é dado por  $y'(x) = y(x)$  e  $y(0) = 1$ , cuja solução é  $y(x) = e^x$ .



**Figura 1 – Módulo 4 – Interpretação geométrica do método de Euler**

No caso geral, considera-se a reta  $r_{i-1}(x)$  que passa pelo ponto  $(x_{i-1}, y_{i-1})$  e que tem coeficiente angular  $f(x_{i-1}, y_{i-1})$ , isto é,  $r_{i-1}(x) = y_{i-1} + (x - x_{i-1})f(x_{i-1}, y_{i-1})$ . A aproximação de  $y(x_i)$  é dada por  $r_{i-1}(x_i)$ , ou seja,  $y(x_i) \approx r_{i-1}(x_i) = y_{i-1} + (x_i - x_{i-1})f(x_{i-1}, y_{i-1}) = y_{i-1} + hf(x_{i-1}, y_{i-1}) = y_i$ .

## Erro Global do Método de Euler

Já foi dito na Seção 1 que o erro local do método de Euler é o erro de truncamento da série de Taylor da função  $y$ . Isto acontece porque, no cálculo do Erro Local ( $E_{\text{loc}}$ ), considera-se que  $y(t_{i-1}) = y_{i-1}$ . Assim, da equação diferencial dada no PVI, tem-se que  $f(t_{i-1}, y_{i-1}) = f(t_{i-1}, y(t_{i-1})) = y'(t_{i-1})$ . Portanto,

$$\begin{aligned} E_{\text{loc}} &= |y(t_i) - y_i| = |y(t_{i-1}) + hy'(t_{i-1}) + [1/(2!)]h^2y''(c_{i-1}) - y_{i-1} - hf(t_{i-1}, y_{i-1})| \rightarrow \\ \rightarrow E_{\text{loc}} &= |[1/(2!)]h^2y''(c_{i-1})| = O(h^2), \quad c_{i-1} \in (t_{i-1}, t_i). \end{aligned}$$

Como o Erro Local do método de Euler é  $O(h^2)$ , diz-se que este método tem ordem 1.

Para a análise do erro global do Método de Euler, serão consideradas hipóteses sobre a função do PVI,  $f(t, y)$ , e sobre a derivada segunda da função incógnita,  $y(t)$ : (i) Suponha que  $f$  seja uma função Lipschitziana na segunda coordenada ( $y$ ), ou seja, suponha que exista uma constante positiva  $L$  tal que  $|f(t, y) - f(t, Y)| \leq L|y - Y|$ , quaisquer que sejam  $t \in [t_0, T]$ ,  $y$  e  $Y$  reais. (ii) Suponha, também, que  $|(1/2)y''(t)| \leq \kappa$ , para todo  $t \in [t_0, T]$ .

**Observação:** A demonstração do Teorema que será enunciado a seguir, utilizará o seguinte resultado:

$$\begin{aligned} S &= \mu^{n-1} + \mu^{n-2} + \dots + \mu^2 + \mu + 1 \rightarrow \mu S = \mu^n + \mu^{n-1} + \dots + \mu^3 + \mu^2 + \mu \rightarrow \\ \rightarrow S - \mu S &= 1 - \mu^n \rightarrow S(1 - \mu) = 1 - \mu^n \rightarrow S = \frac{1 - \mu^n}{1 - \mu}. \end{aligned}$$

**Teorema 1. (Convergência do Método de Euler)** De acordo com as hipóteses anteriores, o Método de Euler é convergente. Isto é, o Erro Global,  $E_{\text{glob}} = |y(t_n) - y_n|$ , tende a zero quando  $h = (T - t_0)/n$  tende a zero (ou, equivalentemente, quando  $n$  tende a infinito).

**Demonstração:** Lembre-se de que  $y(t_i) = y(t_{i-1}) + hy'(t_{i-1}) + (1/2)h^2y''(c_{i-1}) = y(t_{i-1}) + hf(t_{i-1}, y(t_{i-1})) + (1/2)h^2y''(c_{i-1})$ ,  $c_{i-1} \in (t_{i-1}, t_i)$ . Além disso,  $y_i = y_{i-1} + hf(t_{i-1}, y_{i-1})$ ,  $1 \leq i \leq n$ , e  $y(t_0) = y_0$ . Assim,

$$\begin{aligned} |y(t_1) - y_1| &= |y(t_0) + hy'(t_0) + (1/2)h^2y''(c_0) - y_0 - hf(t_0, y_0)| \rightarrow \\ \rightarrow |y(t_1) - y_1| &= |h^2(1/2)y''(c_0)| \leq h^2\kappa, \text{ pois } c_0 \in (t_0, t_1). \end{aligned}$$

Repetindo o raciocínio anterior e usando as hipóteses dadas no teorema, obtém-se que

$$|y(t_2) - y_2| = |y(t_1) + hf(t_1, y(t_1)) + (1/2)h^2y''(c_1) - y_1 - hf(t_1, y_1)| \rightarrow$$

$$\rightarrow |y(t_2) - y_2| \leq |y(t_1) - y_1| + hL|y(t_1) - y_1| + (1/2)h^2\kappa, \text{ pois } c_1 \in (t_1, t_2).$$

Como  $|y(t_1) - y_1| \leq h^2\kappa$  então  $|y(t_2) - y_2| \leq [1 + hL]h^2\kappa + h^2\kappa$ , ou seja

$$|y(t_2) - y_2| \leq h^2\kappa \{[1 + hL] + 1\}.$$

Continuando o processo, tem-se que

$$|y(t_3) - y_3| = |y(t_2) + hf(t_2, y(t_2)) + (1/2)h^2y''(c_2) - y_2 - hf(t_2, y_2)| \rightarrow$$

$$\rightarrow |y(t_3) - y_3| \leq |y(t_2) - y_2| + hL|y(t_2) - y_2| + h^2\kappa, \text{ pois } c_2 \in (t_2, t_3).$$

Utilizando a desigualdade obtida no passo anterior, tem-se que

$$|y(t_3) - y_3| \leq [1 + hL]\{[1 + hL] + 1\}h^2\kappa + h^2\kappa = h^2\kappa\{[1 + hL]^2 + [1 + hL] + 1\}.$$

Por indução finita, suponha que

$$|y(t_j) - y_j| \leq h^2\kappa \{[1 + hL]^{j-1} + [1 + hL]^{j-2} + \dots + 1\}, 1 \leq j < n. \text{ Portanto,}$$

$$|y(t_{j+1}) - y_{j+1}| = |y(t_j) + hf(t_j, y(t_j)) + (1/2)h^2y''(c_j) - y_j - hf(t_j, y_j)| \rightarrow$$

$$\rightarrow |y(t_{j+1}) - y_{j+1}| \leq |y(t_j) - y_j| [1 + hL] + h^2\kappa, \text{ pois } c_j \in (t_j, t_{j+1}).$$

Pela hipótese de indução, segue-se que

$$|y(t_{j+1}) - y_{j+1}| \leq h^2\kappa \{[1 + hL]^{j-1} + [1 + hL]^{j-2} + \dots + [1 + hL] + 1\}[1 + hL] + h^2\kappa \rightarrow$$

$$|y(t_{j+1}) - y_{j+1}| \leq h^2\kappa \{[1 + hL]^j + [1 + hL]^{j-1} + \dots + [1 + hL]^2 + [1 + hL] + 1\}.$$

Seja  $\mu = 1 + hL$ . Então, por indução finita,

$$|y(t_n) - y_n| \leq h^2\kappa \{ \mu^{n-1} + \mu^{n-2} + \dots + \mu^2 + \mu + 1 \} = h^2\kappa \frac{1 - \mu^n}{1 - \mu} = h\kappa \frac{(1 + hL)^n - 1}{L},$$

pois  $\mu - 1 = hL$ .

Como  $h = (T - t_0)/n$ , então

$$0 \leq |y(t_n) - y_n| \leq \frac{T - t_0}{n} \kappa \frac{\left(1 + \frac{(T - t_0)L}{n}\right)^n - 1}{L}.$$

Note que  $\lim_{n \rightarrow +\infty} \left(1 + \frac{(T-t_0)L}{n}\right)^n = e^{(T-t_0)L}$  e  $\lim_{n \rightarrow +\infty} \frac{(T-t_0)L}{n} = 0$ . Portanto,

$$\lim_{n \rightarrow +\infty} |y(t_n) - y_n| = 0.$$



### 3.2 O Método do Ponto Médio

Como foi mostrado na Introdução (Seção 1), o Método do Ponto Médio é um método de ordem 2 e, portanto, tem uma ordem de precisão a mais do que o Método de Euler. Além disso, é um método de passo múltiplo; de passo 2, para ser mais específico, ou seja, o esquema numérico precisará de 2 valores para ser iniciado:  $y_0$  e  $y_1$ .

A dedução deste método é feita através do seguinte desenvolvimento de Taylor:

$$\begin{aligned} y(t_i) - y(t_{i-2}) &= y(t_{i-1} + h) - y(t_{i-1} - h) = y(t_{i-1}) + hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_1 - \{y(t_{i-1}) - \\ &hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_2\} = 2hy'(t_{i-1}) + \text{Erro}, \text{ onde Erro} = E_1 - E_2 = \frac{h^3}{3!} \{y'''(c_{i-1}) + \\ &y'''(d_{i-1})\} = O(h^3); c_{i-1} \in (t_{i-1}, t_i) \text{ e } d_{i-1} \in (t_{i-2}, t_{i-1}). \end{aligned}$$

Como foi mostrado na Seção 1, o Método do Ponto Médio é dado por:

$$y_i = y_{i-2} + 2h.f(t_{i-1}, y_{i-1}), \quad 2 \leq i \leq n.$$

Note que o valor de  $y_1$  deve ser calculado por algum método que tenha, preferencialmente, a mesma ordem do Método do Ponto Médio. Isso evita a introdução de imprecisões numéricas logo no início do método.

Como, até este momento, não foi deduzido formalmente nenhum método de ordem 2, auto-iniciante (método de passo simples que necessita apenas do valor  $y_0$ , dado no PVI, para ser iniciado), o próximo exemplo exibirá uma aplicação do Método do Ponto Médio que utilizará o valor de  $y_1$  calculado pelo Método de Euler (que possui ordem 1).

**Exemplo 2. (Método do Ponto Médio)** Considere o seguinte PVI:

$$y'(t) = f(t, y(t)) = t + y; \quad y(1) = y(t_0) = y_0 = 2.$$

Seja  $h = 0.05$ . Será obtida uma aproximação da solução em  $T = 1.2$ , utilizando o Método do Ponto Médio e  $y_1 = 2.15$  (veja o **Exemplo 1** da Seção 3.1 deste módulo 4).

De acordo com o enunciado do problema, tem-se que

(i)  $I = [t_0, T] = [1, 1.2]$ ;

(ii)  $h = (T - t_0)/n \rightarrow n = (1.2 - 1)/0.05 = 20/5 = 4$ ; logo  $t_i = t_0 + ih$ ,  $0 \leq i \leq 4$ , ou seja,

$$t_0 = 1; \quad t_1 = 1.05; \quad t_2 = 1.1; \quad t_3 = 1.15; \quad t_4 = 1.2.$$

Sabendo que  $f(t, y) = t + y$ , o método do Ponto Médio,  $y_i = y_{i-2} + 2hf(t_{i-1}, y_{i-1})$ , dará origem aos seguintes valores:

$$y_2 = y_0 + 2hf(t_1, y_1) = 2 + 0.1(t_1 + y_1) = 2 + 0.1(1.05 + 2.15) = 2.32;$$

$$y_3 = y_1 + 2hf(t_2, y_2) = 2.15 + 0.1(t_2 + y_2) = 2.15 + 0.1(1.1 + 2.32) = 2.492;$$

$$y_4 = y_2 + 2hf(t_3, y_3) = 2.32 + 0.1(t_3 + y_3) = 2.32 + 0.1(1.15 + 2.492) = 2.6842.$$

Portanto,  $y(1.2) = y(t_4) \approx y_4 = 2.6842$ . Este valor é mais preciso do que o obtido pelo Método de Euler; basta comparar com o valor fornecido pela solução analítica,  $y(t) = 4e^{(t-1)} - t - 1$ ,  $y(1.2) = 2.685611033$ . ■

## Erro Global do Método do Ponto Médio

Para a análise do erro global do Método do Ponto Médio, serão consideradas hipóteses sobre a função do PVI,  $f(t, y)$ , sobre a derivada terceira da função incógnita,  $y(t)$ , e sobre a aproximação inicial de  $y(t_1)$ , denotada por  $y_1$ : (i) Suponha que  $f$  seja uma função Lipschitziana na segunda coordenada ( $y$ ), ou seja, suponha que exista uma constante positiva  $L$  tal que  $|f(t, y) - f(t, Y)| \leq L|y - Y|$ , quaisquer que sejam  $t \in [t_0, T]$ ,  $y$  e  $Y$  reais. (ii) Suponha, também, que  $|(1/3)y'''(t)| \leq \kappa_1$ , para todo  $t \in [t_0, T]$ . (iii)  $|y(t_1) - y_1| \leq Ch^3$ .

**Teorema 2. (Convergência do Método do Ponto Médio)** De acordo com as hipóteses anteriores, o Método do Ponto Médio é convergente. Isto é, o Erro Global,  $E_{\text{glob}} = |y(t_n) - y_n|$ , tende a zero quando  $h = (T - t_0)/n$  tende a zero (ou, equivalentemente, quando  $n$  tende a infinito).

**Demonstração:** Lembre-se de que  $y(t_i) - y(t_{i-2}) = y(t_{i-1} + h) - y(t_{i-1} - h) = y(t_{i-1}) + hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_1 - \{y(t_{i-1}) - hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + E_2\} = 2hy'(t_{i-1}) + \text{Erro}$ , onde

$\text{Erro} = E_1 - E_2 = \frac{h^3}{3!} \{y'''(c_{i-1}) + y'''(d_{i-1})\} = O(h^3)$ ;  $c_{i-1} \in (t_{i-1}, t_i)$  e  $d_{i-1} \in (t_{i-2}, t_{i-1})$ . Além disso,  $y_i = y_{i-2} + 2hf(t_{i-1}, y_{i-1})$ ,  $2 \leq i \leq n$ ;  $y(t_0) = y_0$  e  $y_1$  foi obtido por um método de ordem 2, ou seja,  $|y(t_1) - y_1| \leq Ch^3$ . Assim,



$$|y(t_i) - y_i| = |y(t_{i-2}) + 2h f(t_{i-1}, y(t_{i-1})) + \frac{h^3}{3!} \{y'''(c_{i-1}) + y'''(d_{i-1})\} - y_{i-2} + 2h f(t_{i-1}, y_{i-1})| \rightarrow$$

$$\rightarrow |y(t_i) - y_i| \leq |y(t_{i-2}) - y_{i-2}| + 2hL|y(t_{i-1}) - y_{i-1}| + h^3\kappa_1.$$

Assim,

$$|y(t_2) - y_2| \leq |y(t_0) - y_0| + 2hL|y(t_1) - y_1| + h^3\kappa_1 \leq 2hLCh^3 + h^3\kappa_1 \leq h^3\kappa\{2hL + 1\},$$

onde  $\kappa = \text{máximo}\{C, \kappa_1\}$ .

Repetindo o raciocínio anterior, obtém-se que

$$|y(t_3) - y_3| \leq |y(t_1) - y_1| + 2hL|y(t_2) - y_2| + h^3\kappa_1 \leq Ch^3 + 2hL\{h^3\kappa[2hL + 1]\} + h^3\kappa_1 \leq$$

$$h^3\kappa + h^3\kappa\{(2hL)^2 + 2hL\} + h^3\kappa \rightarrow$$

$$\rightarrow |y(t_3) - y_3| \leq h^3\kappa + h^3\kappa\{(2hL)^2 + 2hL + 1\}.$$

Continuando o processo, tem-se que

$$|y(t_4) - y_4| \leq |y(t_2) - y_2| + 2hL|y(t_3) - y_3| + h^3\kappa_1 \leq h^3\kappa\{2hL + 1\} + 2hL\{h^3\kappa +$$

$$h^3\kappa[(2hL)^2 + 2hL + 1]\} + h^3\kappa_1 \leq h^3\kappa\{2hL + 1\} + 2hLh^3\kappa + h^3\kappa\{(2hL)^3 + (2hL)^2 +$$

$$2hL\} + h^3\kappa \leq h^3\kappa\{2hL + 1\} + 2hLh^3\kappa + h^3\kappa\{(2hL)^3 + (2hL)^2 + 2hL + 1\} \rightarrow$$

$$\rightarrow |y(t_4) - y_4| \leq p(h) + h^3\kappa\{(2hL)^3 + (2hL)^2 + 2hL + 1\},$$

onde  $p(h) = h^3\kappa\{2hL + 1\} + 2hLh^3\kappa$ . Portanto,  $\lim_{h \rightarrow 0} p(h) = 0$ .

Considere  $h < 1/(2L)$  ou, equivalentemente,  $n > (T - t_0)2L$ . Seja  $\zeta = 2hL < 1$ . Por indução finita, suponha que

$$|y(t_j) - y_j| \leq p^{(j)}(h) + h^3\kappa\{\zeta^{j-1} + \zeta^{j-2} + \dots + 1\}, \quad 2 \leq j < n, \text{ tal que } \lim_{h \rightarrow 0} p^{(j)}(h) = 0.$$

Portanto,

$$|y(t_{j+1}) - y_{j+1}| \leq |y(t_j) - y_j| + 2hL|y(t_j) - y_j| + h^3\kappa_1 \leq p^{(j)}(h) + h^3\kappa\{\zeta^{j-2} + \zeta^{j-3} + \dots + 1\}$$

$$+ \zeta\{p^{(j-1)}(h) + h^3\kappa[\zeta^{j-1} + \zeta^{j-2} + \dots + 1]\} + h^3\kappa \leq p^{(j)}(h) + h^3\kappa\{\zeta^{j-2} + \zeta^{j-3} + \dots + 1\} +$$

$$\zeta p^{(j-1)}(h) + h^3\kappa\{\zeta^j + \zeta^{j-1} + \dots + 1\}. \text{ Além disso, } \lim_{h \rightarrow 0} p^{(j)}(h) = \lim_{h \rightarrow 0} p^{(j-1)}(h) = 0.$$

Assim,

$$|y(t_{j+1}) - y_{j+1}| \leq p^{(j+1)}(h) + h^3 \kappa \{\zeta^j + \zeta^{j-1} + \dots + 1\},$$

onde,  $p^{(j+1)}(h) = p^{(j)}(h) + h^3 \kappa \{\zeta^{j-2} + \zeta^{j-3} + \dots + 1\} + \zeta p^{(j-1)}(h)$ .

Observe que  $1 \leq S(\zeta) = \zeta^{j-2} + \zeta^{j-3} + \dots + 1 = (1 - \zeta^{j-1})/(1 - \zeta) \leq 1/(1 - \zeta)$ , pois  $0 \leq \zeta < 1$ . Como  $\zeta = 2Lh$ , então, pelo teorema do confronto,  $\lim_{h \rightarrow 0} S(\zeta) = 1$ . Portanto,  $\lim_{h \rightarrow 0} p^{(j+1)}(h) = 0$ .

Por indução finita, segue-se que  $0 \leq |y(t_n) - y_n| \leq p^{(n)}(h) + h^3 \kappa \{\zeta^{n-1} + \zeta^{n-2} + \dots + 1\}$ .

Como  $h = (T - t_0)/n$ , então  $\lim_{h \rightarrow 0} p^{(n)}(h) = 0 \rightarrow \lim_{n \rightarrow +\infty} p^{(n)}\left(\frac{T - t_0}{n}\right) = 0$ . De fato,

dado  $\varepsilon > 0$  existe  $\delta > 0$  tal que  $|h| < \delta \rightarrow |p^{(n)}(h)| < \varepsilon$ . Seja  $n_0 > \text{máximo}\{(T - t_0)/\delta, (T - t_0)2L\}$ . Assim, se  $n > n_0$  então  $h < \delta \rightarrow |p^{(n)}(h)| < \varepsilon \rightarrow |p^{(n)}[(T - t_0)/\delta]| < \varepsilon$ . Observe que:  $n > n_0 \rightarrow h < 1/(2L) \rightarrow \zeta < 1$ .

Agora, observe que  $1 \leq S(\zeta) = \zeta^{n-1} + \zeta^{n-2} + \dots + 1 = (1 - \zeta^n)/(1 - \zeta) \leq 1/(1 - \zeta)$ . Como  $\zeta = 2Lh = 2L(T - t_0)/n$ , então  $\lim_{n \rightarrow +\infty} S(\zeta) = 1$ . Além disso,  $\lim_{n \rightarrow +\infty} \left(\frac{(T - t_0)^3}{n^3}\right) = 0$ . Dessa forma,  $\lim_{n \rightarrow +\infty} |y(t_n) - y_n| = 0$ . □

### 3.3 O Método de Taylor de Ordem 2

O Método de Taylor de ordem dois é deduzido a partir da seguinte expansão:

$$y(t_i) = y(t_{i-1} + h) = y(t_{i-1}) + hy'(t_{i-1}) + \frac{h^2}{2!} y''(t_{i-1}) + \frac{h^3}{3!} y'''(c_{i-1}), c_{i-1} \in (t_{i-1}, t_i).$$

Sabendo que  $y'(t) = f(t, y(t))$  e utilizando a regra de derivação conhecida como Regra da Cadeia (\*), tem-se que  $y''(t) = f_t(t, y(t)) + f_y(t, y(t)).y'(t)$ , onde  $f_t = \frac{\partial f}{\partial t}$ ,  $f_y = \frac{\partial f}{\partial y}$  e  $y'(t) = f(t, y(t))$ .



(\*) Para obter informações sobre a Regra da Cadeia, consulte as referências abaixo e o material didático desenvolvido para a disciplina Cálculo III.

[3] BOULOS, P. *Introdução ao cálculo*. Volume 2. São Paulo, Edgard Blucher Ltda, 1974.

[4] LEITHOLD, L. *O Cálculo com geometria analítica*. 2 Volumes, 3ª Edição. São Paulo, Harbra, 1994.

Na expansão de Taylor dada anteriormente, desprezando-se o erro ( $\frac{h^3}{3!} y'''(c_{i-1}) = O(h^3)$ ,  $c_{i-1} \in (t_{i-1}, t_i)$ ) e trocando-se os valores exatos pelos valores aproximados (veja os passos (i – iv) apresentados na Seção 1 deste módulo), obtém-se o Método de Taylor de ordem 2:

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) + \frac{h^2}{2!} \{f_t(t_{i-1}, y_{i-1}) + f_y(t_{i-1}, y_{i-1})f(t_{i-1}, y_{i-1})\}, 1 \leq i \leq n.$$

O erro local do Método de Taylor de ordem 2 é o erro de truncamento da série de Taylor da função  $y$ . Isso acontece porque no cálculo do Erro Local ( $E_{loc}$ ) considera-se que  $y(t_{i-1}) = y_{i-1}$ . Assim, da equação diferencial dada no PVI, tem-se que  $f(t_{i-1}, y_{i-1}) = f(t_{i-1}, y(t_{i-1})) = y'(t_{i-1})$ ;  $f_t(t_{i-1}, y_{i-1}) = f_t(t_{i-1}, y(t_{i-1}))$  e  $f_y(t_{i-1}, y_{i-1}) = f_y(t_{i-1}, y(t_{i-1}))$ . Portanto,

$$E_{loc} = |y(t_i) - y_i| = |O(h^3)| = \frac{h^3}{3!} |y'''(c_{i-1})|, c_{i-1} \in (t_{i-1}, t_i).$$

Como o Erro Local é  $O(h^3)$ , diz-se que este Método de Taylor tem ordem 2.

**Observação:** O Método de Taylor de ordem 2, bem como o Método de Euler, é um método auto-iniciante, ou seja, a partir do valor inicial  $y_0$ , fornecido no PVI, é possível calcular todas as aproximações de  $y(t_i)$ ,  $y_i$ ,  $1 \leq i \leq n$ .

**Observação:** A grande desvantagem do Método de Taylor está relacionada ao esforço computacional devido aos cálculos envolvendo as derivadas parciais da função  $f$  dada no PVI. Como os métodos de Euler e do Ponto Médio não envolvem derivadas parciais e utilizam apenas a função  $f$  do PVI, demandam menos esforço computacional quando comparados ao Método de Taylor de ordem 2.

**Exemplo 3. (Método do Ponto Médio com valor  $y_1$  calculado pelo Método de Taylor de ordem 2)** Considere o seguinte PVI:

$$y'(t) = f(t, y(t)) = t + y; \quad y(1) = y(t_0) = y_0 = 2.$$

Seja  $h = 0.05$  e  $t_1 = t_0 + h = 1.05$ . O Método de Taylor de ordem 2 será utilizado para se calcular  $y_1$ , que é uma aproximação de  $y(t_1)$ . A aproximação da solução do PVI, em  $T = 1.2$ , será obtida através do Método do Ponto Médio.

De acordo com o enunciado do problema, tem-se que

(i)  $I = [t_0, T] = [1, 1.2]$ ;

(ii)  $h = (T - t_0)/n \rightarrow n = (1.2 - 1)/0.05 = 20/5 = 4$ ; logo,  $t_i = t_0 + ih$ ,  $0 \leq i \leq 4$ , ou seja,

$$t_0 = 1; \quad t_1 = 1.05; \quad t_2 = 1.1; \quad t_3 = 1.15; \quad t_4 = 1.2.$$

Sabendo que  $f(t, y) = t + y$ , então  $f_t(t, y) = 1$  e  $f_y(t, y) = 1$ . Portanto, o Método de Taylor de ordem 2 é dado por:  $y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) + \frac{h^2}{2!} \{1 + f(t_{i-1}, y_{i-1})\}$ . Assim,  $y_1 = y_0 + hf(t_0, y_0) + \frac{h^2}{2!} \{1 + f(t_0, y_0)\} = 2 + 0.05(1 + 2) + 0.00125\{1 + (1 + 2)\} = 2.155$ .

O método do Ponto Médio,  $y_i = y_{i-2} + 2hf(t_{i-1}, y_{i-1})$ , dará origem aos seguintes valores:

$$y_2 = y_0 + 2hf(t_1, y_1) = 2 + 0.1(t_1 + y_1) = 2 + 0.1(1.05 + 2.155) = 2.3205;$$

$$y_3 = y_1 + 2hf(t_2, y_2) = 2.155 + 0.1(t_2 + y_2) = 2.155 + 0.1(1.1 + 2.3205) = 2.49705;$$

$$y_4 = y_2 + 2hf(t_3, y_3) = 2.3205 + 0.1(t_3 + y_3) = 2.3205 + 0.1(1.15 + 2.49705) = 2.685205.$$

Portanto,  $y(1.2) = y(t_4) \approx y_4 = 2.685205$ . Este valor é mais preciso do que aquele obtido no Exemplo 2, no qual, também, considerou-se o Método do Ponto Médio, porém  $y_1$  foi calculado pelo Método de Euler de ordem 1. Lembre-se de que a solução analítica,  $y(t) = 4e^{(t-1)} - t - 1$ , fornece  $y(1.2) = 2.685611033$ . ■

O próximo exercício tem como propósito mostrar que o Método de Taylor de ordem 2 demanda mais contas (mais esforço computacional) do que o Método do Ponto Médio, embora os dois métodos tenham a mesma precisão (ordem 2).

**Exercício 1. (Método de Taylor de ordem 2)** Considere o mesmo problema exibido no Exemplo 3 anterior. Obtenha uma aproximação da solução do PVI, em  $T = 1.2$ , utilizando o Método de Taylor de ordem 2.

**Solução do Exercício 1:** Como foi exibido no exemplo anterior, o método de Taylor de ordem 2 tem a seguinte expressão:

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) + \frac{h^2}{2!} \{f_t(t_{i-1}, y_{i-1}) + f_y(t_{i-1}, y_{i-1}) \cdot f(t_{i-1}, y_{i-1})\} = y_{i-1} + hf(t_{i-1}, y_{i-1}) + \frac{h^2}{2!} \{1 + f(t_{i-1}, y_{i-1})\}.$$

Assim,

$$y_1 = y_0 + hf(t_0, y_0) + \frac{h^2}{2!} \{1 + f(t_0, y_0)\} = 2 + 0.05(1 + 2) + 0.00125\{1 + (1 + 2)\} = 2.155.$$

$$y_2 = y_1 + hf(t_1, y_1) + \frac{h^2}{2!} \{1 + f(t_1, y_1)\} = 2.155 + 0.05(1.05 + 2.155) + 0.00125\{1 + (1.05 + 2.155)\} = 2.32050625.$$

$$y_3 = y_2 + hf(t_2, y_2) + \frac{h^2}{2!} \{1 + f(t_2, y_2)\} = 2.32050625 + 0.05(1.1 + 2.32050625) + 0.00125\{1 + (1.1 + 2.32050625)\} = 2.497057195.$$

$$y_4 = y_3 + hf(t_3, y_3) + \frac{h^2}{2!} \{1 + f(t_3, y_3)\} = 2.497057195 + 0.05(1.15 + 2.497057195) + 0.00125\{1 + (1.15 + 2.497057195)\} = 2.685218876. \quad \blacksquare$$

### 3.4 Métodos de Runge – Kutta

A fórmula geral de um Método de Runge – Kutta de  $s$  estágios (s-RK) é dada por:

$$K_l = y_{i-1} + h \sum_{j=1}^s a_{l,j} f(t_{i-1} + hc_j, K_j); \quad 1 \leq l \leq s;$$

$$y_i = y_{i-1} + h \sum_{j=1}^s b_j f(t_{i-1} + hc_j, K_j); \quad 1 \leq i \leq n.$$

Os parâmetros  $a_{l,j}$ ,  $c_l = \sum_{j=1}^s a_{l,j}$  e  $b_j$  são determinados de modo a garantir a igualdade dos polinômios de Taylor, na variável  $h$ , oriundos do desenvolvimento de Taylor da função  $f$  que aparece tanto na expressão do Método de Runge – Kutta quanto na expressão do Método de Taylor de uma determinada ordem.

Os métodos s-RK dados pela fórmula geral exibida anteriormente são denominados métodos implícitos. Neste material didático, não serão deduzidos os métodos implícitos. Uma referência importante sobre o assunto é a seguinte:

BUTCHER, J.C. *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. New York, NY, USA, Wiley-Interscience, 1987.

Eu utilizei este livro em minha dissertação de Mestrado. Fui orientado pela professora Célia Maria Finazzi de Andrade, ICMC – USP – São Carlos. A professora Célia emprestou o livro dela, novinho em folha, para mim. Era o meu livro de cabeceira, era a minha Bíblia. Café da manhã, almoço e janta, o livro sempre a vista. Chorei, sorri, senti-me aliviado diante dele. Dissertação finalizada, defendida e aprovada, em fevereiro de 1993, devolvi o livro para a professora Célia Finazzi. Ela mostrava com orgulho, para a professora Neide Bertoldi, as folhas encardidas pelo manuseio. Se não me engano, a dissertação foi aprovada com louvor pelos membros da Banca Examinadora: Célia Finazzi, Neide Bertoldi Franco (ICMC – USP) e Sebastião Pereira Martins (IBILCE – UNESP – São José do Rio Preto).

**Observação:** No caso geral, os  $s$  estágios de um método de Runge – Kutta são obtidos através da resolução de um sistema de  $s$  equações não lineares. Este tipo de sistema não será abordado neste material, mas quem tiver interesse no assunto pode consultar, por exemplo, o livro da Vera Lúcia Lopes e Márcia Ruggiero (Bibliografia comentada na Seção 4.2 do módulo 2). Os parâmetros de um  $s$ -RK podem ser exibidos matricialmente como segue.

$c_1 = \sum_{j=1}^s a_{1,j}$	$a_{11}$	$a_{12}$	$a_{13}$	...	$a_{1s}$
$c_2 = \sum_{j=1}^s a_{2,j}$	$a_{21}$	$a_{22}$	$a_{23}$	...	$a_{2s}$
$c_3 = \sum_{j=1}^s a_{3,j}$	$a_{31}$	$a_{32}$	$a_{33}$	...	$a_{3s}$
$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$	$\cdot$ $\cdot$ $\cdot$
$c_s = \sum_{j=1}^s a_{s,j}$	$a_{s1}$	$a_{s2}$	$a_{s3}$	...	$a_{ss}$
	$b_1$	$b_2$	$b_3$	...	$b_s$

**Observação:** Uma classe especial de métodos  $s$ -RK é a dos métodos diagonalmente implícitos. Nesta classe de métodos, os  $s$  estágios também são implícitos, mas as equações do sistema não linear podem ser resolvidas sequencialmente. Primeiro obtém-

se o valor de  $K_1$  resolvendo-se uma equação não linear do tipo  $K_1 = F_1(K_1)$ . Dado  $K_1$  obtém-se o valor de  $K_2$  resolvendo-se a equação  $K_2 = F_2(K_1, K_2)$ , e assim sucessivamente. Dessa forma, dados os valores de  $K_1, \dots, K_{j-1}$ , o valor do estágio  $K_j$ ,  $2 < j \leq s$ , será obtido resolvendo-se a equação  $K_j = F_j(K_1, \dots, K_{j-1}, K_j)$ . A fórmula de um s-RK diagonalmente implícito é exibida a seguir.

$$K_l = y_{i-1} + h \sum_{j=1}^l a_{l,j} f(t_{i-1} + h c_j, K_j); \quad 1 \leq l \leq s;$$

$$y_i = y_{i-1} + h \sum_{j=1}^s b_j f(t_{i-1} + h c_j, K_j); \quad 1 \leq i \leq n.$$

A representação matricial dos parâmetros envolvidos nesse tipo de método é dada abaixo (note que a matriz com elementos  $a_{ij}$  é triangular inferior).

$c_1 = \sum_{j=1}^1 a_{1,j}$	$a_{11}$	0	0	...	0
$c_2 = \sum_{j=1}^2 a_{2,j}$	$a_{21}$	$a_{22}$	0	...	0
$c_3 = \sum_{j=1}^3 a_{3,j}$	$a_{31}$	$a_{32}$	$a_{33}$	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_s = \sum_{j=1}^s a_{s,j}$	$a_{s1}$	$a_{s2}$	$a_{s3}$	...	$a_{ss}$
	$b_1$	$b_2$	$b_3$	...	$b_s$

Na próxima seção serão deduzidos os Métodos de Runge – Kutta de ordem 2 e dois estágios explícitos. Serão exibidos, também, os métodos explícitos 3-RK e 4-RK de ordens 3 e 4, respectivamente.

## Métodos de Runge – Kutta explícitos

Um s-RK explícito tem a seguinte fórmula:

$$K_l = y_{i-1} + h \sum_{j=1}^{l-1} a_{l,j} f(t_{i-1} + h c_j, K_j); \quad 1 \leq l \leq s;$$

$$y_i = y_{i-1} + h \sum_{j=1}^s b_j f(t_{i-1} + h c_j, K_j); \quad 1 \leq i \leq n.$$

A representação matricial dos parâmetros envolvidos nesse método é exibida a seguir.

$c_1 = 0$	0	0	0	...	0	0
$c_2 = a_{21}$	$a_{21}$	0	0	...	0	0
$c_3 = \sum_{j=1}^2 a_{3,j}$	$a_{31}$	$a_{32}$	0	...	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_s = \sum_{j=1}^{s-1} a_{s,j}$	$a_{s1}$	$a_{s2}$	$a_{s3}$	...	$a_{s,s-1}$	0
	$b_1$	$b_2$	$b_3$	...	$b_{s-1}$	$b_s$

### Método de ordem 2 e 2 estágios

De acordo com a fórmula apresentada anteriormente, o método explícito de dois estágios pode ser reescrito como:

$$K_1 = y_{i-1}; \quad K_2 = y_{i-1} + ha_{21} f(t_{i-1} + hc_1, K_1) = y_{i-1} + ha_{21} f(t_{i-1}, y_{i-1});$$

$$y_i = y_{i-1} + h \{ b_1 f(t_{i-1}, y_{i-1}) + b_2 f(t_{i-1} + hc_2, y_{i-1} + ha_{21} f(t_{i-1}, y_{i-1})) \}, \quad 1 \leq i \leq n.$$

Considere o desenvolvimento de Taylor da função  $f$  em torno do ponto  $(t_{i-1}, y_{i-1})$  de modo que a expressão de  $y_i$  seja dada em função de um polinômio de grau 2 na variável  $h$ ; os termos  $O(h^3)$  farão parte do erro de truncamento da série de Taylor. Dessa forma,

$$y_i = y_{i-1} + h \{ b_1 f(t_{i-1}, y_{i-1}) + b_2 f(t_{i-1} + hc_2, y_{i-1} + ha_{21} f(t_{i-1}, y_{i-1})) \} = y_{i-1} + hb_1 f(t_{i-1}, y_{i-1})$$

$$+ h b_2 \{ f(t_{i-1}, y_{i-1}) + f_t(t_{i-1}, y_{i-1}) \cdot hc_2 + f_y(t_{i-1}, y_{i-1}) \cdot ha_{21} f(t_{i-1}, y_{i-1}) + O(h^2) \} =$$

$$y_{i-1} + hf(t_{i-1}, y_{i-1}) [b_1 + b_2] + h^2 [b_2 c_2 f_t(t_{i-1}, y_{i-1}) + b_2 a_{21} f_y(t_{i-1}, y_{i-1}) f(t_{i-1}, y_{i-1})] + O(h^3).$$

Como  $a_{21} = c_2$ , então

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) [b_1 + b_2] + h^2 b_2 c_2 [f_t(t_{i-1}, y_{i-1}) + f_y(t_{i-1}, y_{i-1}) f(t_{i-1}, y_{i-1})] + O(h^3).$$

Os parâmetros do Método 2-RK serão obtidos a partir da igualdade do polinômio de grau 2, na variável  $h$ , dado na expressão anterior e o polinômio dado no Método de Taylor de ordem 2:

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) + \frac{h^2}{2!} [f_t(t_{i-1}, y_{i-1}) + f_y(t_{i-1}, y_{i-1}) f(t_{i-1}, y_{i-1})].$$



Dessa forma,

$$b_1 + b_2 = 1 \quad \text{e} \quad b_2 c_2 = 1/2.$$

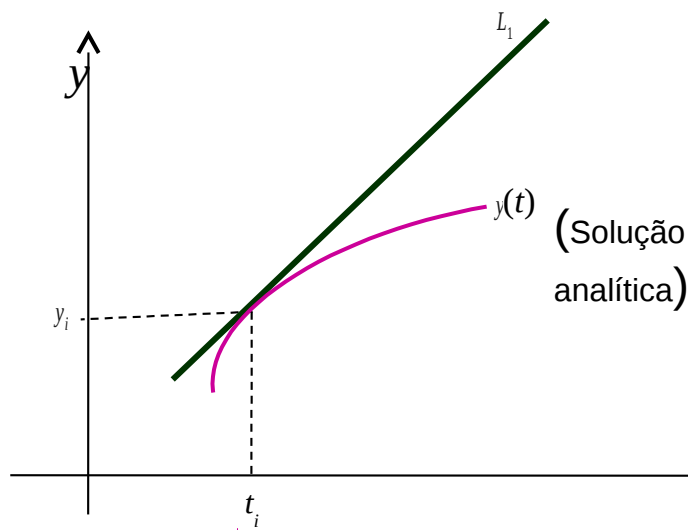
Portanto, existem infinitos métodos de Runge – Kutta de ordem dois e dois estágios. Supondo-se  $c_2 \neq 0$  então  $b_2 = 1/(2c_2)$  e  $b_1 = 1 - 1/(2c_2)$ .

Os dois métodos mais utilizados são o Método de Euler Melhorado e o Método de Euler Modificado. No Método de Euler Melhorado (ou Aperfeiçoado), tem-se que  $c_2 = 1$ ;  $b_2 = 1/(2c_2) = 1/2$  e  $b_1 = 1 - 1/(2c_2) = 1/2$ . No Método de Euler Modificado, tem-se que  $c_2 = 1/2$ ;  $b_2 = 1/(2c_2) = 1$  e  $b_1 = 1 - 1/(2c_2) = 0$ . Dessa forma, as expressões dos métodos são as seguintes:

Euler Aperfeiçoado:  $y_i = y_{i-1} + (h/2) \{ f(t_{i-1}, y_{i-1}) + f(t_i, y_{i-1} + h f(t_{i-1}, y_{i-1})) \}, 1 \leq i \leq n.$

Euler Modificado:  $y_i = y_{i-1} + h f(t_{i-1} + h/2, y_{i-1} + (h/2) f(t_{i-1}, y_{i-1})), 1 \leq i \leq n.$

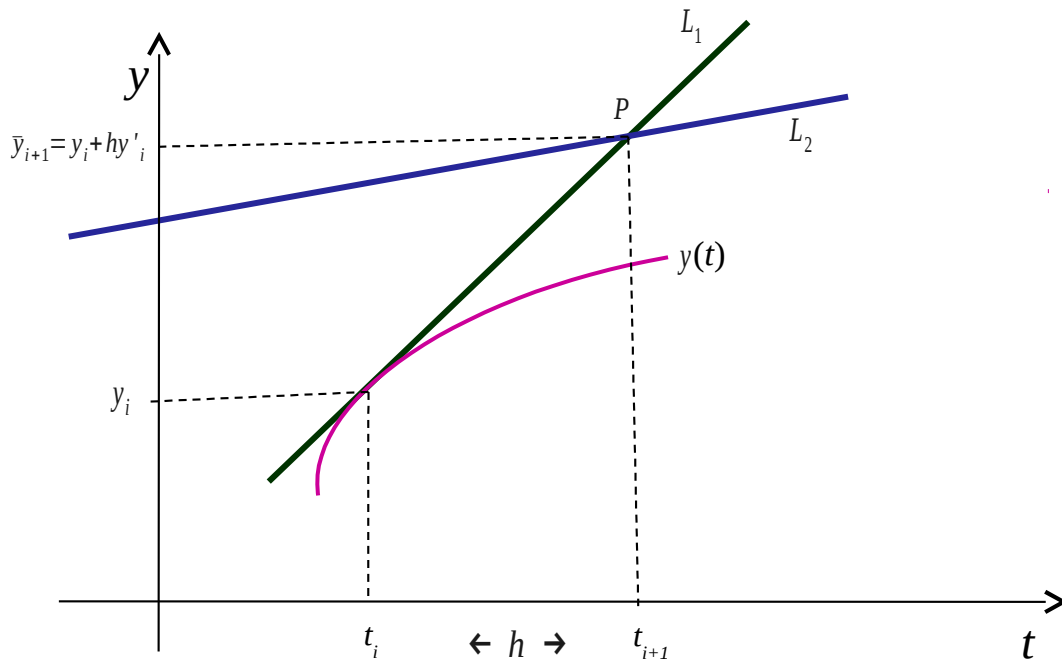
A seguir será exibida uma interpretação geométrica do Método de Euler Aperfeiçoado. Considere o ponto  $(t_i, y_i)$ ,  $y_i \approx y(t_i)$ . Suponha a situação ideal em que a curva desenhada com linha cheia seja a solução  $y(t)$  do PVI. Por  $(t_i, y_i)$  traçamos a reta  $L_1$  que tem coeficiente angular dado por  $f(t_i, y_i)$ ,  $L_1(t) = y_i + f(t_i, y_i)(t - t_i)$ , conforme a **Figura 2** abaixo.



**Figura 2 – Módulo 4 – 1ª parte da interpretação geométrica – Euler Melhorado**

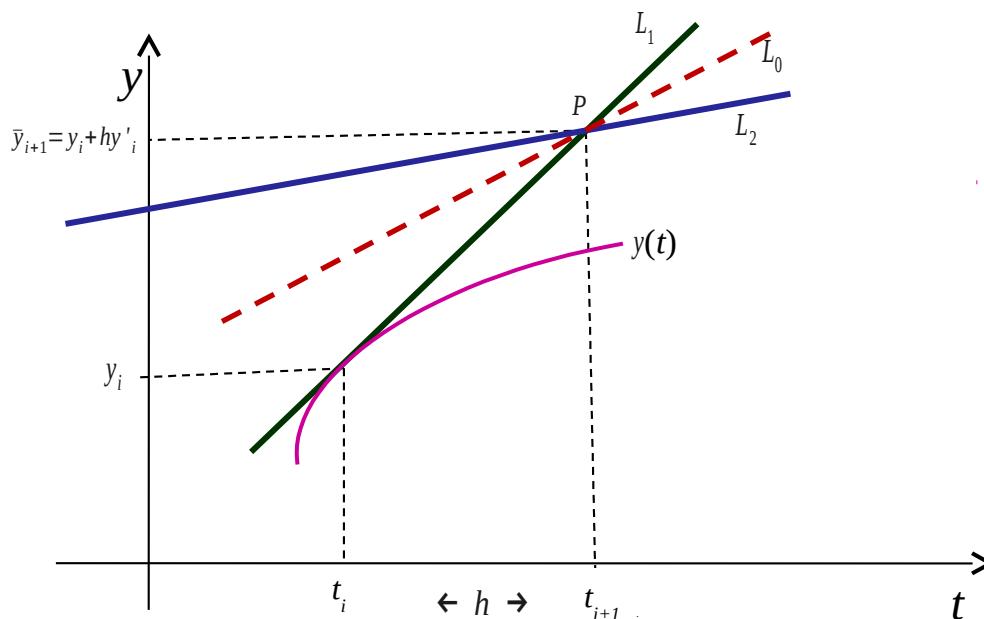
Note que  $L_1(t_i + h) = y_i + hf(t_i, y_i)$  é uma aproximação do Método de Euler de ordem 1, a qual será denotada por  $\bar{y}_{i+1}$ .

Agora considere o ponto  $P = (t_{i+1}, \bar{y}_{i+1})$ . A reta passando por  $P$  e que tem coeficiente angular dado por  $f(t_{i+1}, \bar{y}_{i+1})$  é dada por  $L_2(t) = \bar{y}_{i+1} + f(t_{i+1}, \bar{y}_{i+1})(t - t_{i+1})$  (Veja a **Figura 3**).



**Figura 3 – Módulo 4 – 2ª parte da interpretação geométrica – Euler Melhorado**

Na terceira etapa da construção geométrica do Método de Euler Melhorado, traçaremos a reta pontilhada  $L_0$  que passa por  $P$  e tem inclinação dada pela média aritmética das inclinações das retas  $L_1$  e  $L_2$ ,  $\frac{f(t_i, y_i) + f(t_{i+1}, \bar{y}_{i+1})}{2}$ , conforme a **Figura 4**.



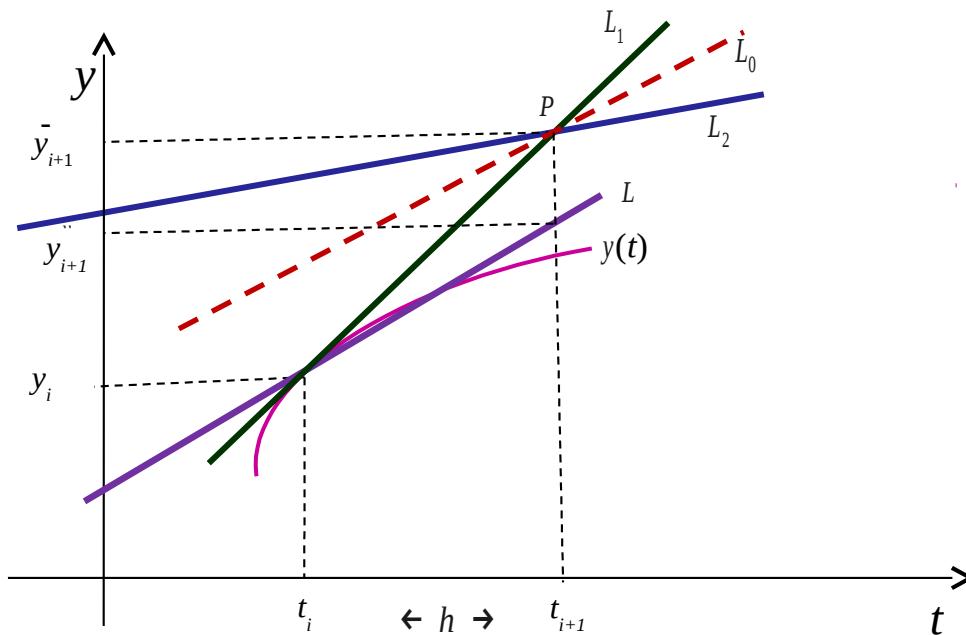
**Figura 4 – Módulo 4 – 3ª parte da interpretação geométrica – Euler Melhorado**

Finalmente, a última parte da interpretação geométrica está exibida na **Figura 5**. A reta  $L$  que passa por  $(t_i, y_i)$  e é paralela à reta  $L_0$  tem equação dada por:

$$L(t) = y_i + \frac{f(t_i, y_i) + f(t_{i+1}, \bar{y}_{i+1})}{2} (t - t_i).$$

Note que  $L(t_{i+1})$  é justamente igual à aproximação  $y_{i+1}$  do Método de Euler Melhorado:

$$y_{i+1} = y_i + (h/2) \{f(t_i, y_i) + f(t_{i+1}, y_i + h f(t_i, y_i))\}.$$



**Figura 5 – Módulo 4 – 4ª parte da interpretação geométrica – Euler Melhorado**

Como exercício, faça a interpretação geométrica do Método de Euler Modificado.

**Exemplo 4. (Método de Euler Modificado)** Considere o seguinte PVI:

$$y'(t) = f(t, y(t)) = t + y; \quad y(1) = y(t_0) = y_0 = 2.$$

Seja  $h = 0.05$  e  $t_1 = t_0 + h = 1.05$ . O Método de Euler Modificado de ordem 2 será utilizado para se obter a aproximação do PVI no ponto  $T = 1.2$ .

Como já foi mostrado nos exemplos anteriores,

$$t_0 = 1; \quad t_1 = 1.05; \quad t_2 = 1.1; \quad t_3 = 1.15; \quad t_4 = 1.2.$$

Assim, o Método de Euler Modificado,  $y_i = y_{i-1} + h f(t_{i-1} + h/2, y_{i-1} + (h/2) f(t_{i-1}, y_{i-1}))$ , dará origem aos seguintes valores:

$$y_1 = y_0 + h f(t_0 + h/2, y_0 + (h/2) f(t_0, y_0)) = 2 + 0.05 \{ t_0 + h/2 + y_0 + (h/2)[t_0 + y_0] \} = 2 + 0.05 \{ 1 + 0.025 + 2 + 0.025 [1 + 2] \} = 2.155;$$

$$y_2 = y_1 + h f(t_1 + h/2, y_1 + (h/2) f(t_1, y_1)) = 2.155 + 0.05 \{ t_1 + h/2 + y_1 + (h/2)[t_1 + y_1] \} = 2.155 + 0.05 \{ 1.05 + 0.025 + 2.155 + 0.025 [1.05 + 2.155] \} = 2.32050625;$$

$$y_3 = y_2 + hf(t_2 + h/2, y_2 + (h/2) f(t_2, y_2)) = 2.32050625 + 0.05 \{ t_2 + h/2 + y_2 + (h/2)[t_2 + y_2] \} \\ = 2.32050625 + 0.05 \{ 1.1 + 0.025 + 2.32050625 + 0.025 [1.1 + 2.32050625] \} = 2.497057195;$$

$$y_4 = y_3 + hf(t_3 + h/2, y_3 + (h/2) f(t_3, y_3)) = 2.497057195 + 0.05 \{ t_3 + h/2 + y_3 + (h/2)[t_3 + y_3] \} \\ = 2.497057195 + 0.05 \{ 1.15 + 0.025 + 2.497057195 + 0.025 [1.15 + 2.497057195] \} = 2.685218876.$$

Portanto,  $y(1.2) = y(t_4) \approx y_4 = 2.685218876$ . Note que todos os valores obtidos pelo Método de Euler Modificado coincidem com os valores obtidos, no **Exercício 1** da Seção 3.3, pelo Método de Taylor de ordem 2; porém o Método de Euler Modificado não utiliza derivadas parciais. Lembre-se de que a solução analítica,  $y(t) = 4e^{(t-1)} - t - 1$ , fornece  $y(1.2) = 2.685611033$ . ■

### Método de ordem 3 e 3 estágios e Método de ordem 4 e 4 estágios

Utilizando os mesmos procedimentos descritos para se deduzir os métodos 2-RK explícitos de ordem 2, os seguintes métodos podem ser obtidos:

- o método de ordem 3 e 3 estágios -

$$K_1 = hf(t_{i-1}, y_{i-1}); \quad K_2 = hf(t_{i-1} + h/2, y_{i-1} + K_1/2); \quad K_3 = hf(t_{i-1} + (3h)/4, y_{i-1} + (3K_2)/4); \\ y_i = y_{i-1} + (2/9)K_1 + (1/3)K_2 + (4/9)K_3, \quad 1 \leq i \leq n;$$

- o método de ordem 4 e 4 estágios -

$$K_1 = hf(t_{i-1}, y_{i-1}); \quad K_2 = hf(t_{i-1} + h/2, y_{i-1} + K_1/2); \\ K_3 = hf(t_{i-1} + h/2, y_{i-1} + K_2/2); \quad K_4 = hf(t_{i-1} + h, y_{i-1} + K_3); \\ y_i = y_{i-1} + (1/6)[K_1 + 2K_2 + 2K_3 + K_4], \quad 1 \leq i \leq n.$$

Os algoritmos desses métodos serão exibidos em uma próxima seção.



## Síntese

Os Métodos de Runge – Kutta de ordem  $p$  são de passo um, ou seja, são auto-iniciantes; não exigem o cálculo de derivadas parciais da função  $f(t,y)$  dada no PVI; necessitam apenas do cálculo de  $f(t,y)$  em determinados pontos, que são obtidos através de uma comparação com o Método de Taylor de ordem  $p$ ; são análogos ao Método de Taylor de ordem  $p$  no seguinte sentido: expandindo-se a função  $f$  em série de Taylor em torno do ponto  $(t_{i-1}, y_{i-1})$  e agrupando-se os termos em relação às potências de  $h$ , de 0 até  $p$ , a expressão do método de Runge – Kutta (polinômio de grau  $p$  na variável  $h$ ) coincide com a expressão do Método de Taylor.

## 4. Métodos Baseados em Integração Numérica

Nesta seção, serão estudados os métodos conhecidos como Métodos de Adams – Bashforth e Métodos de Adams – Moulton. Nesses métodos, o Problema de Valor Inicial:  $y'(t) = f(t, y(t))$  e  $y(t_0) = y_0$  é transformado na seguinte formulação integral:

$$\int_{t_{i-1}}^{t_i} y'(t) dt = \int_{t_{i-1}}^{t_i} f(t, y(t)) dt \quad \rightarrow \quad y(t_i) = y(t_{i-1}) + \int_{t_{i-1}}^{t_i} f(t, y(t)) dt \quad \text{e} \quad y(t_0) = y_0.$$

Analogamente às etapas propostas na Introdução deste módulo (após a primeira observação da Seção 1), observe que a partição do intervalo  $I = [t_0, T]$  considerou pontos igualmente espaçados:  $t_i = t_0 + ih$ ,  $0 \leq i \leq n$ , onde  $h = (T - t_0)/n$ .

Após a formulação integral do PVI, a construção dos métodos numéricos de Adams deve se basear nas seguintes etapas:

(i) Para  $t \in [t_0, T]$ , defina a função  $G(t) = f(t, y(t))$ .

(ii) Construa os polinômios interpoladores de Lagrange de grau  $m$  para a função  $G$  considerando os seguintes  $(m + 1)$  pontos de interpolação: (I)  $t_i, t_{i-1}, \dots, t_{i-m}$  ou (II)  $t_{i-1}, t_{i-2}, \dots, t_{i-(m+1)}$ .

(iii) Considere as seguintes aproximações: (I)  $\int_{t_{i-1}}^{t_i} G(t, y(t)) dt \approx \int_{t_{i-1}}^{t_i} p^{(I)}(t) dt$  ; (II)

$$\int_{t_{i-1}}^{t_i} G(t, y(t)) dt \approx \int_{t_{i-1}}^{t_i} p^{(II)}(t) dt; \quad \text{desprezando-se os erros de integração.}$$

(iv) Substitua  $y(t_i)$  e  $y(t_{i-1})$  pelos valores aproximados  $y_i$  e  $y_{i-1}$ . Nas regras de integração, substitua  $G(t_{i-j}) = f(t_{i-j}, y(t_{i-j}))$  pelo valor aproximado  $f(t_{i-j}, y_{i-j})$ , para cada  $j$ ,  $0 \leq j \leq m + 1$ .

**Observação:** O erro local dos métodos de Adams coincide com o erro de integração numérica:

$$E^{(I)} = \int_{t_{i-1}}^{t_i} G(t, y(t)) dt - \int_{t_{i-1}}^{t_i} p^{(I)}(t) dt \quad \text{ou}$$

$$E^{(II)} = \int_{t_{i-1}}^{t_i} G(t, y(t)) dt - \int_{t_{i-1}}^{t_i} p^{(II)}(t) dt.$$

**Observação:** Os métodos que são construídos com a utilização do polinômio  $p^{(I)}$  são denominados métodos implícitos (Adams – Moulton) e os que utilizam o polinômio  $p^{(II)}$  são denominados métodos explícitos (Adams – Bashforth).

#### 4.1. Métodos de Adams – Bashforth – Moulton de ordem 2

Considere o polinômio de Lagrange de grau 1 sobre os pontos: (I)  $x_{0,I} = t_{i-1}$  e  $x_{1,I} = t_i$  e sobre os pontos: (II)  $x_{0,II} = t_{i-2}$  e  $x_{1,II} = t_{i-1}$ . Esses polinômios serão representados por  $p^{(\vartheta)}(t) = G(x_{0,\vartheta})L_{0,\vartheta}(t) + G(x_{1,\vartheta})L_{1,\vartheta}(t)$ , onde  $L_{0,\vartheta}(t) = \frac{(t-x_{1,\vartheta})}{(x_{0,\vartheta}-x_{1,\vartheta})}$ ;  $L_{1,\vartheta}(t) = \frac{(t-x_{0,\vartheta})}{(x_{1,\vartheta}-x_{0,\vartheta})}$ ;  $\vartheta = I$  ou  $\vartheta = II$  e  $G(t) = f(t, y(t))$ .

O valor de  $R_I = \int_{t_{i-1}}^{t_i} p^{(I)}(t) dt$  deve ser calculado para se deduzir o método implícito.

Porém esse cálculo é exatamente o mesmo que foi realizado na obtenção da Regra do Trapézio (Seção 5.1 do Módulo 3). Portanto,

$$R_I = (h/2)\{G(t_{i-1}) + G(t_i)\} \rightarrow R_I = (h/2)\{f(t_{i-1}, y(t_{i-1})) + f(t_i, y(t_i))\}.$$

No método explícito, é preciso calcular  $R_{II} = \int_{t_{i-1}}^{t_i} p^{(II)}(t) dt$ . Fazendo a troca de variáveis  $t = t_{i-1} + uh$ , obtém-se que

$$R_{II} = \int_0^1 p^{(II)}(t_{i-1} + uh) h du = G(t_{i-2}) \int_0^1 -u h du + G(t_{i-1}) \int_0^1 (u+1) h du \rightarrow$$

$$\rightarrow R_{II} = (h/2)\{-G(t_{i-2}) + 3G(t_{i-1})\} \rightarrow R_{II} = (h/2)\{-f(t_{i-2}, y(t_{i-2})) + 3f(t_{i-1}, y(t_{i-1}))\}.$$

Dessa forma, os métodos de Adams implícito e explícito de ordem 2 são dados, respectivamente, por:

$$y_i = y_{i-1} + (h/2)\{f(t_{i-1}, y_{i-1}) + f(t_i, y_i)\}; \quad 1 \leq i \leq n, \quad (\text{Adams – Moulton})$$

e

$$y_i = y_{i-1} + (h/2)\{-f(t_{i-2}, y_{i-2}) + 3f(t_{i-1}, y_{i-1})\}; \quad 2 \leq i \leq n, \quad (\text{Adams – Bashforth}).$$

**Observação:** O primeiro método anterior é implícito porque a sua expressão é da forma  $y_i = F_I(y_i)$ , onde  $F_I(y_i) = y_{i-1} + (h/2)\{f(t_{i-1}, y_{i-1}) + f(t_i, y_i)\}$ . Portanto,  $y_i$  está em função dele mesmo, ou seja,  $y_i$  está definido de modo implícito. O valor de  $y_i$ , em geral, é calculado através de um método apropriado para obter zero de função, tal como o Método do Ponto Fixo (ou Método Iterativo Linear).

**Observação:** O segundo método anterior é explícito e de passo 2 porque a sua expressão é da forma  $y_i = F_{II}(y_{i-2}, y_{i-1}) = y_{i-1} + (h/2)\{-f(t_{i-2}, y_{i-2}) + 3f(t_{i-1}, y_{i-1})\}$ . Portanto, dados os dois valores iniciais,  $y_0$  e  $y_1$ , os demais valores são calculados, explicitamente, sem a necessidade de métodos iterativos.

## Fórmulas dos Erros de Integração – métodos de ordem 2

Os métodos implícito e explícito apresentados anteriormente são de ordem 2 porque os erros associados às regras de integração empregadas em suas deduções são do tipo  $O(h^3)$ , como será mostrado a seguir.

Lembre-se de que o erro de integração é a integral do erro de interpolação (veja a Seção 5 do módulo 3). Dessa forma,  $E^{(I)} = \int_{t_{i-1}}^{t_i} N(t) \frac{G^{(2)}(t)}{2!} dt$ , onde  $N(t) = (t - t_{i-1})(t - t_i) < 0$ ,  $\forall t \in (t_{i-1}, t_i)$ .

Supondo que a derivada segunda  $G^{(2)}$  seja contínua então  $E^{(I)}$  vai ter a mesma expressão do erro da Regra do Trapézio (veja a Seção 5.1 do Módulo 3). Assim,  $E^{(I)} = -\frac{h^3 G^{(2)}(c)}{12}$ , onde  $c \in (t_{i-1}, t_i)$ .

Além disso,  $E^{(II)} = \int_{t_{i-1}}^{t_i} N(t) \frac{G^{(2)}(t)}{2!} dt$ , onde  $N(t) = (t - t_{i-2})(t - t_{i-1}) > 0$ ,  $\forall t \in (t_{i-1}, t_i)$ .

Considerando-se a mesma hipótese anterior sobre  $G^{(2)}$  e usando o Teorema do Valor Intermediário para integrais (veja as observações apresentadas na Seção 5.1 do Módulo 3), tem-se que:  $E^{(II)} = \frac{G^{(2)}(c)}{2} \int_{t_{i-1}}^{t_i} N(t) dt$ . Utilizando a troca de variáveis  $t = t_{i-1} + uh$ ,

obtem-se que

$$E^{(II)} = \frac{G^{(2)}(c)}{2} \int_0^1 N(t_{i-1} + uh) h du = \frac{h^3 G^{(2)}(c)}{2} \int_0^1 (u+1)u du = \frac{5 h^3 G^{(2)}(c)}{12},$$

$c \in (t_{i-1}, t_i)$ .

**Observação:** Como  $G(c) = f(c, y(c)) = y'(c)$ , então  $G^{(2)}(c) = f^{(2)}(c, y(c)) = y^{(3)}(c)$ .



## Método Previsor – Corretor de ordem 2

Utilizando-se os métodos de Adams implícito e explícito, pode-se construir um esquema numérico denominado método previsor – corretor, que é aplicado na resolução de um PVI. As etapas desse procedimento são exibidas a seguir.

(i) Obtenha os valores iniciais do método previsor (método explícito de passo múltiplo – Método de Adams-Bashforth), utilizando um método auto-iniciante que tenha a mesma ordem que os métodos implícito e explícito.

(ii) Calcule o valor de  $y_i$  pelo método previsor e denote este valor por  $y_i^{(0)}$ .

(iii) Use o valor  $y_i^{(0)}$  no lado direito da expressão do método corretor (método implícito – Método de Adams-Moulton). Mais especificamente, com a notação utilizada anteriormente (veja a primeira observação da Seção 4.1 deste módulo), substitua  $F_I(y_i)$  por  $F_I(y_i^{(0)})$ .

(iv) Execute quantas correções forem necessárias para obter a aproximação desejada para  $y_i$ . Note que, para cada  $i$ , está sendo executado o Método do Ponto Fixo (análogo ao estudado no tópico Zero de Função):  $y_i^{(k)} = F_I(y_i^{(k-1)})$ ,  $k \geq 1$ .

O esquema previsor – corretor será aplicado no próximo exercício.

**Exercício 2. (Método Previsor – Corretor)** Considere o seguinte PVI:

$$y'(t) = f(t, y(t)) = t + y; \quad y(1) = y(t_0) = y_0 = 2.$$

Seja  $h = 0.05$ . Obtenha uma aproximação da solução do PVI, em  $T = 1.2$ . Utilize o par previsor – corretor de Adams-Bashforth-Moulton de ordem 2.

**Solução do Exercício 2:**

Sabemos que  $I = [t_0, T] = [1, 1.2]$ ;  $h = (T - t_0)/n \rightarrow n = (1.2 - 1)/0.05 = 20/5 = 4$ . Além disso,  $t_i = t_0 + ih$ ,  $0 \leq i \leq 4$ , ou seja,  $t_0 = 1$ ;  $t_1 = 1.05$ ;  $t_2 = 1.1$ ;  $t_3 = 1.15$  e  $t_4 = 1.2$ .

Seguindo as etapas do esquema previsor – corretor, tem-se que

(i) O método previsor tem expressão dada por  $y_i = y_{i-1} + (h/2)\{-f(t_{i-2}, y_{i-2}) + 3f(t_{i-1}, y_{i-1})\}$ , logo, precisa de dois valores iniciais  $y_0$  e  $y_1$ . De acordo com o PVI,  $y_0 = 2$ . O valor de  $y_1$  deve ser calculado por um método de ordem 2, auto-iniciante (ou de passo 1). Utilizando o Método de Runge-Kutta de ordem 2, considere  $y_1 = 2.155$ .

(ii) Dado  $y_1$ , o método previsor nos fornece

$$y_2^{(0)} = y_1 + (h/2)\{-f(t_0, y_0) + 3f(t_1, y_1)\} = 2.155 + 0.025\{-(t_0 + y_0) + 3(t_1 + y_1)\} \rightarrow$$

$$\rightarrow y_2^{(0)} = 2.155 + 0.025\{-(1 + 2) + 3(1.05 + 2.155)\} = 2.320375.$$

(iii e iv) A partir de  $y_2^{(0)}$  serão realizadas iterações (neste exercício, serão realizadas apenas duas iterações) com o método corretor (método implícito):  $y_2^{(k)} = F_I(y_2^{(k-1)})$ ,  $k = 1$  e  $k = 2$ . Como o método iterativo é dado por  $y_2^{(k)} = y_1 + (h/2)f(t_1, y_1) + (h/2)f(t_2, y_2^{(k-1)})$ , então, durante as iterações, o valor de  $y_1 + (h/2)f(t_1, y_1)$  não se altera. Dessa forma, o método corretor fornecerá as seguintes aproximações:

$$(\text{Corretor: } y_i = y_{i-1} + (h/2)\{f(t_{i-1}, y_{i-1}) + f(t_i, y_i)\})$$

valor constante:

$$y_1 + (h/2)f(t_1, y_1) = 2.155 + 0.025(1.05 + 2.155) = 2.235125.$$

iterações:

$$y_2^{(1)} = 2.235125 + (h/2)f(t_2, y_2^{(0)}) = 2.235125 + 0.025(t_2 + y_2^{(0)}) \rightarrow$$

$$\rightarrow y_2^{(1)} = 2.235125 + 0.025(1.1 + 2.320375) = 2.320634375.$$

$$y_2^{(2)} = 2.235125 + (h/2)f(t_2, y_2^{(1)}) = 2.235125 + 0.025(t_2 + y_2^{(1)}) \rightarrow$$

$$\rightarrow y_2^{(2)} = 2.235125 + 0.025(1.1 + 2.320634375) = 2.320640859.$$

Repetindo-se as etapas anteriores, obtêm-se:

$$(\text{Previsor: } y_i = y_{i-1} + (h/2)\{-f(t_{i-2}, y_{i-2}) + 3f(t_{i-1}, y_{i-1})\})$$

$$y_3^{(0)} = y_2 + (h/2)\{-f(t_1, y_1) + 3f(t_2, y_2)\} = 2.320640859 + 0.025\{-(t_1 + y_1) + 3(t_2 + y_2)\} \rightarrow$$

$$y_3^{(0)} = 2.320640859 + 0.025\{-(1.05 + 2.155) + 3(1.1 + 2.320640859)\} = 2.497063923.$$

$$(\text{Corretor: } y_i = y_{i-1} + (h/2)\{f(t_{i-1}, y_{i-1}) + f(t_i, y_i)\})$$

valor constante:

$$y_2 + (h/2)f(t_2, y_2) = 2.320640859 + 0.025(1.1 + 2.320640859) = 2.40615688.$$

iterações:

$$y_3^{(1)} = 2.40615688 + (h/2)f(t_3, y_3^{(0)}) = 2.40615688 + 0.025(t_3 + y_3^{(0)}) \rightarrow$$

$$\rightarrow y_3^{(1)} = 2.40615688 + 0.025(1.15 + 2.497063923) = 2.497333479.$$

$$y_3^{(2)} = 2.40615688 + (h/2)f(t_3, y_3^{(1)}) = 2.40615688 + 0.025(t_3 + y_3^{(1)}) \rightarrow$$

$$\rightarrow y_3^{(2)} = 2.40615688 + 0.025(1.15 + 2.497333479) = 2.497340217.$$

( Previsor:  $y_i = y_{i-1} + (h/2)\{-f(t_{i-2}, y_{i-2}) + 3f(t_{i-1}, y_{i-1})\}$  )

$$y_4^{(0)} = y_3 + (h/2)\{-f(t_2, y_2) + 3f(t_3, y_3)\} = 2.497340217 + 0.025\{-(t_2 + y_2) + 3(t_3 + y_3)\} \rightarrow$$

$$y_4^{(0)} = 2.497340217 + 0.025\{-(1.1 + 2.320640859) + 3(1.15 + 2.497340217)\} = 2.685374712.$$

( Corretor:  $y_i = y_{i-1} + (h/2)\{f(t_{i-1}, y_{i-1}) + f(t_i, y_i)\}$  )

valor constante:

$$y_3 + (h/2)f(t_3, y_3) = 2.497340217 + 0.025(1.15 + 2.497340217) = 2.588523722.$$

iterações:

$$y_4^{(1)} = 2.588523722 + (h/2)f(t_4, y_4^{(0)}) = 2.587116135 + 0.025(t_4 + y_4^{(0)}) \rightarrow$$

$$\rightarrow y_4^{(1)} = 2.588523722 + 0.025(1.2 + 2.685374712) = 2.68565809.$$

$$y_4^{(2)} = 2.588523722 + (h/2)f(t_4, y_4^{(1)}) = 2.587116135 + 0.025(t_4 + y_4^{(1)}) \rightarrow$$

$$\rightarrow y_4^{(2)} = 2.588523722 + 0.025(1.2 + 2.68565809) = 2.685665174.$$

Lembre-se de que a solução analítica,  $y(t) = 4e^{(t-1)} - t - 1$ , fornece  $y(1.2) = 2.685611033$ . Portanto, a aproximação anterior foi a melhor obtida com um método de ordem 2. Perceba que a quantidade de cálculos efetuados com o método previsor-corretor foi bem maior do que a quantidade efetuada por todos os outros métodos de ordem 2. O esforço computacional é o preço que se paga para se obter métodos mais precisos. ■

## 4.2. Métodos de Adams – Bashforth – Moulton de ordem 3

Considere o polinômio de Lagrange de grau 2 sobre os pontos: (I)  $x_{0,I} = t_{i-2}$ ,  $x_{1,I} = t_{i-1}$  e  $x_{2,I} = t_i$  e sobre os pontos: (II)  $x_{0,II} = t_{i-3}$ ,  $x_{1,II} = t_{i-2}$  e  $x_{2,II} = t_{i-1}$ . Usando a mesma notação da Seção 4.1, esses polinômios serão representados por  $p^{(\vartheta)}(t) = G(x_{0,\vartheta})L_{0,\vartheta}(t) + G(x_{1,\vartheta})L_{1,\vartheta}(t) + G(x_{2,\vartheta})L_{2,\vartheta}(t)$ , onde  $L_{0,\vartheta}(t) = \frac{(t-x_{1,\vartheta})(t-x_{2,\vartheta})}{(x_{0,\vartheta}-x_{1,\vartheta})(x_{0,\vartheta}-x_{2,\vartheta})}$ ;  $L_{1,\vartheta}(t) = \frac{(t-x_{0,\vartheta})(t-x_{2,\vartheta})}{(x_{1,\vartheta}-x_{0,\vartheta})(x_{1,\vartheta}-x_{2,\vartheta})}$ ;  $L_{2,\vartheta}(t) = \frac{(t-x_{0,\vartheta})(t-x_{1,\vartheta})}{(x_{2,\vartheta}-x_{0,\vartheta})(x_{2,\vartheta}-x_{1,\vartheta})}$ ;  $\vartheta = I$  ou  $\vartheta = II$  e  $G(t) = f(t, y(t))$ .

Os métodos implícito e explícito serão deduzidos de forma análoga às deduções que foram feitas, anteriormente, para os métodos de ordem 2, na Seção 4.1. Dessa forma, será preciso calcular as integrais dos polinômios de grau 2,  $p^{(\vartheta)}(t)$ ,  $\vartheta = I$  e  $\vartheta = II$ , no intervalo  $[t_{i-1}, t_{i+1}]$ .

Utilizando a troca usual de variáveis,  $t = t_{i-1} + uh$ , obtém-se que:

$$R_I = \int_{t_{i-1}}^{t_i} p^{(I)}(t) dt = \int_0^1 p^{(I)}(t_{i-1} + uh) h du = G(t_{i-2}) \int_0^1 \frac{uh(u-1)h}{-h(-2h)} h du + \\ G(t_{i-1}) \int_0^1 \frac{(u+1)h(u-1)h}{h(-h)} h du + G(t_i) \int_0^1 \frac{(u+1)h uh}{2h(h)} h du \rightarrow$$

$$\rightarrow R_I = (h/12)[-1G(t_{i-2}) + 8G(t_{i-1}) + 5G(t_i)].$$

$$R_{II} = \int_{t_{i-1}}^{t_i} p^{(II)}(t) dt = \int_0^1 p^{(II)}(t_{i-1} + uh) h du = G(t_{i-3}) \int_0^1 \frac{(u+1)h uh}{-h(-2h)} h du + \\ G(t_{i-2}) \int_0^1 \frac{(u+2)h uh}{h(-h)} h du + G(t_{i-1}) \int_0^1 \frac{(u+2)h(u+1)h}{2h(h)} h du \rightarrow$$

$$\rightarrow R_{II} = (h/12)[5G(t_{i-3}) - 16G(t_{i-2}) + 23G(t_{i-1})].$$

Dessa forma, os métodos de Adams implícito e explícito de ordem 3 são dados, respectivamente, por:

$$y_i = y_{i-1} + (h/12)\{-1f(t_{i-2}, y_{i-2}) + 8f(t_{i-1}, y_{i-1}) + 5f(t_i, y_i)\}; \quad 2 \leq i \leq n, \quad (\text{Adams – Moulton})$$

e

$$y_i = y_{i-1} + (h/12)\{5f(t_{i-3}, y_{i-3}) - 16f(t_{i-2}, y_{i-2}) + 23f(t_{i-1}, y_{i-1})\}; \quad 3 \leq i \leq n, \quad (\text{Adams – Bashforth}).$$

**Observação:** Os dois métodos anteriores são de passo múltiplo. O método implícito (Adams – Moulton) é de passo 2 porque necessita de dois valores iniciais,  $y_0$  e  $y_1$ , enquanto que o método explícito (Adams – Bashforth) é de passo 3 porque precisa de 3 valores iniciais,  $y_0$ ,  $y_1$  e  $y_2$ .

### Fórmulas dos Erros de Integração – métodos de ordem 3

Os métodos implícito e explícito apresentados anteriormente são de ordem 3 porque os erros associados às regras de integração empregadas em suas deduções são do tipo  $O(h^4)$ , como será mostrado a seguir.

Lembre-se de que o erro de integração é a integral do erro de interpolação (veja a Seção 5 do módulo 3). Dessa forma, utilizando a notação apresentada na Seção 4 anterior,  $E^{(I)}$

$$= \int_{t_{i-1}}^{t_i} N_I(t) \frac{G^{(3)}(t)}{3!} dt, \quad \text{onde } N_I(t) = (t - t_{i-2})(t - t_{i-1})(t - t_i) < 0, \quad \forall t \in (t_{i-1}, t_i);$$

$$E^{(II)} = \int_{t_{i-1}}^{t_i} N_{II}(t) \frac{G^{(3)}(t)}{3!} dt, \text{ onde } N_{II}(t) = (t - t_{i-3})(t - t_{i-2})(t - t_{i-1}) > 0, \forall t \in (t_{i-1}, t_i).$$

Supondo que a derivada terceira  $G^{(3)}$  seja contínua; usando o Teorema do Valor Intermediário para integrais (veja as observações apresentadas na Seção 5.1 do Módulo 3) e a troca de variáveis  $t = t_{i-1} + uh$ , tem-se que:

$$E^{(I)} = \frac{G^{(3)}(c)}{6} \int_{t_{i-1}}^{t_i} N_I(t) dt = \frac{-h^4 G^{(3)}(c)}{24}, \quad c \in (t_{i-1}, t_i).$$

$$E^{(II)} = \frac{G^{(3)}(C)}{6} \int_{t_{i-1}}^{t_i} N_{II}(t) dt = \frac{9 h^4 G^{(3)}(C)}{24}, \quad C \in (t_{i-1}, t_i).$$

**Observação:** Como  $G(c) = f(c, y(c)) = y'(c)$  então  $G^{(3)}(c) = f^{(3)}(c, y(c)) = y^{(4)}(c)$ .

Já deu para perceber que a resolução numérica de um PVI através dos métodos de Adams de passo múltiplo, baseados em integração numérica, vai demandar muito esforço computacional. Por essa razão, não é conveniente ficar fazendo os vários cálculos exigidos nas iterações do corretor e nas sucessivas avaliações da função  $f$ , exigidas tanto no método predictor quanto no corretor, apenas com o auxílio de uma calculadora. O ideal é elaborar um programa computacional que faça todos estes cálculos. Esse programa computacional será exibido na Seção 6.

## 5. Atividades do Módulo 4

### Atividade 1 – Adquirindo e testando conhecimentos através da resolução da lista de exercícios

#### Enunciado da atividade

Em alguns exercícios desta lista você precisará lidar com o desenvolvimento de Taylor de funções de uma ou duas variáveis; portanto, é fundamental que você recorde a teoria vista sobre este assunto na disciplina Cálculo III.

#### Quarta Lista de Exercícios de Cálculo Numérico

1) Integrando-se a EDO de um PVI no intervalo  $[x_{i-1}, x_i]$ , obtém-se que  $y(x_i) = y(x_{i-1}) + \int_{x_{i-1}}^{x_i} f(x, y(x)) dx$ . Seja  $G(x) = f(x, y(x))$ . A integral de  $G(x)$  pode ser aproximada pelos polinômios interpoladores sobre os pontos  $x_0 = x_{i-2}$ ,  $x_1 = x_{i-1}$  e  $x_2 = x_i$  ou sobre os

pontos:  $x_0 = x_{i-3}$ ,  $x_1 = x_{i-2}$  e  $x_2 = x_{i-1}$ . A notação  $f_i$  denotará  $f(x_i, y_i)$ . O procedimento anterior dará origem aos seguintes métodos:

$$\text{I) } y_k = y_{k-1} + (h/12) [-f_{k-2} + 8f_{k-1} + 5f_k], \text{ com erro local } O_I(h^4) = (-1/24).h^4.y^{(4)}(\eta);$$

$$\text{II) } y_k = y_{k-1} + (h/12) [5f_{k-3} - 16f_{k-2} + 23f_{k-1}], \text{ com erro local } O_{II}(h^4) = (3/8).h^4.y^{(4)}(\mu);$$

$\eta$  e  $\mu$  são constantes. Considere  $y_1 = 2,07626$ ;  $y_2 = 2,15508$  e o PVI:  $y' = x + y$ ;  $y(1) = 2$ , com  $x \in I = [1, 1.2]$ . Seja  $h = 0.025$  o comprimento dos subintervalos de  $I$ .

i) Calcule o valor de  $y_3$  pelo método implícito; para isso, use o valor inicial fornecido pelo método explícito.

ii) Obtenha uma das fórmulas do erro:  $\int_{x_{i-1}}^{x_i} \frac{G'''(\xi(x))}{3!} N(x) dx$ , onde  $N(x) = (x-x_0)(x-x_1)(x-x_2)$ .

Sugestão: Mostre que  $N(x)$  não muda de sinal no intervalo dado e use o Teorema do Valor Intermediário, como na dedução do erro na regra de integração do Trapézio.

iii) Ambos os métodos são de passo múltiplo. Diga qual é a multiplicidade do passo de cada um deles. Justifique.

iv) Comente a seguinte afirmação: “Se o PVI a ser resolvido tiver equação  $y' = x + y$ , então os valores iniciais dos métodos I e II poderiam ser calculados pelo método de Taylor de ordem 3, sem comprometer o esforço computacional. Porém, se o PVI tiver equação  $y' = (2/x)y + x^2e^x$ , então o método de Runge-Kutta de ordem 3 seria mais adequado para o cálculo dos valores iniciais”.

2) Considere o PVI  $y' = yx^2 - y$ ;  $y(0) = 1$ , com  $x \in I = [0, 2]$ . Seja  $h = 0.25$  o comprimento dos subintervalos de  $I$ . O método de Euler tem equação  $y_i = y_{i-1} + h f_{i-1}$ , onde  $f_{i-1} = f(x_{i-1}, y_{i-1})$ ; o método do ponto médio tem equação dada por  $y_i = y_{i-2} + 2h f_{i-1}$ , e possui erro local da forma  $O(h^3)$ . O método de Runge-Kutta explícito de dois estágios e ordem dois, denominado Euler Modificado, tem equação da forma:  $y_i = y_{i-1} + hf(x_{i-1} + h/2, y_{i-1} + [h/2]f_{i-1})$ . O método de Taylor de ordem dois tem equação  $y_i = y_{i-1} + hf_{i-1} + [h^2/2] \{ f_x(x_{i-1}, y_{i-1}) + f_{i-1} f_y(x_{i-1}, y_{i-1}) \}$ . Os métodos baseados na regra do trapézio possuem erro de truncamento do tipo  $O(h^3)$  e têm equações dadas por:  $y_i = y_{i-1} + (h/2)[f_{i-1} + f_i] \equiv F(y_i)$  e  $y_i = y_{i-1} + (h/2)\{-f_{i-2} + 3f_{i-1}\}$ .

i) Considerando-se o valor de  $h$  dado acima, qual deve ser o valor de  $n$  para se obter  $y_n \approx y(2)$ ?

ii) Escreva a equação do método de Taylor de ordem 2 para o PVI acima.

- iii) Quais dos métodos anteriores são auto-iniciantes? Justifique.
- iv) Dos métodos exibidos acima, quais são de passo 2? Seria adequado usar o método de Euler para se obter o valor inicial  $y_1$  dos métodos de passo 2?
- v) Em relação à precisão, existe alguma diferença entre o método de Taylor e o de Runge-Kutta exibidos acima? Justifique a sua resposta. Qual a principal diferença entre eles?
- vi) Use um dos métodos anteriores para obter o valor de  $y_1$  que seja adequado para ser usado nos métodos de passo 2 dados anteriormente.
- vii) Obtenha  $(y_2)^{(0)}$ ,  $y_2$  e  $(y_3)^{(0)}$  usando o par previsor-corretor dado pelos métodos explícito e implícito do trapézio. Faça apenas uma iteração com o corretor (método implícito).

3) O professor de Cálculo Numérico, percebendo que os seus alunos andavam muito displicentes nas aulas de laboratório, resolveu pregar uma peça em um dos alunos, o *mais cara-de-pau*. O “*cara-de-pau com força*” (que será denotado por *m\_cp.cf*) nem ao menos trocou o nome dos arquivos que estavam no disquete de sua colega de classe (Sim, disquete! Naquele tempo ainda não existia *pen-drive*), a qual havia participado minutos antes da aula destinada aos 30 primeiros alunos da turma. O logro acabou virando um teste que foi estendido a todos os alunos daquela turma.

O “*m\_cp.cf*” possuía dois arquivos no disquete, um deles foi nomeado de *mari.m* e continha uma rotina do método previsor-corretor de Adams-Moulton. O professor solicitou ao aluno que alterasse a função do PVI estudado. A partir daí seguiram-se alguns questionamentos:

(i) Um código calcula a aproximação de  $y' = f(x,y)$ , no ponto  $x = 2.3$ , partindo do valor inicial  $y(2) = 0.5$  e usando  $h = 0.015$ . Alterando-se o valor inicial para  $y(1) = 2$ , quantos passos são necessários para se obter a aproximação do problema modificado, no mesmo ponto  $x = 2.3$ , considerando subintervalos de comprimento  $h \leq 0.015$  ?

(ii) Obtido  $y_p$ , calculado pelo método previsor:  $y_i = y_{i-1} + \frac{h}{24} (55 f_{i-1} - 59 f_{i-2} + 37 f_{i-3} - 9 f_{i-4})$ , o método corretor,  $y_i = y_{i-1} + \frac{h}{24} (9 f_i + 19 f_{i-1} - 5 f_{i-2} + 1 f_{i-3})$ , em que  $f_i = f(x_i, y_i)$ , é empregado de modo eficiente, com esforço mínimo, da seguinte forma:

$$y_p = y_i; \quad y = y_{i-1}; \quad xx = x_{i-1} + h; \quad func(5) = f(xx, y_p);$$

$$y_c = y + (h/24)(19 func(4) - 5 func(3) + func(2));$$

*for k=1:1:10*

$$y = y_c + \frac{9h}{24} \text{func}(5);$$

$$\text{func}(5) = f(x, y);$$

*end*

*%y será a aproximação para  $y(x_i)$ .*

Explique o procedimento anterior. Indique quantas iterações foram feitas. Quais os valores que estão sendo armazenados nas componentes *func(4)*, *func(3)* e *func(2)*? Justifique por que o esforço será mínimo. Qual é o passo de cada um dos métodos de passo múltiplo? Indique os valores iniciais.

4) O método de Runge-Kutta explícito de dois estágios e ordem dois possui equação dada por:

$$y_k = y_{k-1} + h\{b_1 f_{k-1} + b_2 f(x_{k-1} + hc_2, y_{k-1} + hc_2 f_{k-1})\}, \text{ onde}$$

$b_1 + b_2 = 1$  e  $b_2 c_2 = 1/2$ . Considere  $c_2 = 1$  e o PVI:  $y' - x + y - 2 = 0$ ;  $y(0) = 2$ . Use  $h = 0.1$  e o método acima para obter  $y_1 \cong y(0.1)$ .

5) O modelo de Gompertz é adequado para traduzir crescimento celular (plantas, bactérias, tumores etc). A EDO  $y' = b \cdot y(t) \cdot e^{-at}$  que representa este modelo é facilmente resolvida pelo método de separação de variáveis. Considere  $b = 1.5$ ,  $a = 0.5$  e  $y(0) = 1$ .

i) Para se obter uma estimativa para  $y(1.5)$ , usando  $h = 0.1$ , quantos passos devem ser executados?

Da regra do trapézio obtém-se a fórmula  $y_k = y_{k-1} + (h/2) [f_{k-1} + f_k] \equiv F(y_k)$ , que é um método implícito de passo simples: necessita de apenas um valor inicial, porém  $y_k$  está em função dele mesmo ( $y_k = F(y_k)$ ). Para resolver um PVI utilizando tal método, necessita-se de um processo predictor-corretor. Esse processo consiste em obter uma estimativa para  $y_k$  (denotada por  $y_k^{(p)}$ ) através de um método explícito de mesma ordem que o método dado. Em seguida, resolvendo-se  $y_k = F(y_k^{(p)})$ , corrige-se esta estimativa.

ii) Considere o PVI apresentado anteriormente; faça um passo do predictor-corretor, com uma única correção. Para o predictor, escolha o mais adequado dos seguintes métodos: Euler ou Runge-Kutta de 2ª ordem, justifique a escolha.

6) Considere o seguinte método para a solução numérica de PVI:



$$y_{k+1} = y_{k-3} + (4h/3) [2f_{k-2} - f_{k-1} + 2f_k],$$

com erro local  $E_{loc} = (14/45)h^5 f^{(4)}(c)$ .

i) O método é de passo múltiplo. Quais são os valores iniciais necessários, além de  $y_0$ ? Justifique. Os valores iniciais devem ser calculados por um método de qual ordem?

ii) Dado o PVI:  $y' - x^2 + y^2 = 0$ ;  $y(1) = 1$ , obtenha uma aproximação para  $y(1.8)$ , com o método dado; use  $h = 0.2$  e os valores iniciais  $y(1.2) = 0.96$ ,  $y(1.4) = 0.89$  e  $y(1.6) = 0.79$ .

7) O método de Runge-Kutta explícito de dois estágios e ordem dois possui equação dada por:  $y_k = y_{k-1} + h \{b_1 f_{k-1} + b_2 f(x_{k-1} + hc_2, y_{k-1} + h c_2 f_{k-1})\}$ , onde,  $b_1 + b_2 = 1$  e  $b_2 c_2 = 1/2$ .

Considere o PVI:  $y' = -20y$ ,  $y(0) = y_0 = 1$ .

i) Mostre que  $y_k = y_{k-1} p(h)$ , onde  $p(h) = 1 - 20h + 200h^2$ .

ii) Use o item anterior para concluir que  $y_k = y_0 [p(h)]^k$ .

iii) Mostre que  $p(h) > 1 \Leftrightarrow h > 0.1$ .

iv) Prove que  $h > 0.1 \Rightarrow \lim_{k \rightarrow +\infty} y_k = +\infty$ .

v) O que acontece com a solução numérica se  $h = 0.1$ ?

vi) Sabendo que a solução analítica do PVI é  $y(x) = e^{-20x}$  e que  $\lim_{x \rightarrow +\infty} y(x) = 0$ , explique por que  $h < 0.1$  é uma escolha adequada para se obter aproximações com este método. Lembre-se de que  $\lim_{k \rightarrow +\infty} a^k = 0$ , se  $|a| < 1$ .

vii) Obtenha uma aproximação para  $y(1/20)$ , usando  $h = 0.001$ .

## Atividade 2 – Método do Ponto Médio

**Prezado(a) aluno(a),** Você poderá encontrar vários problemas práticos relacionados a problemas de valor inicial na página da Faculdade de Matemática da UFU: [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) (Laboratório - Unidade 6).

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre Métodos Baseados em série de Taylor: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 6), ambos localizados em [www.portal.famat.ufu.br/node/278](http://www.portal.famat.ufu.br/node/278) e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) A equação diferencial  $y'(t) = f(t, y(t)) = \frac{1}{\ln(y(t))}$ , com  $y(t) > 0$ , pode ser resolvida analiticamente utilizando o método das variáveis separáveis, visto em Cálculo 3. A função incógnita não tem uma forma explícita, ou seja, ela é dada implicitamente por meio da seguinte equação:  $y(t)[\ln(y(t)) - 1] = t + K$ , onde  $K$  é uma constante. Por exemplo, se considerarmos a condição inicial  $y(1) = e > 1$ , a constante  $K$  será igual a  $-1$ . Nesse caso, podemos assumir que a função  $y(t)$  é crescente em uma vizinhança de  $t = 1$ .

(a) Obtenha uma aproximação para  $y(2)$  usando o método do ponto médio:

$$y_i = y_{i-2} + 2h.f(t_{i-1}, y_{i-1}), \text{ com } n = 4.$$

Utilize um método de mesma ordem para obter o valor inicial  $y_1$ . Observe que a função  $f(t, y) = 1/(\ln(y))$  não depende diretamente da variável  $t$ . Nesse caso, a equação é dita autônoma.

(b) Encontre  $y(2)$  utilizando o método de Newton – Raphson. Para isso, resolva o problema de zero de função:  $F(y(t)) = y(t) \cdot [\ln(y(t)) - 1] - t - K = 0$ , com  $t = 2$  e  $K = -1$ . Mostre que a função  $F(y) = y \cdot [\ln(y) - 1] - 1$  muda de sinal no intervalo  $[3, 4]$  e que  $F'(y) > 0$ . Portanto, nesse intervalo existe uma única raiz da equação  $F(y) = 0$ .

**Observação:** A precisão da solução obtida pelo Método do Ponto Médio é melhorada se for considerada uma quantidade maior de subintervalos ( $n$ ), ou seja, se for considerado um valor menor de  $h$  (note que  $h = 1/n$ ).

### Informações sobre a Atividade 2.

(a1.1) Obtenção do valor de  $h$  e de  $t_0$ .

O valor inicial do PVI é dado no ponto  $t_0 = 1$  e o valor final é dado no ponto  $T = 2$ . De acordo com o enunciado do exercício, o intervalo  $[t_0, T] = [1, 2]$  será particionado em  $n = 4$  subintervalos. Nesse caso, o valor do comprimento dos subintervalos é dado por

$$h = \frac{T - t_0}{n} = \frac{2 - 1}{4} = 0.25.$$

(a1.2) Obtenção de  $t_1$ ;  $t_2$ ;  $t_3$  e  $t_4$ .

$$t_1 = t_0 + h = 1.25; \quad t_2 = t_0 + 2h = 1.50; \quad t_3 = t_0 + 3h = 1.75 \quad \text{e} \quad t_4 = t_0 + 4h = 2.0.$$

(a1.3) Obtenção de  $y_0$ .

O valor inicial do PVI é dado por  $y(1) = e \approx 2,71828$ ; portanto,  $y(t_0) = y_0 = e$ .

(a2) Cálculo de  $y_1$ .

Como o Método do Ponto Médio tem ordem 2, então o mais adequado é que o valor inicial  $y_1$  seja calculado por um método auto-iniciante de ordem 2. Vamos utilizar o Método de Euler Modificado:

$$y_i = y_{i-1} + h.f(t_{i-1} + h/2, y_{i-1} + (h/2).f(t_{i-1}, y_{i-1})).$$

Note que a função do PVI é dada por  $f(t, y) = \frac{1}{\ln(y)}$ ; assim

$$f(t_0, y_0) = f(0, e) = \frac{1}{\ln(e)} = 1.$$

$$\text{Portanto, } y_1 = y_0 + \frac{0.25}{\ln(y_0 + 0.125f(t_0, y_0))} = e + \frac{0.25}{\ln(e + 0.125)} = 2.9575 \ 25674.$$

(a3) Fórmula do ponto Médio:  $y_i = y_{i-2} + 2h.f(t_{i-1}, y_{i-1})$ .

$$y_2 = y_0 + 2h.f(t_1, y_1) = e + \frac{0.5}{\ln(y_1)} = 3.179386285;$$

$$y_3 = y_1 + 2h.f(t_2, y_2) = y_1 + \frac{0.5}{\ln(y_2)} = 3.389794289;$$

$$y_4 = y_2 + 2h.f(t_3, y_3) = y_2 + \frac{0.5}{\ln(y_3)} = 3.588964102.$$

b1) Teorema do Valor Intermediário (TVI).

Observe que a função  $F(y) = y[\ln(y) - 1] - 1$  satisfaz as hipóteses do TVI no intervalo  $[3, 4]$ :  $F$  é contínua neste intervalo e  $F(3).F(4) < 0$ , pois  $F(3) < 0$  e  $F(4) > 0$ .

b2) Método de Newton – Raphson.

$$y^{(n)} = \varphi(y^{(n-1)}), \text{ onde } \varphi(y) = y - \frac{F(y)}{F'(y)} \text{ e } y^{(0)} \text{ é dado.}$$

Note que  $F'(y) = \ln(y)$ ; assim,

$$\varphi(y) = \frac{y+1}{\ln(y)}.$$

Vamos utilizar  $y^{(0)} = 3.5$ , que é o ponto médio do intervalo que contém a raiz da equação  $F(y) = 0$ . Dessa forma, obtemos:  $y^{(1)} \approx 3.592060201$ ;  $y^{(2)} \approx 3.591121573$ ;  $y^{(3)} \approx 3.591121477$ . Observe que  $F(y^{(3)}) \approx 0.42 \times 10^{-9}$ ; portanto,  $y^{(3)}$  é uma aproximação de  $y^{(2)}$  com precisão de 9 casas decimais.

### Atividade 3 – Método Iterativo do tipo Previsor - Corretor

**Prezado(a) aluno(a),**

Vimos que o Método do Ponto Médio,  $y_i = y_{i-2} + 2h.f(x_{i-1}, y_{i-1})$ , precisa de 2 valores iniciais,  $y_0$  e  $y_1$ , para ser inicializado e, portanto, é um método de passo múltiplo (passo 2). Outros métodos de passo múltiplo podem ser deduzidos considerando-se integração numérica. Um esquema numérico deste tipo é obtido como segue. Integrando-se a EDO

de um PVI no intervalo  $[x_{i-1}, x_i]$ , obtém-se que  $y(x_i) = y(x_{i-1}) + \int_{x_{i-1}}^{x_i} f(x, y(x)) dx$ . Seja

$G(x) = f(x, y(x))$ . A integral de  $G(x)$  pode ser aproximada pelos polinômios interpoladores sobre os pontos  $x_0 = x_{i-2}$ ,  $x_1 = x_{i-1}$  e  $x_2 = x_i$  ou sobre os pontos:  $x_0 = x_{i-3}$ ,  $x_1 = x_{i-2}$  e  $x_2 = x_{i-1}$ . A notação  $f_i$  denotará  $f(x_i, y_i)$ . O procedimento anterior dará origem aos seguintes métodos:

$$\text{I) } y_i = y_{i-1} + (h/12) [-f_{i-2} + 8f_{i-1} + 5f_i] = F(y_i), \text{ com erro local } O_I(h^4) = (-1/24).h^4.y^{(4)}(\eta);$$

$$\text{II) } y_i = y_{i-1} + (h/12) [5f_{i-3} - 16f_{i-2} + 23f_{i-1}] = F(y_i), \text{ com erro local } O_{II}(h^4) = (3/8).h^4.y^{(4)}(\mu).$$

O primeiro método tem passo 2 (precisa de dois valores iniciais:  $y_0$  e  $y_1$ ) e é implícito ( $y_i$  aparece tanto no lado esquerdo como no lado direito da equação I). O segundo método é explícito e tem passo 3, isto é, dados os valores iniciais  $y_0$ ,  $y_1$  e  $y_2$ , pode-se encontrar explicitamente o valor de  $y_3$ .

O esquema previsor – corretor é utilizado da seguinte forma: (i) obtenha os valores iniciais do método previsor (explícito); (ii) calcule o valor de  $y_i$  pelo método previsor e denote este valor por  $y_i^{(0)}$ ; (iii) use o valor  $y_i^{(0)}$  no lado direito do método corretor (implícito), mais especificamente, no lado direito de I, substitua  $f_i$  por  $f(x_i, y_i^{(0)})$ ; (iv) execute quantas correções forem necessárias para obter a aproximação desejada para  $y_i$ .

Note que, para cada  $i$ , está sendo executado o Método do Ponto Fixo (análogo ao estudado no tópico Zero de Função):  $y_i^{(n)} = F(y_i^{(n-1)}) = y_{i-1} + (h/12) \cdot [-f_{i-2} + 8f_{i-1}] + (5h/12) \cdot f(x_i, y_i^{(n-1)})$ . Sob certas condições (convergência do Método do Ponto Fixo:

$$\frac{5h}{12} \frac{\partial f(x_i, y)}{\partial y} < 1), y_i^{(n)} \text{ convergirá para } y_i \text{ (que é a aproximação de } y(x_i)).$$

### Enunciado da atividade

(I) Leia atentamente um dos seguintes materiais didáticos sobre métodos de passo múltiplo: (1) apostila do curso; (2) apostila do professor Castilho e slides (Aula - Unidade 6), ambos localizados na página da FAMAT: <http://www.portal.famat.ufu.br/node/278> e (3) livro de Cálculo Numérico da professora Neide Maria Bertoldi Franco.

(II) O PVI apresentado a seguir (livro da Neide) está associado a um problema de engenharia mecânica e representa a resposta (positiva) de uma certa válvula hidráulica sujeita a uma entrada de variação senoidal.

$$\frac{dy}{dt} = \sqrt{2 \cdot \left(1 - \frac{y^2}{\sin^2(t)}\right)}, \quad y_0 = y(t_0) = y(0) = 0.$$

(i) Suponha que  $y(t)$  seja contínua em  $t = 0$  e tenha derivada,  $y'(0) = \lim_{t \rightarrow 0} \frac{y(t) - y(0)}{t - 0}$ , neste ponto. Sabendo-se que  $y(0) = 0$  e  $\lim_{t \rightarrow 0} \frac{\sin(t)}{t} = 1$ , mostre que  $\lim_{t \rightarrow 0} \frac{y(t)}{\sin(t)} = y'(0)$ .

(ii) Supondo-se que  $y'(t)$  seja contínua em  $t = 0$ , mostre que  $y'(0) = \sqrt{\frac{2}{3}}$ . Para isso, use o item anterior para concluir que  $y'(0) = \lim_{t \rightarrow 0} \frac{dy}{dt} = \sqrt{2 - 2(y'(0))^2}$ .

(iii) Observe que  $f(t, y) = \sqrt{2 \cdot \left(1 - \frac{y^2}{\sin^2(t)}\right)}$ ; assim,  $f(0,0) = y'(0) = \sqrt{\frac{2}{3}}$ . Considere  $h = 0.1$  e os valores iniciais  $y_0 = 0$ ,  $y_1 = 0.081619463$  e  $y_2 = 0.162886278$  (calculados pelo método de Runge - Kutta de ordem 3). Use o método previsor–corretor dado

anteriormente (I: corretor e II: previsor), com duas iterações do corretor, para calcular  $y_3$  (que será aproximadamente 0.243477727).

### Informações sobre a Atividade 3.

(1) Como  $y(0) = 0$  e  $\lim_{t \rightarrow 0} \frac{\text{sen}(t)}{t} = 1 \rightarrow \lim_{t \rightarrow 0} \frac{t}{\text{sen}(t)} = 1$ , então

$$\lim_{t \rightarrow 0} \frac{y(t)}{\text{sen}(t)} = \lim_{t \rightarrow 0} \frac{t y(t)}{t \text{sen}(t)} = \lim_{t \rightarrow 0} \frac{y(t) - y(0)}{t - 0} \lim_{t \rightarrow 0} \frac{t}{\text{sen}(t)} = y'(0).$$

(2) Observe que

$$y'(0) = \lim_{t \rightarrow 0} \frac{dy}{dt} = \sqrt{\lim_{t \rightarrow 0} \left[ 2 \cdot \left( 1 - \frac{y^2(t)}{\text{sen}^2(t)} \right) \right]} = \sqrt{2 - 2 \lim_{t \rightarrow 0} \frac{y^2(t)}{\text{sen}^2(t)}}.$$

Do item anterior (i), segue-se que

$$y'(0) = \sqrt{2 - 2 \lim_{t \rightarrow 0} \frac{y^2(t)}{\text{sen}^2(t)}} = \sqrt{2 - 2(y'(0))^2}.$$

Assim,  $[y'(0)]^2 = 2 - 2[y'(0)]^2$ , ou seja,  $3[y'(0)]^2 = 2$ . Logo,  $y'(0) = \sqrt{\frac{2}{3}}$ .

(3) O método previsor é dado por  $y_i = y_{i-1} + (h/12) [5f_{i-3} - 16f_{i-2} + 23f_{i-1}]$ . Utilizando os valores iniciais:  $y_0 = 0$ ,  $y_1 = 0.081619463$  e  $y_2 = 0.162886278$ ; a função do PVI,  $f(t, y) =$

$\sqrt{2 \cdot \left( 1 - \frac{y^2}{\text{sen}^2(t)} \right)}$ ;  $h = 0.1$ ;  $t_0 = 0$ ;  $t_1 = 0.1$ ;  $t_2 = 0.2$ , encontramos o seguinte valor:

$$y_3^{(0)} = y_2 + (0.1/12) [5 \cdot f(t_0, y_0) - 16 \cdot f(t_1, y_1) + 23 \cdot f(t_2, y_2)] = 0.196906968 - 0.108583002 + 0.155187595 = 0.243511561.$$

(4) As iterações com o método corretor são dadas por

$$y_i^{(n)} = F(y_i^{(n-1)}) = y_{i-1} + (h/12) \cdot [-f_{i-2} + 8 \cdot f_{i-1}] + (5h/12) \cdot f(t_i, y_i^{(n-1)}).$$

Como  $f(t, y) = \sqrt{2 \cdot \left( 1 - \frac{y^2}{\text{sen}^2(t)} \right)}$ ;  $h = 0.1$ ;  $t_0 = 0$ ;  $t_1 = 0.1$ ;  $t_2 = 0.2$ ;  $y_0 = 0$ ;

$$y_1 = 0.081619463; \quad y_2 = 0.162886278; \quad y_3^{(0)} = 0.243511561, \text{ então}$$

$$y_3^{(1)} = y_2 + (h/12)[-f_1 + 8f_2] + (5h/12)f(t_3, y_3^{(0)}) = 0.162886278 - (0.1/12) 0.814372517 + (0.8/12) 0.809674413 + (0.5/12) 0.801258719 = 0.243463914;$$

$$y_3^{(2)} = y_2 + (h/12)[-f_1 + 8f_2] + (5h/12)f(t_3, y_3^{(1)}) = 0.162886278 - (0.1/12) 0.814372517 + (0.8/12) 0.809674413 + (0.5/12) 0.801590236 = 0.243477727.$$

## 6. Atividade suplementar



### Atividade 1

**Parte 1.** Deduza o Método de Adams – Moulton (implícito) e o Método de Adams – Bashforth (explícito), de ordem 4. Para isso, considere o polinômio de Lagrange de grau 3 sobre os pontos: (I)  $x_{0,I} = t_{i-3}$ ,  $x_{1,I} = t_{i-2}$ ,  $x_{2,I} = t_{i-1}$  e  $x_{3,I} = t_i$  e sobre os pontos: (II)  $x_{0,II} = t_{i-4}$ ,  $x_{1,II} = t_{i-3}$ ,  $x_{2,II} = t_{i-2}$  e  $x_{3,II} = t_{i-1}$ . Usando a mesma notação da Seção 4.1, estes polinômios serão representados por  $p^{(\vartheta)}(t) = G(x_{0,\vartheta})L_{0,\vartheta}(t) + G(x_{1,\vartheta})L_{1,\vartheta}(t) + G(x_{2,\vartheta})L_{2,\vartheta}(t) + G(x_{3,\vartheta})L_{3,\vartheta}(t)$ , onde

$$L_{0,\vartheta}(t) = \frac{(t-x_{1,\vartheta})(t-x_{2,\vartheta})(t-x_{3,\vartheta})}{(x_{0,\vartheta}-x_{1,\vartheta})(x_{0,\vartheta}-x_{2,\vartheta})(x_{0,\vartheta}-x_{3,\vartheta})};$$

$$L_{1,\vartheta}(t) = \frac{(t-x_{0,\vartheta})(t-x_{2,\vartheta})(t-x_{3,\vartheta})}{(x_{1,\vartheta}-x_{0,\vartheta})(x_{1,\vartheta}-x_{2,\vartheta})(x_{1,\vartheta}-x_{3,\vartheta})};$$

$$L_{2,\vartheta}(t) = \frac{(t-x_{0,\vartheta})(t-x_{1,\vartheta})(t-x_{3,\vartheta})}{(x_{2,\vartheta}-x_{0,\vartheta})(x_{2,\vartheta}-x_{1,\vartheta})(x_{2,\vartheta}-x_{3,\vartheta})};$$

$$L_{3,\vartheta}(t) = \frac{(t-x_{0,\vartheta})(t-x_{1,\vartheta})(t-x_{2,\vartheta})}{(x_{3,\vartheta}-x_{0,\vartheta})(x_{3,\vartheta}-x_{1,\vartheta})(x_{3,\vartheta}-x_{2,\vartheta})};$$

$\vartheta = I$  ou  $\vartheta = II$  e  $G(t) = f(t, y(t))$ .

Os métodos implícito e explícito serão deduzidos de forma análoga às deduções que foram feitas, anteriormente, para os métodos de ordens 2 e 3, nas Seções 4.1 e 4.2.

Dessa forma, será preciso calcular as integrais dos polinômios de grau 3,  $p^{(\vartheta)}(t)$ ,  $\vartheta = I$  e  $\vartheta = II$ , no intervalo  $[t_{i-1}, t_i]$ . Não se esqueça de utilizar a troca de variáveis:  $t = t_{i-1} + uh$ .

Os métodos obtidos deverão ter as seguintes expressões:

Adams – Moulton:

$$y_i = y_{i-1} + (h/24)\{f(t_{i-3}, y_{i-3}) - 5f(t_{i-2}, y_{i-2}) + 19f(t_{i-1}, y_{i-1}) + 9f(t_i, y_i)\}; \quad 3 \leq i \leq n,$$

e

Adams – Bashforth:

$$y_i = y_{i-1} + (h/24)\{-9f(t_{i-4}, y_{i-4}) + 37f(t_{i-3}, y_{i-3}) - 59f(t_{i-2}, y_{i-2}) + 55f(t_{i-1}, y_{i-1})\}; \quad 4 \leq i \leq n.$$

**Parte 2.** Os métodos implícito e explícito apresentados anteriormente são de ordem 4 porque os erros associados às regras de integração empregadas em suas deduções são do tipo  $O(h^5)$ . Obtenha as fórmulas dos erros utilizando a notação apresentada na Seção 4 anterior. Ou seja, calcule

$$E^{(I)} = \int_{t_{i-1}}^{t_i} N_I(t) \frac{G^{(4)}(t)}{4!} dt, \quad \text{onde } N_I(t) = (t - t_{i-3})(t - t_{i-2})(t - t_{i-1})(t - t_i) < 0, \quad \forall t \in (t_{i-1}, t_i);$$

e

$$E^{(II)} = \int_{t_{i-1}}^{t_i} N_{II}(t) \frac{G^{(4)}(t)}{4!} dt, \quad \text{onde } N_{II}(t) = (t - t_{i-4})(t - t_{i-3})(t - t_{i-2})(t - t_{i-1}) > 0, \quad \forall t \in (t_{i-1}, t_i).$$

Suponha que a derivada quarta  $G^{(4)}$  seja contínua; use o Teorema do Valor Intermediário para integrais (veja as observações apresentadas na Seção 5.1 do Módulo 3) e a troca de variáveis  $t = t_{i-1} + uh$ .

O resultado esperado é o seguinte:

$$E^{(I)} = \frac{G^{(4)}(c)}{24} \int_{t_{i-1}}^{t_i} N_I(t) dt = \frac{-19 h^5 G^{(4)}(c)}{720}, \quad c \in (t_{i-1}, t_i).$$

$$E^{(II)} = \frac{G^{(4)}(C)}{24} \int_{t_{i-1}}^{t_i} N_{II}(t) dt = \frac{251 h^5 G^{(4)}(C)}{720}, \quad C \in (t_{i-1}, t_i).$$

**Parte 3.** Considere o Problema de Valor Inicial exibido na atividade 3 da Seção 5 anterior:



$$\frac{dy}{dt} = \sqrt{2 \cdot \left(1 - \frac{y^2}{\sin^2(t)}\right)}, \quad y_0 = y(t_0) = y(0) = 0.$$

Este PVI está associado a um problema de engenharia mecânica e representa a resposta (positiva) de uma certa válvula hidráulica sujeita a uma entrada de variação senoidal (referência 6 do Módulo 1: Neide Bertoldi Franco – Cálculo Numérico).

O código computacional apresentado a seguir é a implementação do método predictor corretor de Adams – Moulton de ordem 4. Os valores iniciais são calculados pelo método de Runge-Kutta de ordem 4.

Entenda o código computacional e desenvolva o seu próprio código para resolver outros Problemas de Valor Inicial.

**Observação:** Como foi mostrado na atividade 3 deste módulo (Seção 5), a função do PVI anterior no ponto  $t = 0$  é dado por  $f(0,0) = \sqrt{\frac{2}{3}}$ .

**CÓDIGO:** Adams – Moulton de ordem 4

%Previsor - corretor de Adams-Moulton, com valores iniciais calculados  
%pelo método de Runge-Kutta de quarta ordem.

%Entrada de dados.

% Entre com um valor para x0 associado ao PVI y'= f(x,y); y(x0) = y0;

x0 = 0;

% Entre com um valor para y0, valor inicial do PVI y'= f(x,y); y(x0) = y0;

y0 = 0;

% Forneça o valor de x, ponto em que será calculado o valor aproximado de y(x);

x = 1.0;

% Entre com o número máximo de subintervalos de [x0, x];

n = 100;

% Entre com o número de iterações do método corretor;

max=9;

```

% {Runge-Kutta de quarta ordem}

% o vetor func armazenará os valores de f(x,y) para x e y associados a cada
% um dos quatro valores iniciais (y0,y1,y2,y3) necessários para o início do
% método previsor-corretor e para x e y associado à estimativa para o previsor (y4).

func(1)=0;

func(2)=0;

func(3)=0;

func(4)=0;

func(5)=0;

h=(x-x0)/n % comprimento dos subintervalos.

func(1)=f_pc(x0,y0);

for i=2:1:4
    % Cálculo dos 4 Estágios do método de R-K

    k1 = h*f_pc(x0,y0);

    k2 = h*f_pc(x0+h/2,y0+k1/2);

    k3 = h*f_pc(x0+h/2,(y0+(k2/2)));

    k4 = h*f_pc(x0+h,(y0+k3));

    % Cálculo aproximado de y(x) pelo método de R-K

    y = y0 + (1/6.)*(k1 + 2*k2 + 2*k3 + k4);

    y0 = y;

    x0 = x0 + h;

    func(i) = f_pc(x0,y0);
end

```

```

for j=5:1:n+1

    %{Adams-Bashforth: PREVISOR}

    yp = y + (h/24.0)*(55.0*func(4) - 59.0*func(3) + 37.0*func(2) - 9.0*func(1));

    xx=x0+h;

    func(5) = f_pc(xx,yp);

    %valor que se mantem fixo durante o processo corretor:
    yc = y + (h/24.0)*(19.0*func(4) - 5.0*func(3) + func(2));

    for k=1:1:max

        %{Adams-Mouton: CORRETOR}

        y = yc+(((h*9.0)/24)*func(5));

        func(5) = f_pc(xx,y);
    end

    func(1) = func(2);

    func(2) = func(3);

    func(3) = func(4);

    func(4) = func(5);

    x0 = x0 + h;

end

fprintf('Para ');

x

fprintf('O valor estimado para y(x) eh dado por y = %12.20f\n', y);

%Fim do Código

```

**Observação:** No código anterior, a função do PVI foi definida no arquivo f\_pc.m, o qual tem o seguinte conteúdo:

```
function g = f_pc(x,y)

    if(x == 0 && y == 0)
        g = sqrt(2/3);
    else
        g = y/sin(x);
        g = g*g;
        g = sqrt(2*(1 - g));
    end

    %Fim da Rotina
```

**Observação:** O valor da solução do PVI em  $t = 1$ , considerando 100 subintervalos e 9 iterações com o método corretor, em cada passo, é dada por  $y_{100} = 0.75596114186817875424$ .

## Atividade 2

Faça um código computacional implementando o Método Previsor Corretor de ordem 3 que foi deduzido na Seção 4.2. Use o Método de Runge – Kutta de ordem 3 (veja a Seção 3.4) para calcular os valores iniciais do método previsor.

## Referências Bibliográficas

ÁVILA, G. *Análise Matemática para a Licenciatura*, São Paulo, Edgard Blucher, 2006.

ÁVILA, G. *Introdução à análise matemática*, São Paulo, Edgard Blucher, 1992.

BASSANEZI, R. C., *Ensino-aprendizagem com modelagem matemática*, São Paulo, Contexto, 2002.

BOLDRINI, J. L, COSTA, S. I. R., RIBEIRO, V. L. F. F. E WETZLER, H.G., *Álgebra Linear*, 3ª Edição, São Paulo, Harper & Row do Brasil, 1980.

BURDEN, R.L., FAIRES, J.D., *Análise Numérica* - tradução da 8ª edição norte-americana, São Paulo, Cengage Learning, 721 p., 2008.

BUTCHER, J.C. *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. New York, NY, USA, Wiley-Interscience, 1987.

CHAPRA, S.C., CANALE, R.P., *Métodos Numéricos para Engenharia*, São Paulo, McGraw-Hill, 809 p., 2008.

FIGUEIREDO, Djairo Guedes de, *Análise de Fourier e Equações Diferenciais Parciais*. 2ª Ed., Rio de Janeiro, IMPA, 1987.

FIGUEIREDO, D. G. *Análise 1*. 2ª Edição, São Paulo, Livros Técnicos e Científicos Editora S/A, 1996.

FRANCO, Neide Bertoldi, *Cálculo Numérico*, São Paulo: Pearson Prentice Hall, 2006.

ISSACSON, E. and KELLER, H. B., *Analysis of numerical methods*, New York, John Wiley and Sons, 1982.

LIMA, E. L. *Curso de análise*. Volume 1. Projeto Euclides, Rio de Janeiro, SBM, 2000.

LIMA, E. L. *Análise real*. Volume 1. Coleção Matemática Universitária, Rio de Janeiro, SBM, 2001.

LIMA, Elon Lages, *Curso de Análise*. Volume 2 (Projeto Euclides). Rio de Janeiro, Instituto de Matemática Pura e Aplicada, CNPq, 2000.

ORTEGA, J. M., *Numerical analysis; a second course*, New York, Academic Press, 1972.

RUGGIERO, M.A.G., LOPES, V.L.R., *Cálculo Numérico: Aspectos Teóricos e Computacionais*, 2ª Edição, São Paulo, Pearson Education do Brasil, 1996.