

ZEROS REAIS DE FUNÇÕES REAIS

2.1 INTRODUÇÃO

Nas mais diversas áreas das ciências exatas ocorrem, freqüentemente, situações que envolvem a resolução de uma equação do tipo $f(x) = 0$.

Consideremos, por exemplo, o seguinte circuito:

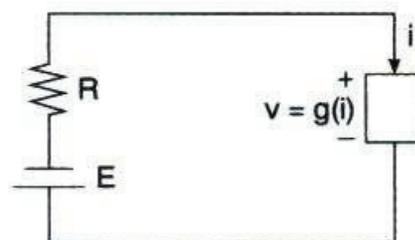


Figura 2.1

A figura acima representa um dispositivo não linear, isto é, a função g que dá a tensão em função da corrente é não linear. Dados E e R e supondo conhecida a característica do dispositivo $v = g(i)$, se quisermos saber a corrente que vai fluir no circuito temos de resolver a equação $E - Ri - g(i) = 0$ (pela lei de Kirchoff). Na prática, $g(i)$ tem o aspecto de um polinômio do terceiro grau.

Queremos então resolver a equação $f(i) = E - Ri - g(i) = 0$.

O objetivo deste capítulo é o estudo de métodos numéricos para resolução de equações não lineares como a acima.

Um número real ξ é um zero da função $f(x)$ ou uma raiz da equação $f(x) = 0$ se $f(\xi) = 0$.

Em alguns casos, por exemplo, de equações polinomiais, os valores de x que anulam $f(x)$ podem ser reais ou complexos. Neste capítulo, estaremos interessados somente nos zeros reais de $f(x)$.

Graficamente, os zeros reais são representados pelas abscissas dos pontos onde uma curva intercepta o eixo \vec{ox} .

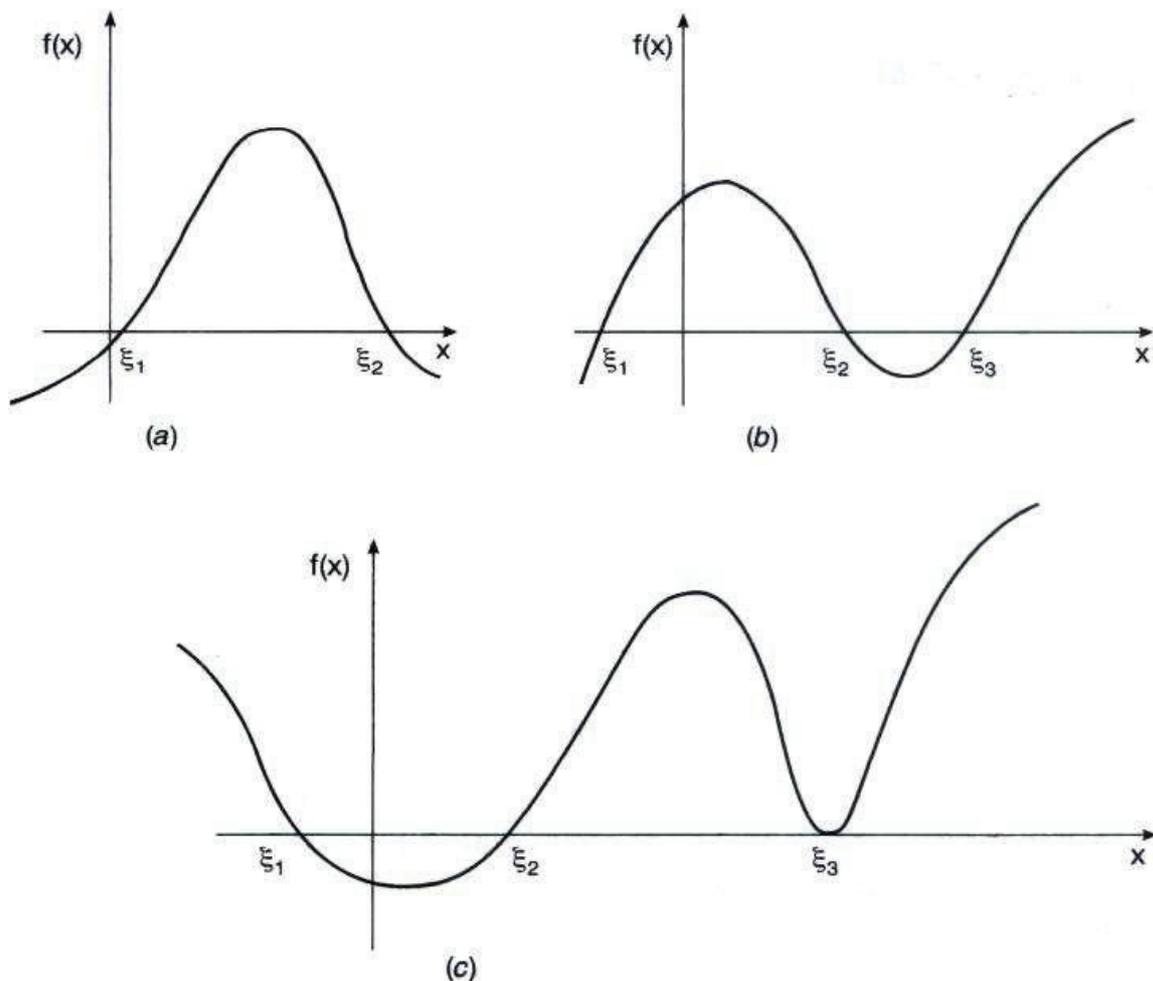


Figura 2.2

Como obter raízes reais de uma equação qualquer?

Sabemos que, para algumas equações, como por exemplo as equações polinomiais de segundo grau, existem fórmulas explícitas que dão as raízes em função dos coeficientes. No entanto, no caso de polinômios de grau mais alto e no caso de funções mais complicadas, é praticamente impossível se achar os zeros exatamente. Por isso, temos de nos contentar em encontrar apenas aproximações para esses zeros; mas isto não é uma limitação muito séria, pois, com os métodos que apresentaremos, conseguimos, a menos de limitações de máquinas, encontrar os zeros de uma função com qualquer precisão prefixada.

A idéia central destes métodos é partir de uma aproximação inicial para a raiz e em seguida refinar essa aproximação através de um processo iterativo.

Por isso, os métodos constam de duas fases:

FASE I: Localização ou isolamento das raízes, que consiste em obter um intervalo que contém a raiz;

FASE II: Refinamento, que consiste em, escolhidas aproximações iniciais no intervalo encontrado na Fase I, melhorá-las sucessivamente até se obter uma aproximação para a raiz dentro de uma precisão ϵ prefixada.

2.2 FASE I: ISOLAMENTO DAS RAÍZES

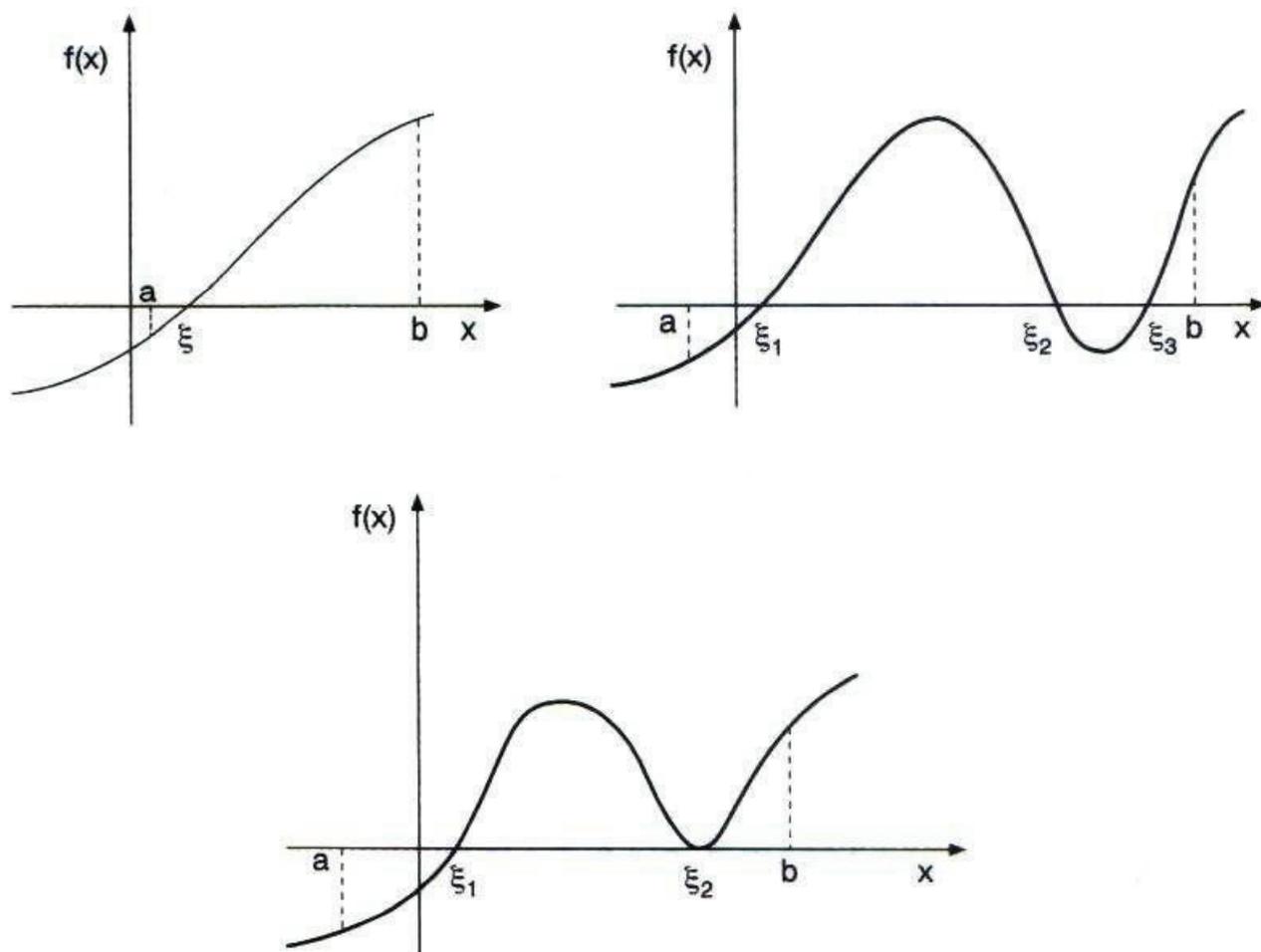
Nesta fase é feita uma análise teórica e gráfica da função $f(x)$. É importante ressaltar que o sucesso da Fase II depende fortemente da precisão desta análise.

Na análise teórica usamos freqüentemente o teorema:

TEOREMA 1

Seja $f(x)$ uma função contínua num intervalo $[a, b]$.

Se $f(a)f(b) < 0$ então existe pelo menos um ponto $x = \xi$ entre a e b que é zero de $f(x)$.

GRAFICAMENTE**Figura 2.3**

Conforme vemos, a interpretação gráfica deste teorema é extremamente simples (e uma demonstração deste resultado pode ser encontrada em [11]).

OBSERVAÇÃO

Sob as hipóteses do teorema anterior, se $f'(x)$ existir e preservar sinal em (a, b) , então este intervalo contém um único zero de $f(x)$.

GRAFICAMENTE

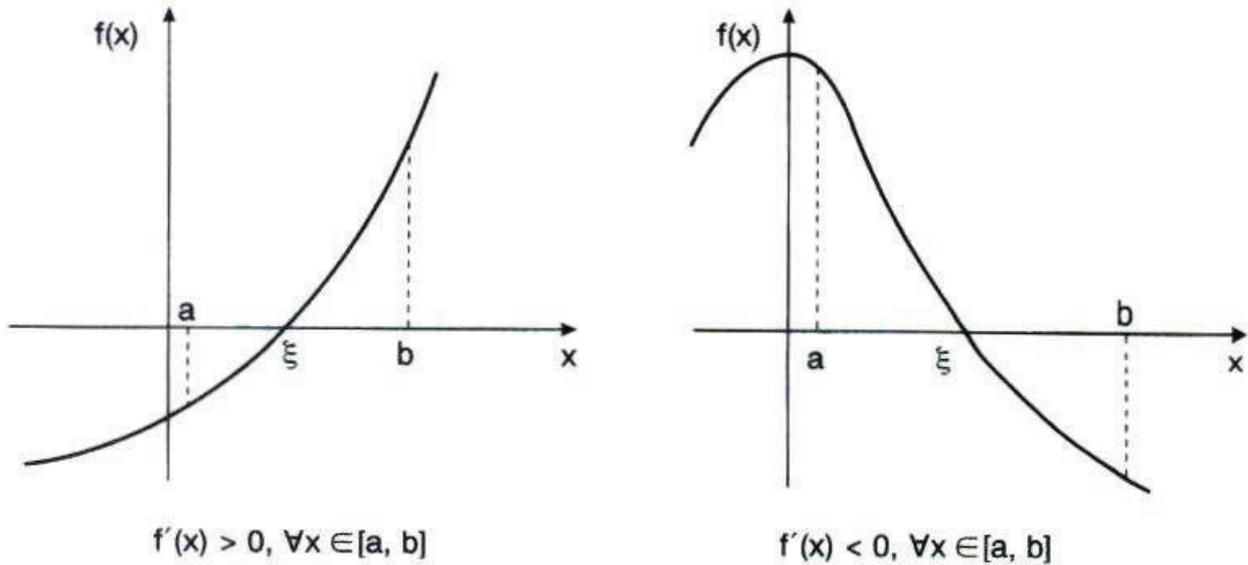


Figura 2.4

Uma forma de se isolar as raízes de $f(x)$ usando os resultados anteriores é tabelar $f(x)$ para vários valores de x e analisar as mudanças de sinal de $f(x)$ e o sinal da derivada nos intervalos em que $f(x)$ mudou de sinal.

Exemplo 1

a) $f(x) = x^3 - 9x + 3$

Construindo uma tabela de valores para $f(x)$ e considerando apenas os sinais, temos:

x	$-\infty$	-100	-10	-5	-3	-1	0	1	2	3	4	5
f(x)	-	-	-	-	+	+	+	-	-	+	+	+

Sabendo que $f(x)$ é contínua para qualquer x real e observando as variações de sinal, podemos concluir que cada um dos intervalos $I_1 = [-5, -3]$, $I_2 = [0, 1]$ e $I_3 = [2, 3]$ contém pelo menos um zero de $f(x)$.

Como $f(x)$ é polinômio de grau 3, podemos afirmar que cada intervalo contém um único zero de $f(x)$; assim, localizamos todas as raízes de $f(x) = 0$.

$$b) f(x) = \sqrt{x} - 5e^{-x}$$

Temos que $D(f) = \mathbb{R}^+$ ($D(f) \equiv$ domínio de $f(x)$)

Construindo uma tabela de valores com o sinal de $f(x)$ para determinados valores de x temos:

x	0	1	2	3	...
$f(x)$	-	-	+	+	...

Analisando a tabela, vemos que $f(x)$ admite pelo menos um zero no intervalo $(1, 2)$.

Para se saber se este zero é único neste intervalo, podemos usar a observação anterior, isto é, analisar o sinal de $f'(x)$:

$$f'(x) = \frac{1}{2\sqrt{x}} + 5e^{-x} > 0, \quad \forall x > 0.$$

Assim, podemos concluir que $f(x)$ admite um único zero em todo seu domínio de definição e este zero está no intervalo $(1, 2)$.

OBSERVAÇÃO

Se $f(a)f(b) > 0$ então podemos ter várias situações no intervalo $[a, b]$, conforme mostram os gráficos:

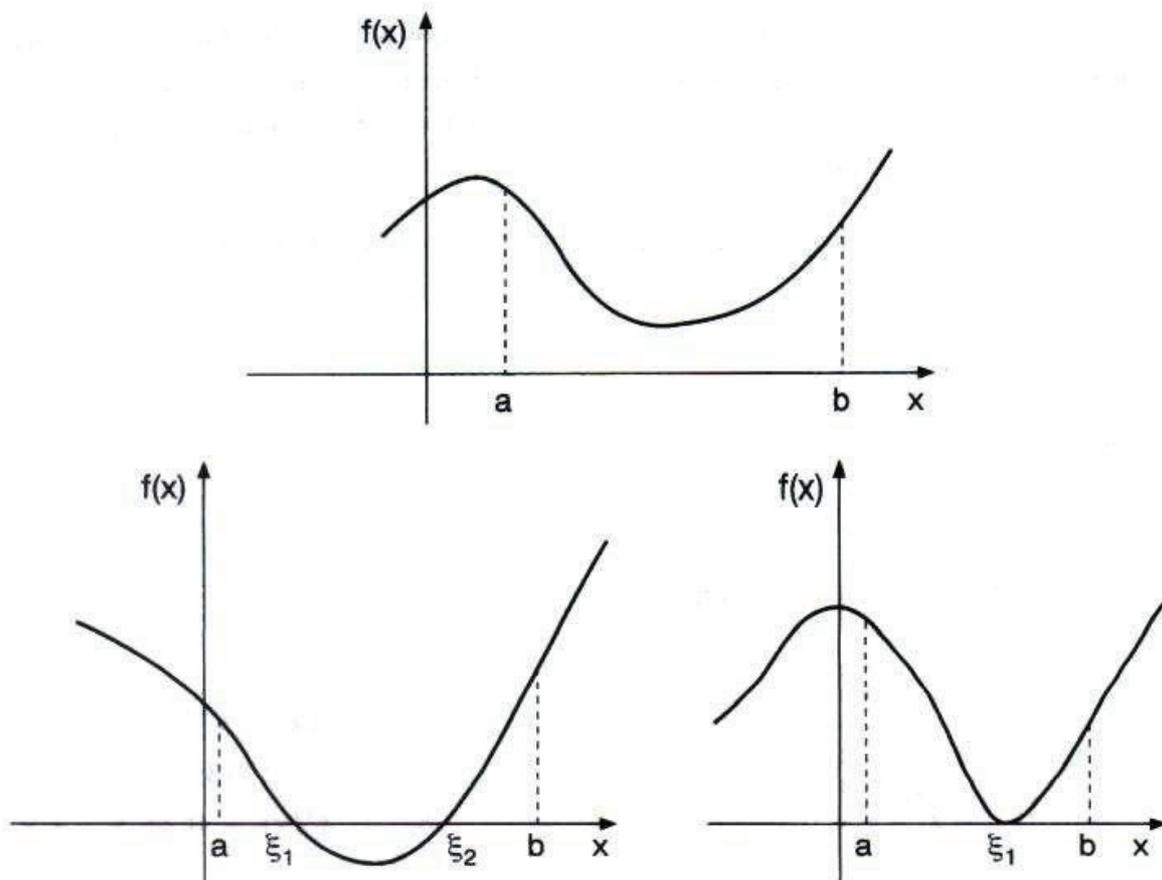


Figura 2.5

A análise gráfica da função $f(x)$ ou da equação $f(x) = 0$ é fundamental para se obter boas aproximações para a raiz.

Para tanto, é suficiente utilizar um dos seguintes processos:

- i)* esboçar o gráfico da função $f(x)$ e localizar as abcissas dos pontos onde a curva intercepta o eixo \vec{ox} ;
- ii)* a partir da equação $f(x) = 0$, obter a equação equivalente $g(x) = h(x)$, esboçar os gráficos das funções $g(x)$ e $h(x)$ no mesmo eixo cartesiano e localizar os pontos x onde as duas curvas se interceptam, pois neste caso $f(\xi) = 0 \Leftrightarrow g(\xi) = h(\xi)$;
- iii)* usar os programas que traçam gráficos de funções, disponíveis em algumas calculadoras ou softwares matemáticos.

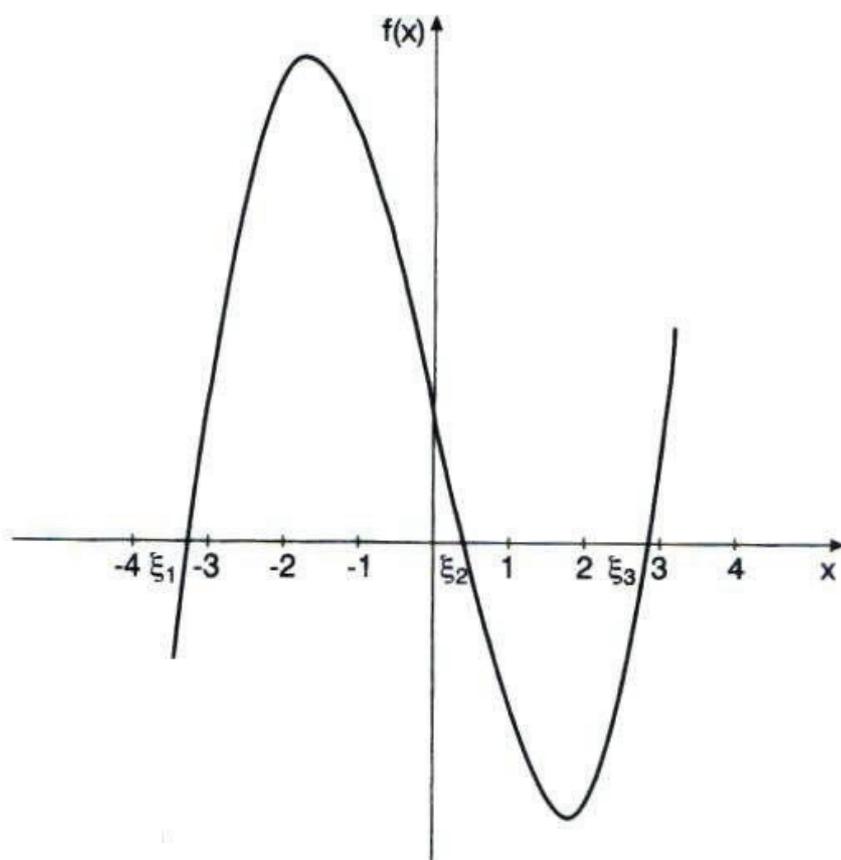
O esboço do gráfico de uma função requer um estudo detalhado do comportamento desta função, que envolve basicamente os itens: domínio da função; pontos de descontinuidade; intervalos de crescimento e decrescimento; pontos de máximo e mínimo; concavidade; pontos de inflexão e assíntotas da função.

Este esquema geral de análise de funções e construção de gráficos é encontrado em [16] e [20].

Exemplo 2

a) $f(x) = x^3 - 9x + 3$

Usando o processo (i), temos:



$$\begin{aligned} f(x) &= x^3 - 9x + 3 \\ f'(x) &= 3x^2 - 9 \\ f'(x) = 0 &\Leftrightarrow x = \pm\sqrt{3} \end{aligned}$$

x	f(x)
-4	-25
-3	3
$-\sqrt{3}$	13.3923
-1	11
0	3
1	-5
$\sqrt{3}$	-7.3923
2	-7
3	3

$$\xi_1 \in (-4, -3)$$

$$\xi_2 \in (0, 1)$$

$$\xi_3 \in (2, 3)$$

Figura 2.6

E, usando o processo (ii): da equação $x^3 - 9x + 3 = 0$, podemos obter a equação equivalente $x^3 = 9x - 3$. Neste caso, temos $g(x) = x^3$ e $h(x) = 9x - 3$. Assim,

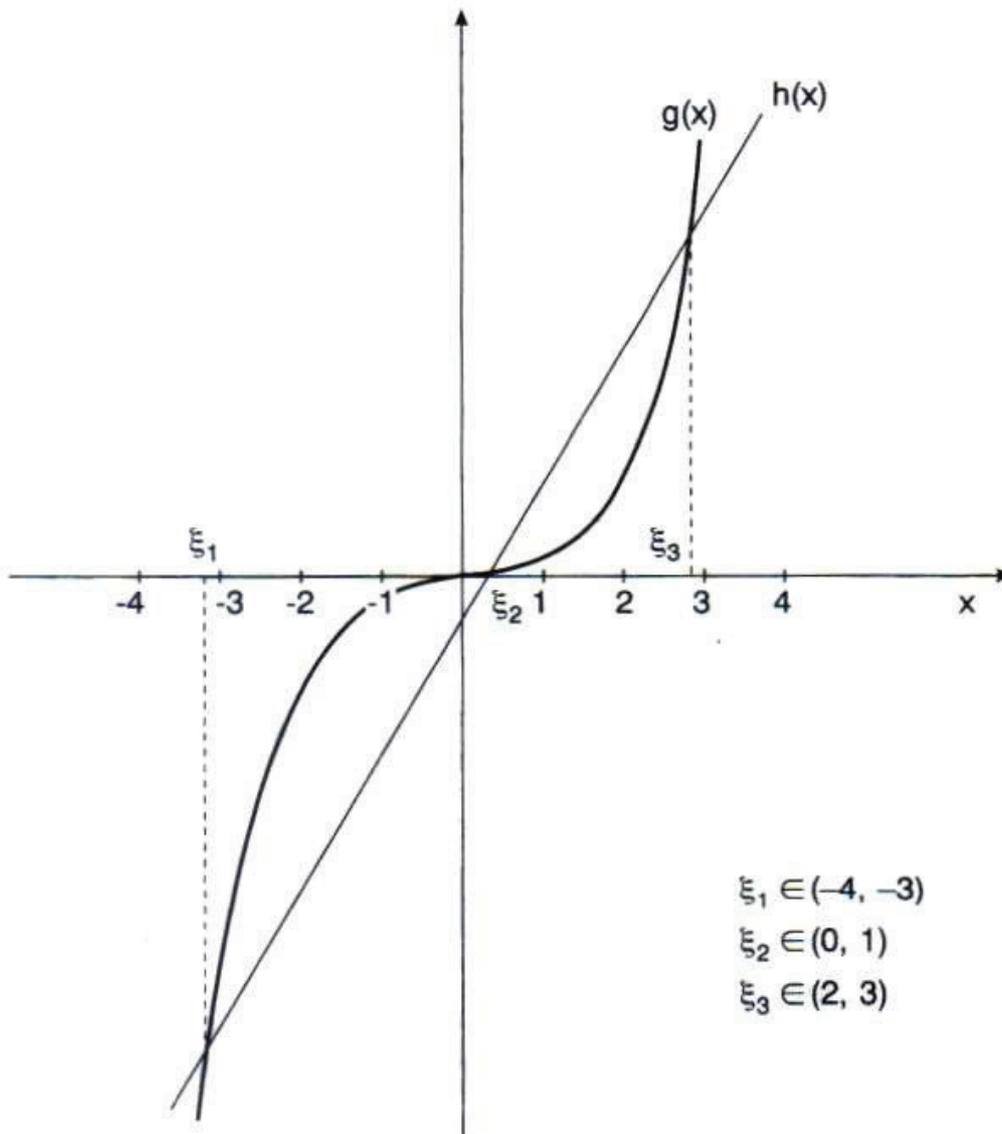


Figura 2.7

$$b) f(x) = \sqrt{x} - 5e^{-x}$$

Neste caso, é mais conveniente usar o processo (ii):

$$\sqrt{x} - 5e^{-x} = 0 \Leftrightarrow \sqrt{x} = 5e^{-x} \Rightarrow g(x) = \sqrt{x} \text{ e } h(x) = 5e^{-x}$$

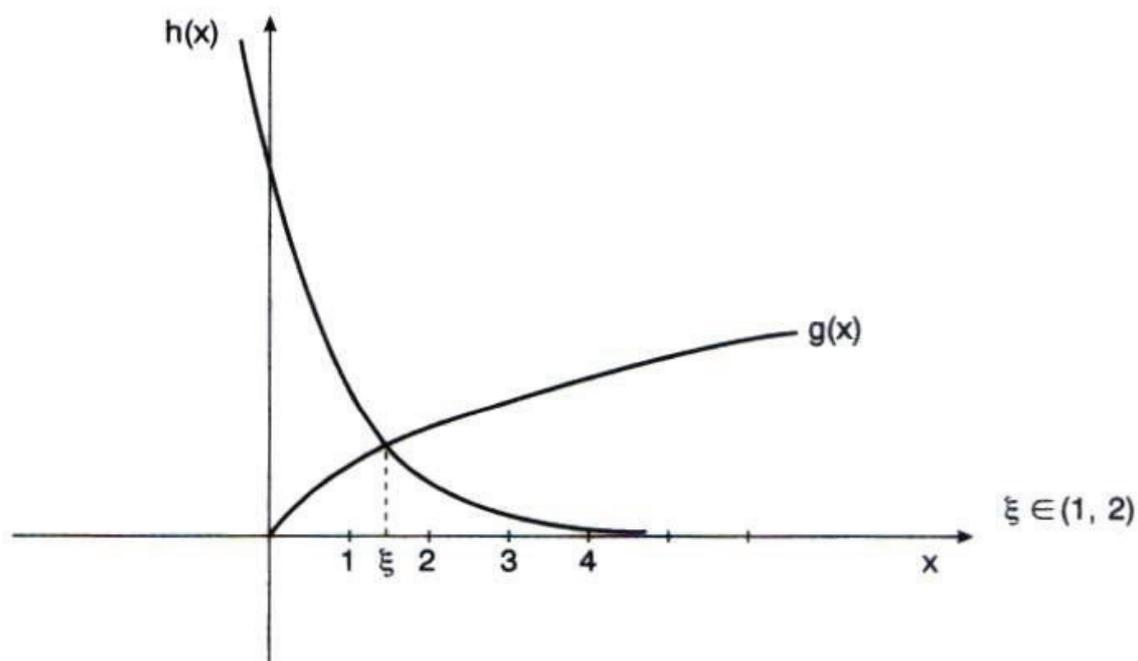


Figura 2.8

$$c) f(x) = x \log(x) - 1$$

Usando novamente o processo (ii) temos que

$$x \log(x) - 1 = 0 \Leftrightarrow \log(x) = \frac{1}{x} \Rightarrow g(x) = \log(x) \text{ e } h(x) = \frac{1}{x}$$

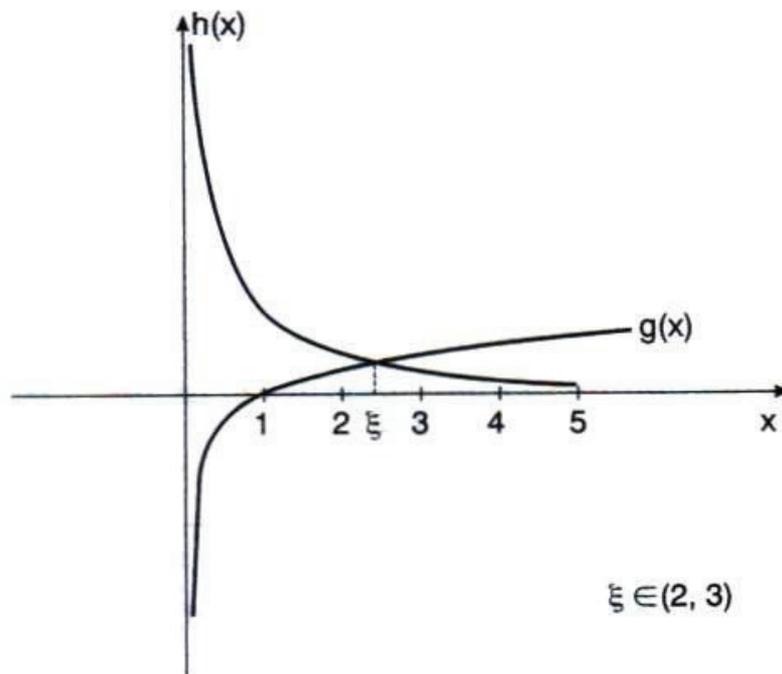


Figura 2.9

2.3 FASE II: REFINAMENTO

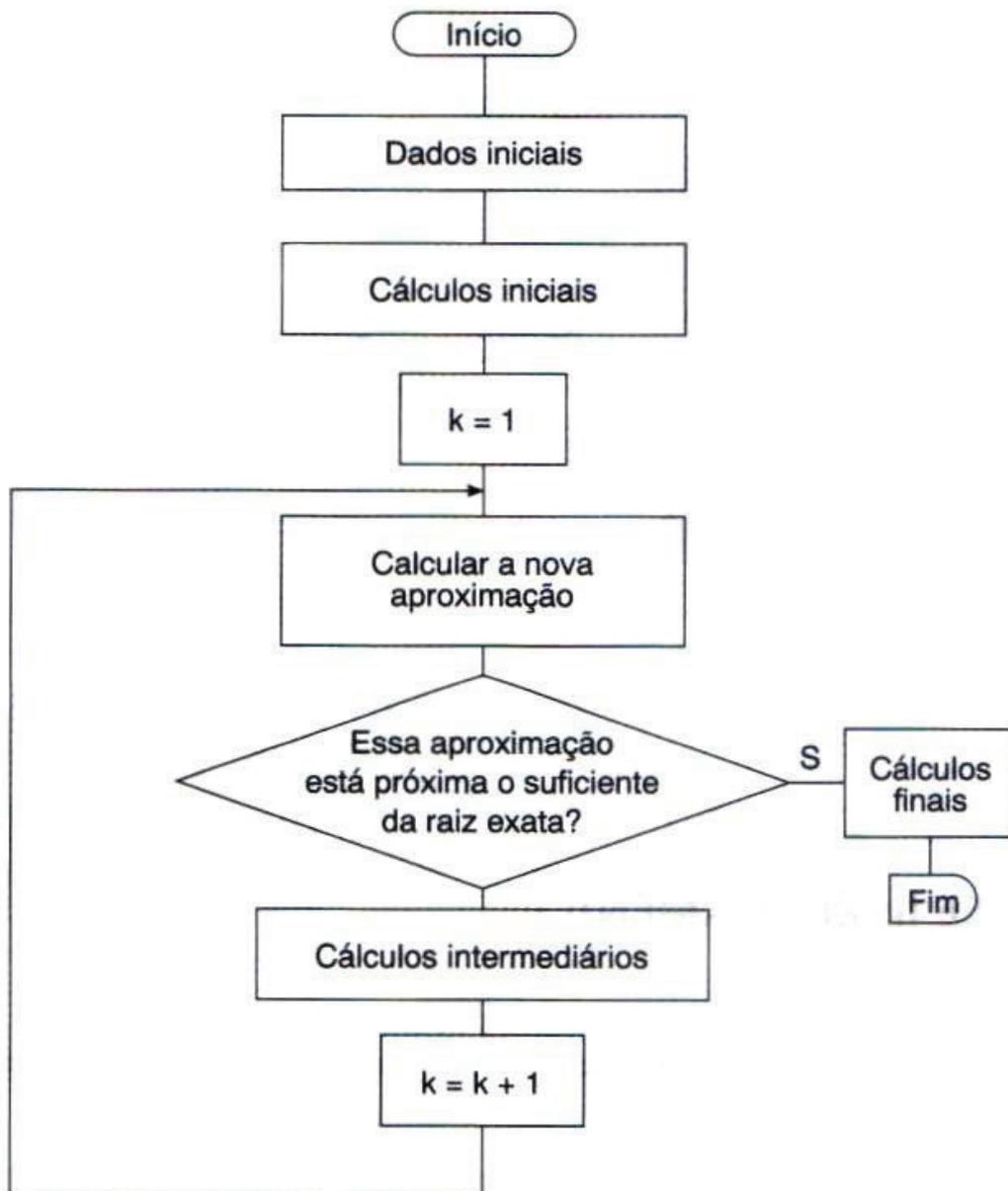
Estudaremos neste item vários métodos numéricos de refinamento de raiz. A forma como se efetua o refinamento é que diferencia os métodos. Todos eles pertencem à classe dos métodos iterativos.

Um *método iterativo* consiste em uma seqüência de instruções que são executadas passo a passo, algumas das quais são repetidas em ciclos.

A execução de um ciclo recebe o nome de *iteração*. Cada iteração utiliza resultados das iterações anteriores e efetua determinados testes que permitem verificar se foi atingido um resultado próximo o suficiente do resultado esperado.

Observamos que os métodos iterativos para obter zeros de funções fornecem apenas uma aproximação para a solução exata.

Os métodos iterativos para refinamento da aproximação inicial para a raiz exata podem ser colocados num diagrama de fluxo:

**Figura 2.10**

2.3.1 CRITÉRIOS DE PARADA

Pelo diagrama de fluxo verifica-se que todos os métodos iterativos para obter zeros de função efetuam um teste do tipo:

x_k está suficientemente próximo da raiz exata?

Que tipo de teste efetuar para se verificar se x_k está suficientemente próximo da raiz exata? Para isto é preciso entender o significado de raiz aproximada.

Existem duas interpretações para raiz aproximada que nem sempre levam ao mesmo resultado:

\bar{x} é raiz aproximada com precisão ϵ se:

i) $|\bar{x} - \xi| < \epsilon$ ou

ii) $|f(\bar{x})| < \epsilon$.

Como efetuar o teste (i) se não conhecemos ξ ?

Uma forma é reduzir o intervalo que contém a raiz a cada iteração. Ao se conseguir um intervalo $[a, b]$ tal que:

$$\left. \begin{array}{l} \xi \in [a, b] \\ e \\ b - a < \epsilon \end{array} \right\} \text{então } \forall x \in [a, b], |x - \xi| < \epsilon. \text{ Portanto, } \forall x \in [a, b] \text{ pode ser tomado como } \bar{x}$$

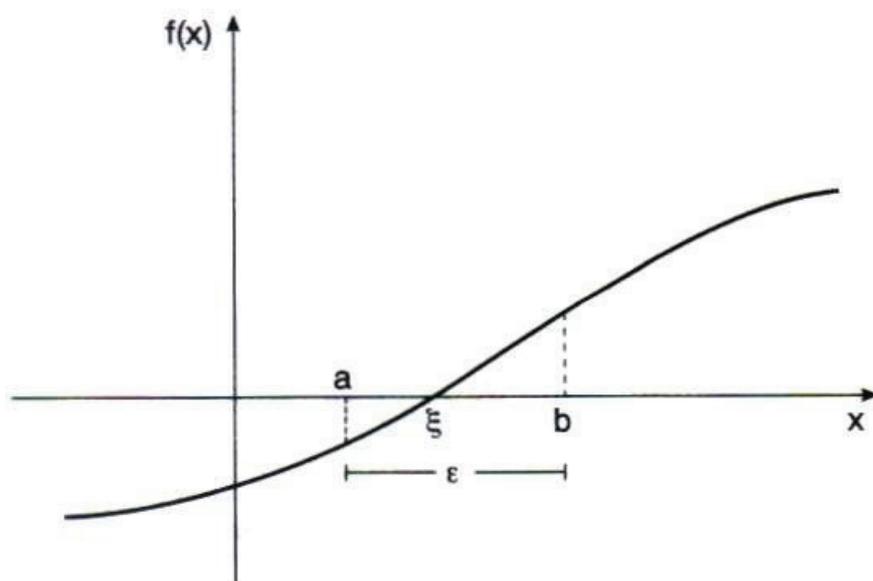


Figura 2.11

Nem sempre é possível ter as exigências (i) e (ii) satisfeitas simultaneamente. Os gráficos a seguir ilustram algumas possibilidades:

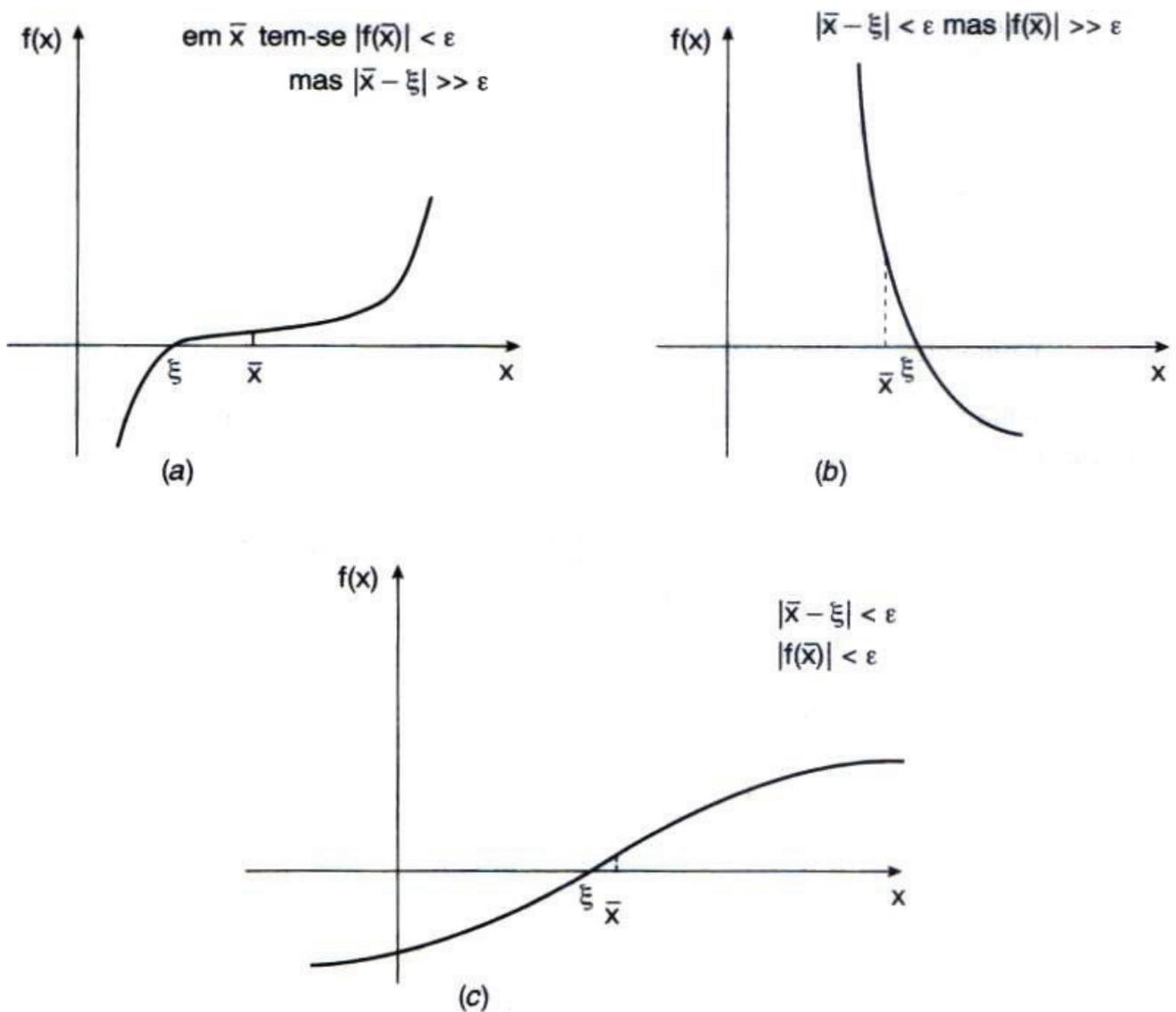


Figura 2.12

Os métodos numéricos são desenvolvidos de forma a satisfazer pelo menos um dos critérios.

Observamos que, dependendo da ordem de grandeza dos números envolvidos, é aconselhável usar teste do erro relativo, como por exemplo, considerar \bar{x} como aproximação de ξ se $\frac{|f(\bar{x})|}{L} < \epsilon$ onde $L = |f'(x)|$ para algum x escolhido numa vizinhança de ξ .

Em programas computacionais, além do teste de parada usado para cada método, deve-se ter o cuidado de estipular um *número máximo de iterações* para se evitar que o programa entre em “looping” devido a erros no próprio programa ou à inadequação do método usado para o problema em questão.

2.3.2 MÉTODOS ITERATIVOS PARA SE OBTER ZEROS REAIS DE FUNÇÕES

I. MÉTODO DA BISSECÇÃO

Seja a função $f(x)$ contínua no intervalo $[a, b]$ e tal que $f(a)f(b) < 0$.

Vamos supor, para simplificar, que o intervalo (a, b) contenha uma única raiz da equação $f(x) = 0$.

O objetivo deste método é reduzir a amplitude do intervalo que contém a raiz até se atingir a precisão requerida: $(b - a) < \epsilon$, usando para isto a sucessiva divisão de $[a, b]$ ao meio.

GRAFICAMENTE

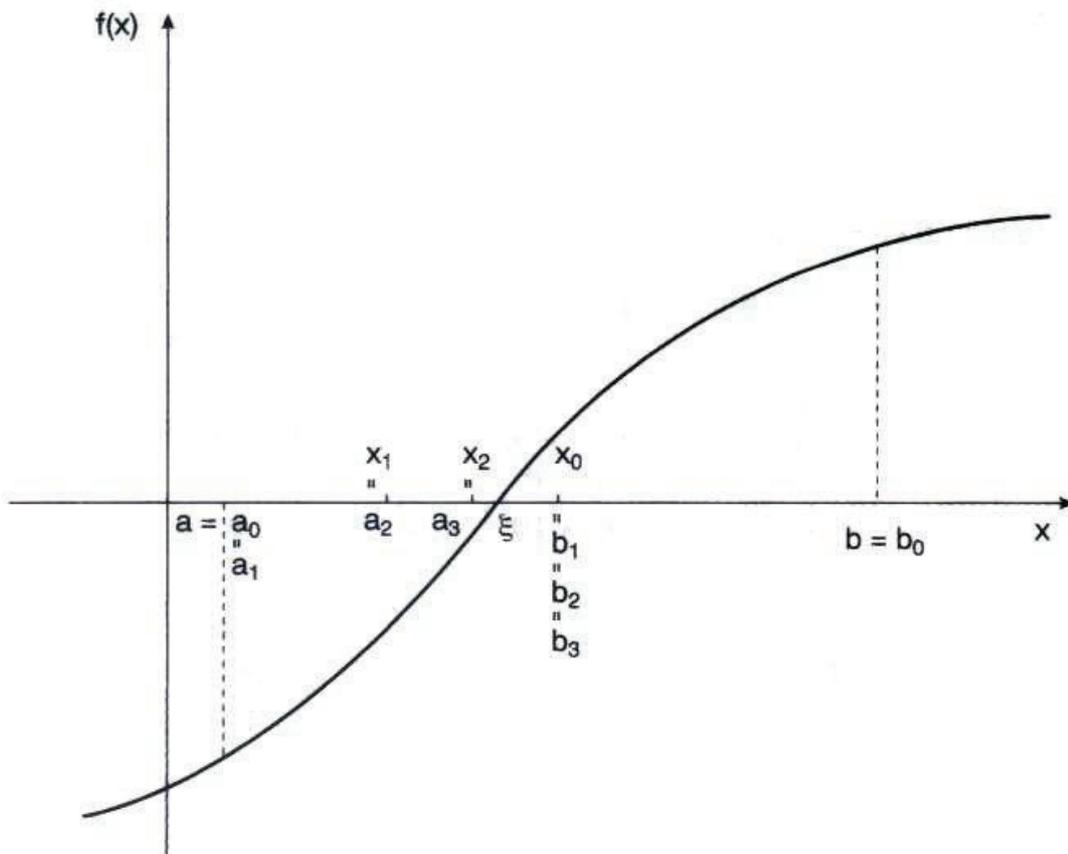


Figura 2.13

As iterações são realizadas da seguinte forma:

$$x_0 = \frac{a_0 + b_0}{2} \quad \left\{ \begin{array}{l} f(a_0) < 0 \\ f(b_0) > 0 \\ f(x_0) > 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \xi \in (a_0, x_0) \\ a_1 = a_0 \\ b_1 = x_0 \end{array} \right.$$

$$x_1 = \frac{a_1 + b_1}{2} \quad \left\{ \begin{array}{l} f(a_1) < 0 \\ f(b_1) > 0 \\ f(x_1) < 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \xi \in (x_1, b_1) \\ a_2 = x_1 \\ b_2 = b_1 \end{array} \right.$$

$$x_2 = \frac{a_2 + b_2}{2} \quad \left\{ \begin{array}{l} f(a_2) < 0 \\ f(b_2) > 0 \\ f(x_2) < 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \xi \in (x_2, b_2) \\ a_3 = x_2 \\ b_3 = b_2 \end{array} \right.$$

⋮
⋮
⋮

Exemplo 3

Já vimos que a função $f(x) = x \log(x) - 1$ tem um zero em $(2, 3)$.

O método da bissecção aplicado a esta função com $[2, 3]$ como intervalo inicial fornece:

$$x_0 = \frac{2 + 3}{2} = 2.5 \quad \left\{ \begin{array}{l} f(2) = -0.3979 < 0 \\ f(3) = 0.4314 > 0 \\ f(2.5) = -5.15 \times 10^{-3} < 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \xi \in (2.5, 3) \\ a_1 = x_0 = 2.5 \\ b_1 = b_0 = 3 \end{array} \right.$$

$$x_1 = \frac{2.5 + 3}{2} = 2.75 \quad \left\{ \begin{array}{l} f(2.5) < 0 \\ f(3) > 0 \\ f(2.75) = 0.2082 > 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \xi \in (2.5, 2.75) \\ a_2 = a_1 = 2.5 \\ b_2 = x_1 = 2.75 \end{array} \right.$$

⋮
⋮
⋮

ALGORITMO 1

Seja $f(x)$ contínua em $[a, b]$ e tal que $f(a)f(b) < 0$.

- 1) Dados iniciais:
 - a) intervalo inicial $[a, b]$
 - b) precisão ε
- 2) Se $(b - a) < \varepsilon$, então escolha para \bar{x} qualquer $x \in [a, b]$. FIM.
- 3) $k = 1$
- 4) $M = f(a)$
- 5) $x = \frac{a + b}{2}$
- 6) Se $Mf(x) > 0$, faça $a = x$. Vá para o passo 8.
- 7) $b = x$
- 8) Se $(b - a) < \varepsilon$, escolha para \bar{x} qualquer $x \in [a, b]$. FIM.
- 9) $k = k + 1$. Volte para o passo 5.

Terminado o processo, teremos um intervalo $[a, b]$ que contém a raiz (e tal que $(b - a) < \varepsilon$) e uma aproximação \bar{x} para a raiz exata.

Exemplo 4

$f(x) = x^3 - 9x + 3$		$I = [0, 1]$	$\epsilon = 10^{-3}$
Iteração	x	$f(x)$	$b - a$
1	.5	-1.375	.5
2	.25	.765625	.25
3	.375	-.322265625	.125
4	.3125	.218017578	.0625
5	.34375	-.0531311035	.03125
6	.328125	.0822029114	.015625
7	.3359375	.0144743919	7.8125×10^{-3}
8	.33984375	-.0193439126	3.90625×10^{-3}
9	.337890625	$-2.43862718 \times 10^{-3}$	1.953125×10^{-3}
10	.336914063	$6.01691846 \times 10^{-3}$	9.765625×10^{-4}

Então $\bar{x} = .337402344$ em dez iterações. Observe que neste exemplo escolhemos $\bar{x} = \frac{a + b}{2}$.

ESTUDO DA CONVERGÊNCIA

É bastante intuitivo perceber que se $f(x)$ é contínua no intervalo $[a, b]$ e $f(a)f(b) < 0$, o método da bissecção vai gerar uma seqüência $\{x_k\}$ que converge para a raiz.

No entanto, a prova analítica da convergência requer algumas considerações. Suponhamos que $[a_0, b_0]$ seja o intervalo inicial e que a raiz ξ seja única no interior desse intervalo. O método da bissecção gera três seqüências:

$\{a_k\}$: não-decrescente e limitada superiormente por b_0 ; então existe $r \in \mathbb{R}$ tal que

$$\lim_{k \rightarrow \infty} a_k = r$$

$\{b_k\}$: não-crescente e limitada inferiormente por a_0 , então existe $s \in \mathbb{R}$ tal que

$$\lim_{k \rightarrow \infty} b_k = s$$

$\{x_k\}$: por construção ($x_k = \frac{a_k + b_k}{2}$), temos $a_k < x_k < b_k, \forall k$.

A amplitude de cada intervalo gerado é a metade da amplitude do intervalo anterior.

$$\text{Assim, } \forall k: b_k - a_k = \frac{b_0 - a_0}{2^k}$$

$$\text{Então } \lim_{k \rightarrow \infty} (b_k - a_k) = \lim_{k \rightarrow \infty} \frac{(b_0 - a_0)}{2^k} = 0.$$

Como $\{a_k\}$ e $\{b_k\}$ são convergentes,

$$\lim_{k \rightarrow \infty} b_k - \lim_{k \rightarrow \infty} a_k = 0 \Rightarrow \lim_{k \rightarrow \infty} b_k = \lim_{k \rightarrow \infty} a_k. \text{ Então } r = s.$$

Seja $\ell = r = s$ o limite das duas seqüências. Dado que para todo k o ponto x_k pertence ao intervalo (a_k, b_k) , o Cálculo Diferencial e Integral nos garante que

$$\lim_{k \rightarrow \infty} x_k = \ell$$

Resta provar que ℓ é o zero da função, ou seja, $f(\ell) = 0$.

Em cada iteração k temos $f(a_k) f(b_k) < 0$. Então

$$\begin{aligned} 0 \geq \lim_{k \rightarrow \infty} f(a_k) f(b_k) &= \lim_{k \rightarrow \infty} f(a_k) \lim_{k \rightarrow \infty} f(b_k) = f(\lim_{k \rightarrow \infty} a_k) f(\lim_{k \rightarrow \infty} b_k) = \\ &= f(r) f(s) = f(\ell) f(\ell) = [f(\ell)]^2 \end{aligned}$$

Assim, $0 \geq [f(\ell)]^2 \geq 0$ donde $f(\ell) = 0$.

Portanto $\lim_{k \rightarrow \infty} x_k = \ell$ e ℓ é zero da função. Das hipóteses iniciais temos que $\ell = \xi$.

Concluimos, pois, que o método da bissecção gera uma seqüência convergente sempre que f for contínua em $[a, b]$ com $f(a)f(b) < 0$.

Ao leitor interessado nos resultados sobre convergência de seqüências de reais utilizados nesta demonstração recomendamos a referência [11].

ESTIMATIVA DO NÚMERO DE ITERAÇÕES

Dada uma precisão ϵ e um intervalo inicial $[a, b]$, é possível saber, *a priori*, quantas iterações serão efetuadas pelo método da bissecção até que se obtenha $b - a < \epsilon$, usando o Algoritmo 1.

Vimos que

$$b_k - a_k = \frac{b_{k-1} - a_{k-1}}{2} = \frac{b_0 - a_0}{2^k}$$

Deve-se obter o valor de k tal que $b_k - a_k < \epsilon$, ou seja,

$$\frac{b_0 - a_0}{2^k} < \epsilon \Rightarrow 2^k > \frac{b_0 - a_0}{\epsilon} \Rightarrow k \log(2) > \log(b_0 - a_0) - \log(\epsilon) \Rightarrow$$

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

Portanto se k satisfaz a relação acima, ao final da iteração k teremos o intervalo $[a, b]$ que contém a raiz ξ , tal que $\forall x \in [a, b] \Rightarrow |x - \xi| \leq b - a < \epsilon$.

Por exemplo, se desejarmos encontrar ξ , o zero da função $f(x) = x \log(x) - 1$ que está no intervalo $[2, 3]$ com precisão $\epsilon = 10^{-2}$, quantas iterações, no mínimo, devemos efetuar?

$$k > \frac{\log(3 - 2) - \log(10^{-2})}{\log(2)} = \frac{\log(1) + 2 \log(10)}{\log(2)} = \frac{2}{0.3010} \approx 6.64 \Rightarrow k = 7$$

OBSERVAÇÕES FINAIS

- conforme demonstramos, satisfeitas as hipóteses de continuidade de $f(x)$ em $[a, b]$ e de troca de sinal em a e b , o método da bissecção gera uma seqüência convergente, ou seja, é sempre possível obter um intervalo que contém a raiz da equação em estudo, sendo que o comprimento deste intervalo final satisfaz a precisão requerida;
- as iterações não envolvem cálculos laboriosos;
- a convergência é muito lenta, pois se o intervalo inicial é tal que $b_0 - a_0 \gg \varepsilon$ e se ε for muito pequeno, o número de iterações tende a ser muito grande, como por exemplo:

$$\left. \begin{array}{l} b_0 - a_0 = 3 \\ \varepsilon = 10^{-7} \end{array} \right\} \Rightarrow k \geq 24.8 \Rightarrow k = 25.$$

O Algoritmo 1 pode incluir também o teste de parada com o módulo da função e o do número máximo de iterações.

II. MÉTODO DA POSIÇÃO FALSA

Seja $f(x)$ contínua no intervalo $[a, b]$ e tal que $f(a)f(b) < 0$.

Supor que o intervalo (a, b) contenha uma única raiz da equação $f(x) = 0$.

Podemos esperar conseguir a raiz aproximada \bar{x} usando as informações sobre os valores de $f(x)$ disponíveis a cada iteração.

No caso do método da bissecção, x é simplesmente a média aritmética entre a e b :

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

No Exemplo 4, temos $f(x) = x^3 - 9x + 3$, $[a, b] = [0, 1]$ e $f(1) = -5 < 0 < 3 = f(0)$. Como $|f(0)|$ está mais próximo de zero que $|f(1)|$, é provável que a raiz esteja mais próxima de 0 que de 1 (pelo menos isto ocorre quando $f(x)$ é linear em $[a, b]$).

Assim, em vez de tomar a média aritmética entre a e b , o método da posição falsa toma a média aritmética ponderada entre a e b com pesos $|f(b)|$ e $|f(a)|$, respectivamente:

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

visto que $f(a)$ e $f(b)$ têm sinais opostos.

Graficamente, este ponto x é a intersecção entre o eixo \vec{ox} e a reta $r(x)$ que passa por $(a, f(a))$ e $(b, f(b))$:

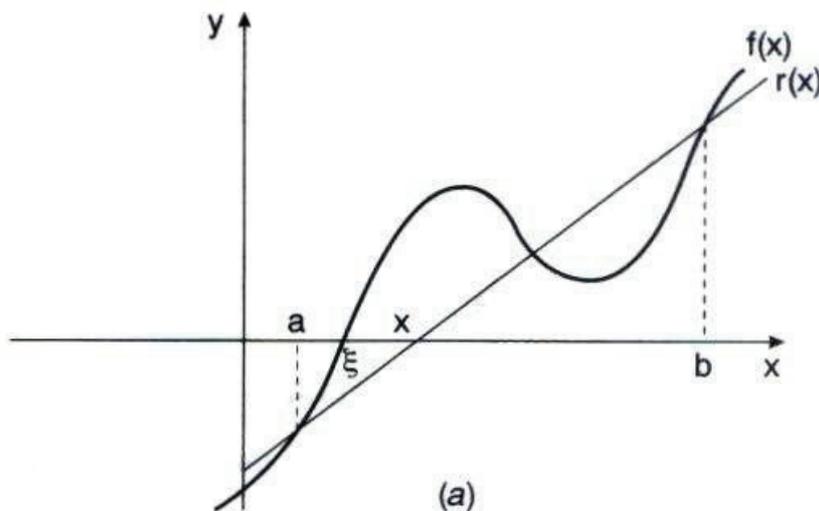


Figura 2.14

E as iterações são feitas assim:

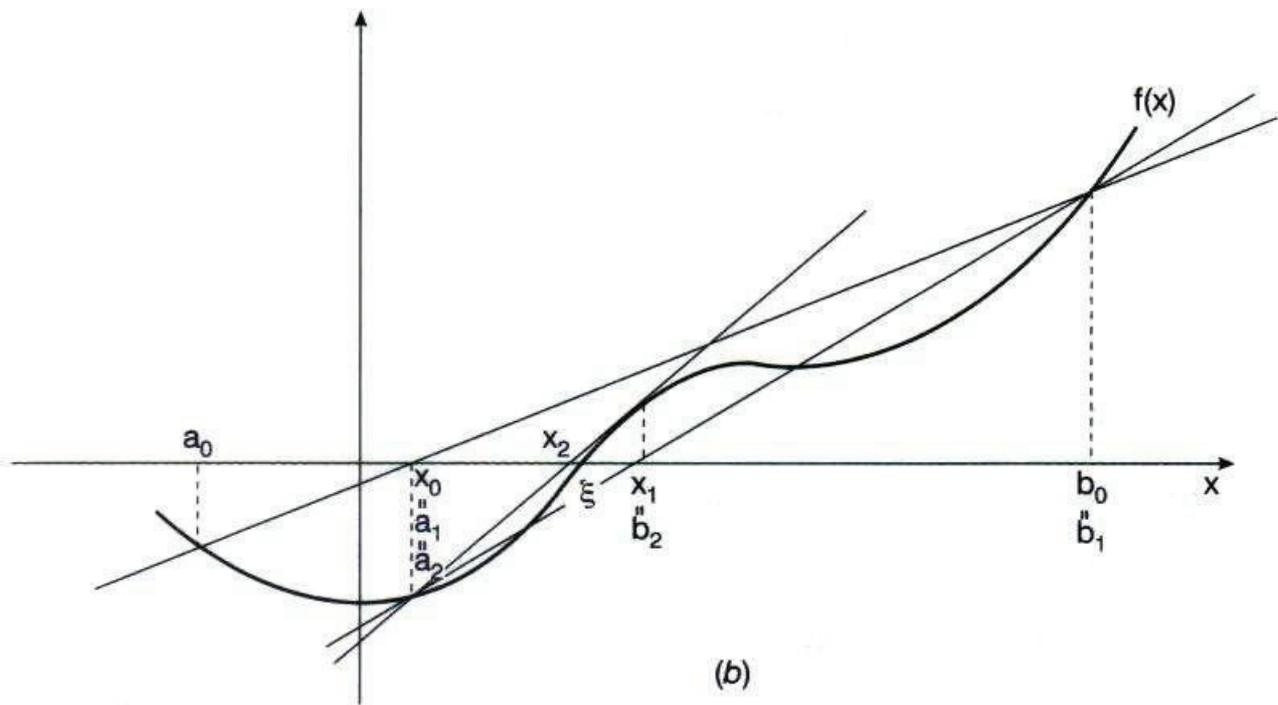


Figura 2.14

Exemplo 5

O método da posição falsa aplicado a $x \log(x) - 1$ em $[a_0, b_0] = [2, 3]$, fica:

$$f(a_0) = -0.3979 < 0$$

$$f(b_0) = 0.4314 > 0$$

$$\Rightarrow x_0 = \frac{af(b) - bf(a)}{f(b) - f(a)} = \frac{2 \times 0.4314 - 3 \times (-0.3979)}{0.4314 - (-0.3979)} = \frac{2.0565}{0.8293} = 2.4798$$

$f(x_0) = -0.0219 < 0$. Como $f(a_0)$ e $f(x_0)$ têm o mesmo sinal,

$$\begin{cases} a_1 = x_0 = 2.4798 & f(a_1) < 0 \\ b_1 = 3 & f(b_1) > 0 \end{cases}$$

$$\Rightarrow x_1 = \frac{2.4798 \times 0.4314 - 3 \times (-0.0219)}{0.4314 - (-0.0219)} = 2.5049 \quad e$$

$f(x_1) = -0.0011$. Analogamente,

$$\begin{cases} a_2 = x_1 = 2.5049 \\ b_2 = b_1 = 3 \end{cases}$$

·
·
·

ALGORITMO 2

Seja $f(x)$ contínua em $[a, b]$ e tal que $f(a)f(b) < 0$.

- 1) Dados iniciais
 - a) intervalo inicial $[a, b]$
 - b) precisões ε_1 e ε_2
- 2) Se $(b - a) < \varepsilon_1$, então escolha para \bar{x} qualquer $x \in [a, b]$. FIM.

$\begin{aligned} &\text{se } f(a) < \varepsilon_2 \\ &\text{ou se } f(b) < \varepsilon_2 \end{aligned}$	$\left. \vphantom{\begin{aligned} &\text{se } f(a) < \varepsilon_2 \\ &\text{ou se } f(b) < \varepsilon_2 \end{aligned}} \right\} \text{escolha a ou b como } \bar{x}. \text{ FIM.}$
-------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
- 3) $k = 1$
- 4) $M = f(a)$
- 5) $x = \frac{af(b) - bf(a)}{f(b) - f(a)}$
- 6) Se $|f(x)| < \varepsilon_2$, escolha $\bar{x} = x$. FIM.
- 7) Se $Mf(x) > 0$, faça $a = x$. Vá para o passo 9.
- 8) $b = x$
- 9) Se $b - a < \varepsilon_1$, então escolha para \bar{x} qualquer $x \in (a, b)$. FIM.
- 10) $k = k + 1$. Volte ao passo 5.

Exemplo 6

$$f(x) = x^3 - 9x + 3 \quad I = [0, 1] \quad \epsilon_1 = \epsilon_2 = 5 \times 10^{-4}$$

Aplicando o método da posição falsa, temos:

Iteração	x	f(x)	b - a
1	.375	-.322265625	1
2	.338624339	$-8.79019964 \times 10^{-3}$.375
3	.337635046	$-2.25883909 \times 10^{-4}$.338624339

E portanto $\bar{x} = 0.337635046$ e $f(\bar{x}) = -2.25 \times 10^{-4}$.

CONVERGÊNCIA

Na referência [30] encontramos demonstrado o seguinte resultado:

“Se $f(x)$ é contínua no intervalo $[a, b]$ com $f(a)f(b) < 0$ então o método da posição falsa gera uma seqüência convergente”.

Embora não façamos aqui a demonstração, observamos que a idéia usada é a mesma aplicada na demonstração da convergência do método da bissecção, ou seja, usando as seqüências $\{a_k\}$, $\{x_k\}$ e $\{b_k\}$. Observamos, ainda, que quando f é derivável duas vezes em $[a, b]$ e $f''(x)$ não muda de sinal nesse intervalo, é bastante intuitivo verificar a convergência graficamente:

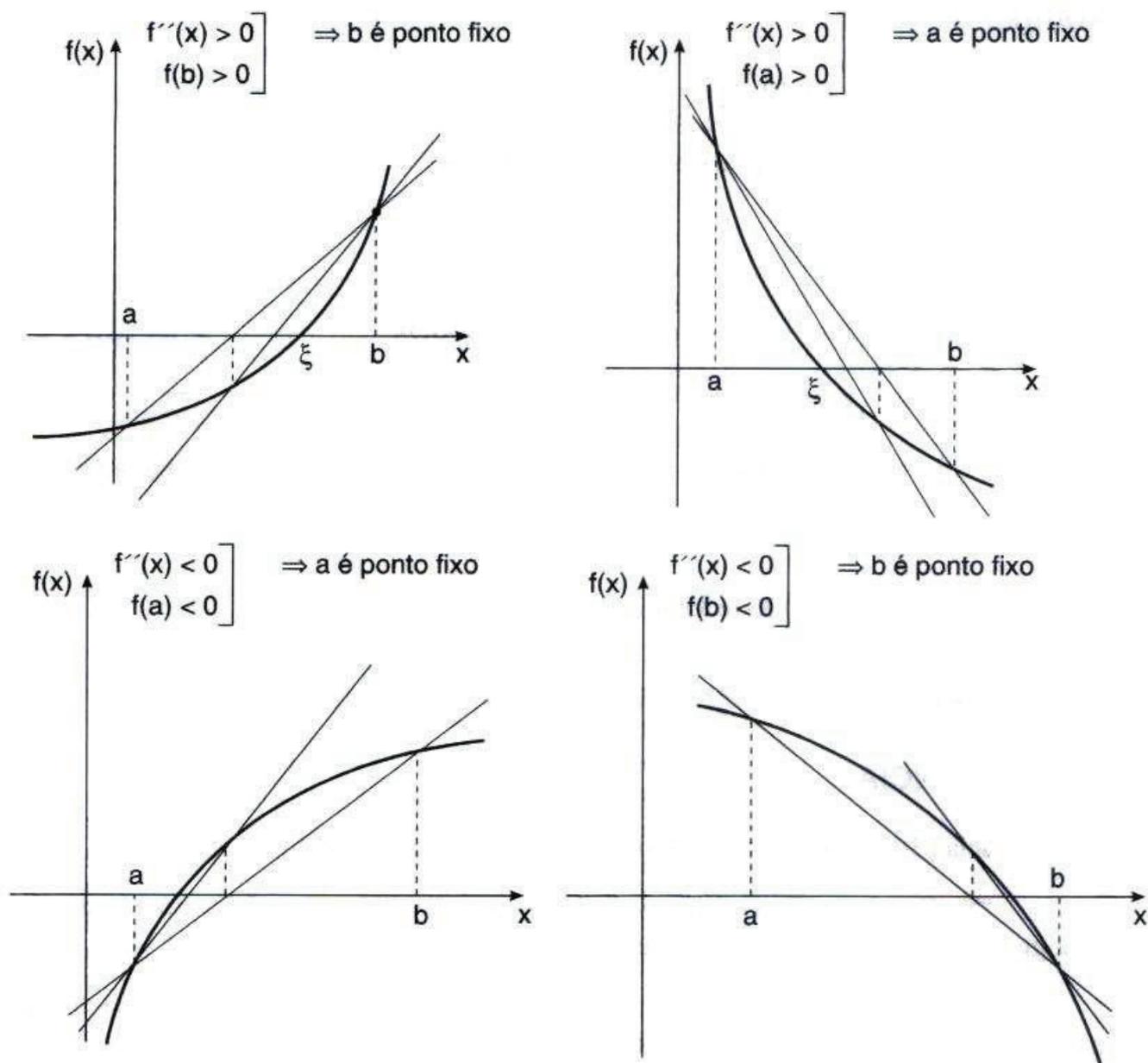


Figura 2.15

Em todos os casos da figura anterior os elementos da seqüência $\{x_k\}$ se encontram na parte do intervalo que fica entre a raiz e o extremo *não*-fixo do intervalo e $\lim_{k \rightarrow \infty} x_k = \xi$.

Analisando ainda estes gráficos, podemos concluir que em geral o método da posição falsa obtém como raiz aproximada um ponto \bar{x} , no qual $|f(\bar{x})| < \varepsilon$, sem que o intervalo $I = [a, b]$ seja pequeno o suficiente. Portanto, se for exigido que os dois critérios de parada sejam satisfeitos simultaneamente, o processo pode exceder um número máximo de iterações.

III. MÉTODO DO PONTO FIXO (MPF)

A importância deste método está mais nos conceitos que são introduzidos em seu estudo que em sua eficiência computacional.

Seja $f(x)$ uma função contínua em $[a, b]$, intervalo que contém uma raiz da equação $f(x) = 0$.

O MPF consiste em transformar esta equação em uma equação equivalente $x = \varphi(x)$ e a partir de uma aproximação inicial x_0 gerar a seqüência $\{x_k\}$ de aproximações para ξ pela relação $x_{k+1} = \varphi(x_k)$, pois a função $\varphi(x)$ é tal que $f(\xi) = 0$ se e somente se $\varphi(\xi) = \xi$. Transformamos assim o problema de encontrar um zero de $f(x)$ no problema de encontrar um ponto fixo de $\varphi(x)$.

Uma função $\varphi(x)$ que satisfaz a condição acima é chamada de *função de iteração* para a equação $f(x) = 0$.

Exemplo 7

Para a equação $x^2 + x - 6 = 0$ temos várias funções de iteração, entre as quais:

a) $\varphi_1(x) = 6 - x^2;$

b) $\varphi_2(x) = \pm \sqrt{6 - x};$

c) $\varphi_3(x) = \frac{6}{x} - 1;$

d) $\varphi_4(x) = \frac{6}{x + 1}.$

A forma geral das funções de iteração $\varphi(x)$ é $\varphi(x) = x + A(x)f(x)$, com a condição que em ξ , ponto fixo de $\varphi(x)$, se tenha $A(\xi) \neq 0$.

Mostremos que $f(\xi) = 0 \Leftrightarrow \varphi(\xi) = \xi$.

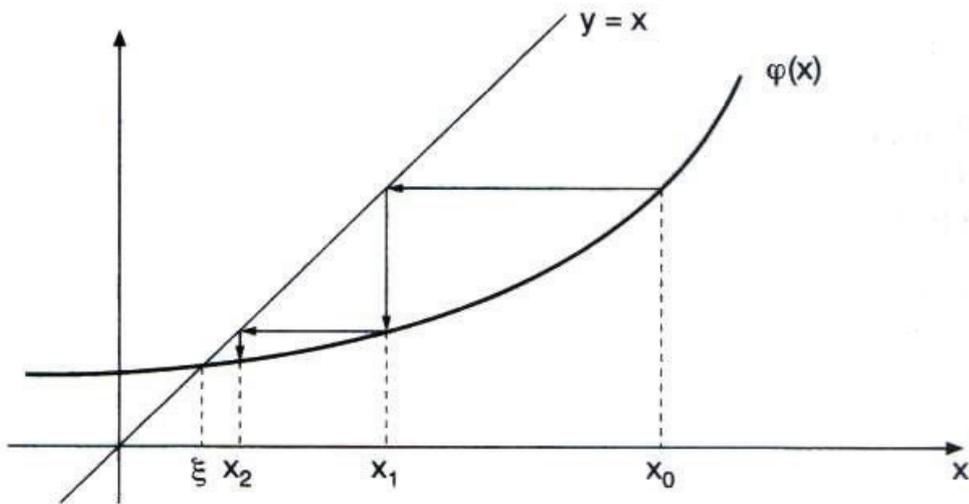
(\Rightarrow) seja ξ tal que $f(\xi) = 0$.

$$\varphi(\xi) = \xi + A(\xi)f(\xi) \Rightarrow \varphi(\xi) = \xi \quad (\text{porque } f(\xi) = 0).$$

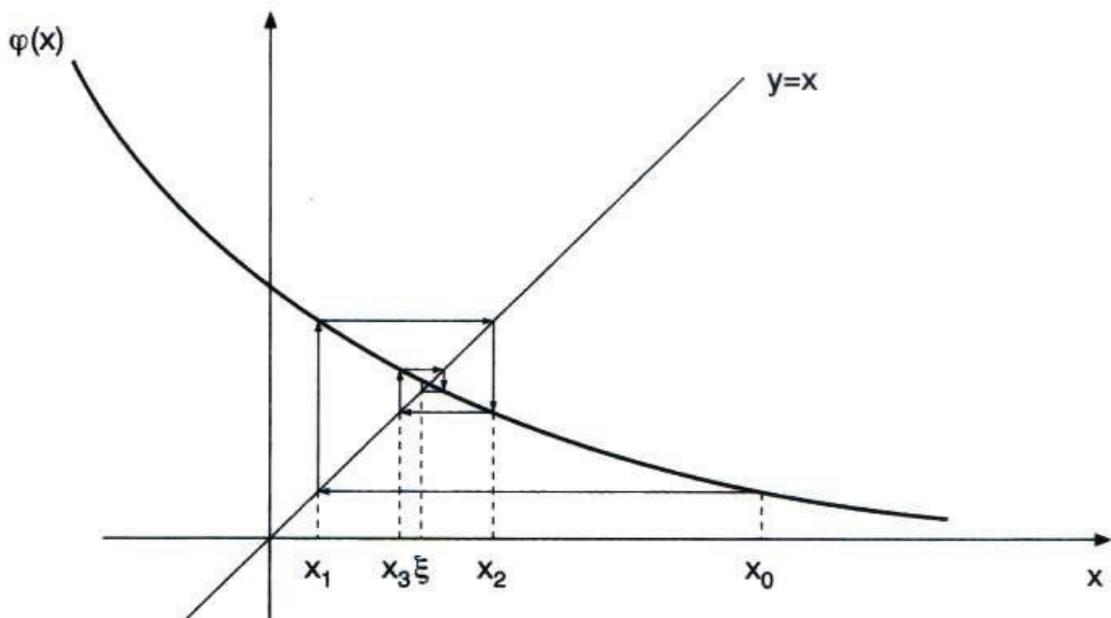
(\Leftarrow) se $\varphi(\xi) = \xi \Rightarrow \xi + A(\xi)f(\xi) = \xi \Rightarrow A(\xi)f(\xi) = 0 \Rightarrow f(\xi) = 0$ (porque $A(\xi) \neq 0$).

Com isto vemos que, dada uma equação $f(x) = 0$, existem infinitas funções de iteração $\varphi(x)$ para a equação $f(x) = 0$.

Graficamente, uma raiz da equação $x = \varphi(x)$ é a abscissa do ponto de intersecção da reta $y = x$ e da curva $y = \varphi(x)$:



(a) $\{x_k\} \rightarrow \xi$ quando $k \rightarrow \infty$



(b) $\{x_k\} \rightarrow \xi$ quando $k \rightarrow \infty$

Figura 2.16

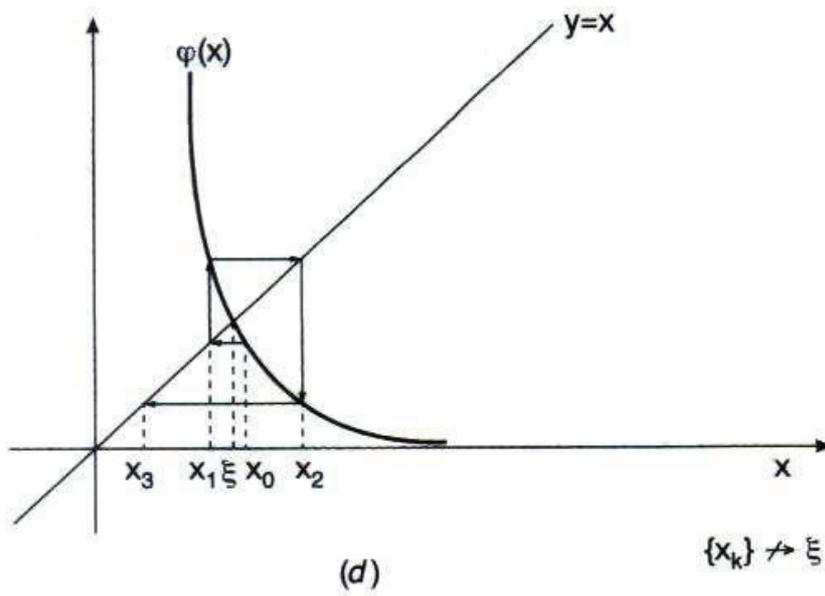
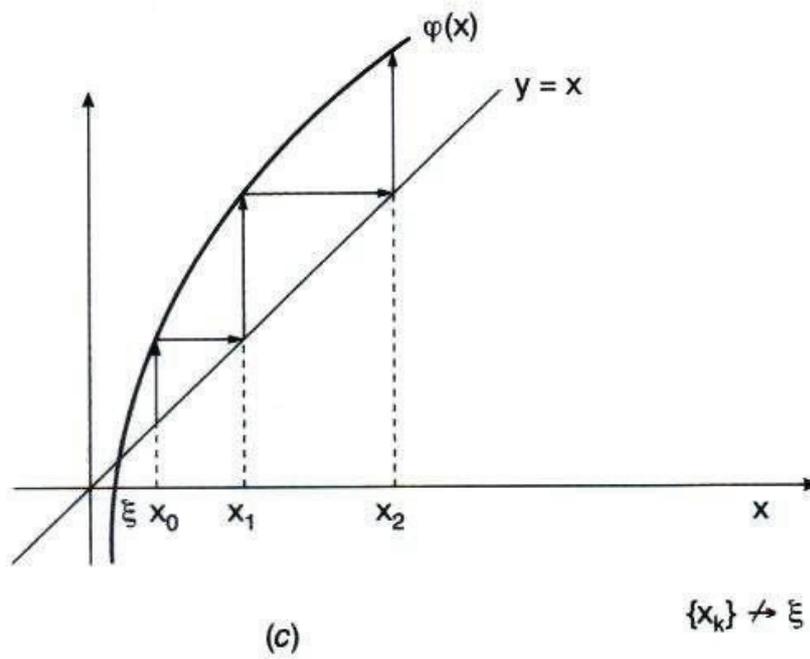


Figura 2.16

Portanto, para certas $\varphi(x)$, o processo pode gerar uma seqüência que diverge de ξ .

ESTUDO DA CONVERGÊNCIA DO MPF

Vimos que, dada uma equação $f(x) = 0$, existe mais de uma função $\varphi(x)$, tal que $f(x) = 0 \Leftrightarrow x = \varphi(x)$.

De acordo com os gráficos da Figura 2.16, não é para qualquer escolha de $\varphi(x)$ que o processo recursivo definido por $x_{k+1} = \varphi(x_k)$ gera uma seqüência que converge para ξ .

Exemplo 8

Embora não seja preciso usar método numérico para se encontrar as duas raízes reais $\xi_1 = -3$ e $\xi_2 = 2$ da equação $x^2 + x - 6 = 0$, vamos trabalhar com duas das funções de iteração dadas no Exemplo 7 para demonstrar numérica e graficamente a convergência ou não do processo iterativo.

Consideremos primeiramente a raiz $\xi_2 = 2$ e $\varphi_1(x) = 6 - x^2$. Tomando $x_0 = 1.5$ temos $\varphi(x) = \varphi_1(x)$ e

$$\begin{aligned}x_1 &= \varphi(x_0) = 6 - 1.5^2 = 3.75 \\x_2 &= \varphi(x_1) = 6 - (3.75)^2 = -8.0625 \\x_3 &= \varphi(x_2) = 6 - (-8.0625)^2 = -59.003906 \\x_4 &= \varphi(x_3) = -(-59.003906)^2 + 6 = -3475.4609 \\&\vdots \\&\vdots \\&\vdots\end{aligned}$$

e podemos ver que $\{x_k\}$ não está convergindo para $\xi_2 = 2$.

GRAFICAMENTE

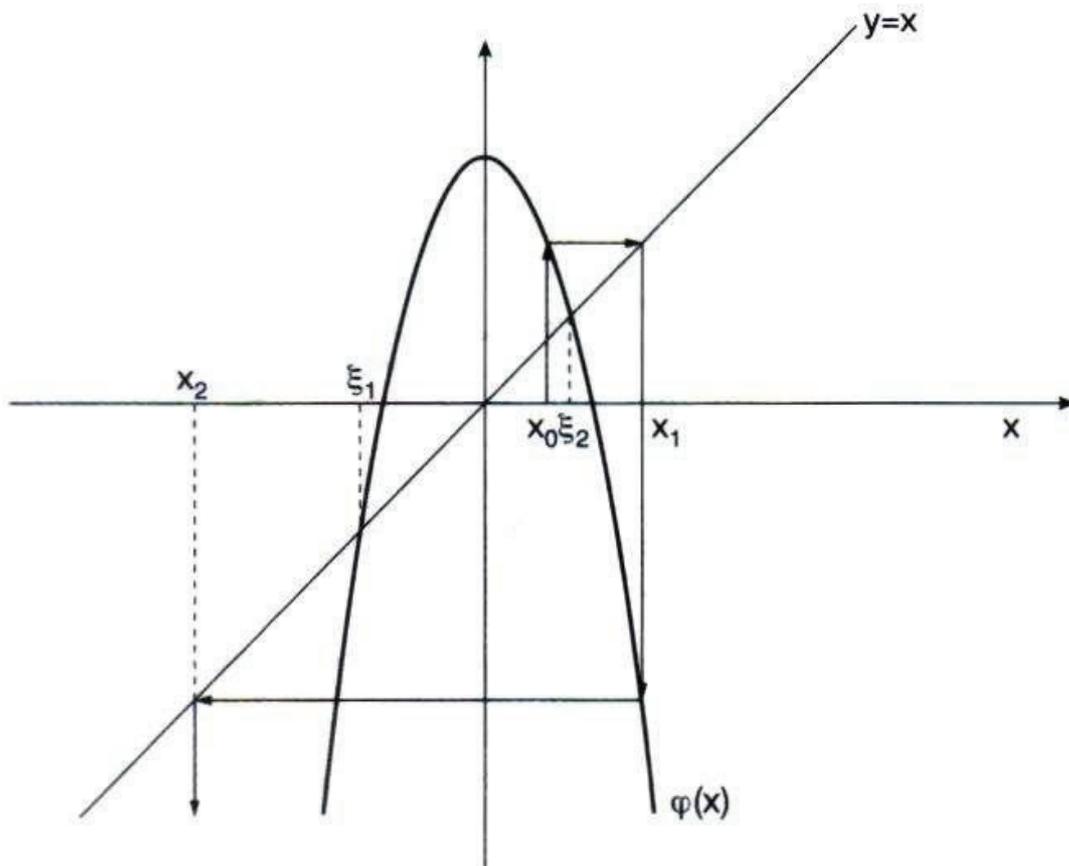


Figura 2.17

Seja agora $\xi_2 = 2$, $\varphi_2(x) = \sqrt{6 - x}$ e novamente $x_0 = 1.5$. Temos, assim, $\varphi(x) = \varphi_2(x)$ e

$$x_1 = \varphi(x_0) = \sqrt{6 - 1.5} = 2.12132$$

$$x_2 = \varphi(x_1) = 1.96944$$

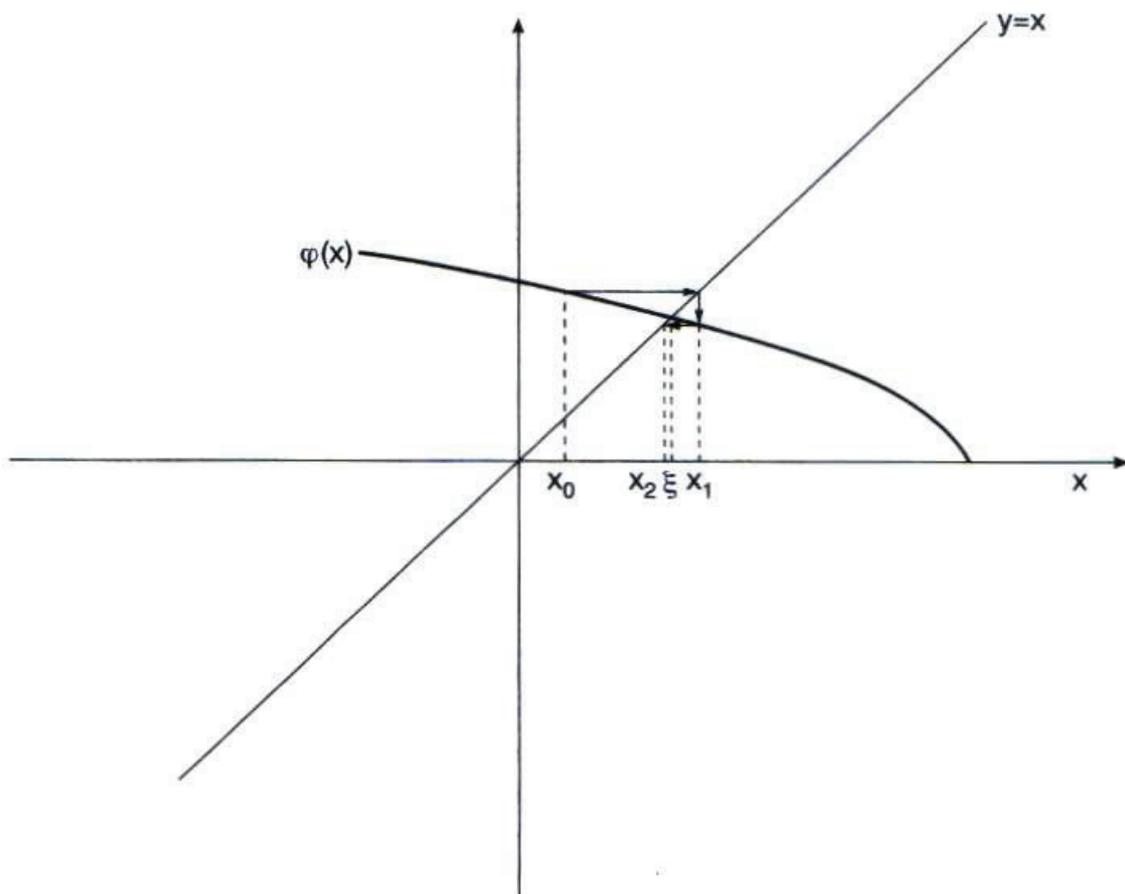
$$x_3 = \varphi(x_2) = 2.00763$$

$$x_4 = \varphi(x_3) = 1.99809$$

$$x_5 = \varphi(x_4) = 2.00048$$

·
·
·

e podemos ver que $\{x_k\}$ está convergindo para $\xi_2 = 2$.

GRAFICAMENTE**Figura 2.18**

O teorema a seguir nos fornece condições suficientes para que o processo seja convergente.

TEOREMA 2

Seja ξ uma raiz da equação $f(x) = 0$, isolada num intervalo I centrado em ξ .

Seja $\varphi(x)$ uma função de iteração para a equação $f(x) = 0$.

Se

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

então a seqüência $\{x_k\}$ gerada pelo processo iterativo $x_{k+1} = \varphi(x_k)$ converge para ξ .

DEMONSTRAÇÃO

A demonstração deste teorema é feita em duas partes:

- 1) prova-se que se $x_0 \in I$, então $x_k \in I, \forall k$;
- 2) prova-se que $\lim_{k \rightarrow \infty} x_k = \xi$.

1) ξ é uma raiz exata da equação $f(x) = 0$.

Assim, $f(\xi) = 0 \Leftrightarrow \xi = \varphi(\xi)$ e,

para qualquer k , temos: $x_{k+1} = \varphi(x_k)$

$$\Rightarrow x_{k+1} - \xi = \varphi(x_k) - \varphi(\xi) \quad (1)$$

Agora, $\varphi(x)$ é contínua e diferenciável em I , então, pelo Teorema do Valor Médio, se $x_k \in I$, existe c_k entre x_k e ξ tal que

$$\varphi'(c_k)(x_k - \xi) = \varphi(x_k) - \varphi(\xi).$$

Portanto, temos

$$x_{k+1} - \xi = \varphi(x_k) - \varphi(\xi) = \varphi'(c_k)(x_k - \xi), \forall k.$$

$$\text{Assim, } x_{k+1} - \xi = \varphi'(c_k)(x_k - \xi) \quad (2)$$

Então, $\forall k$,

$$|x_{k+1} - \xi| = \underbrace{|\varphi'(c_k)|}_{< 1} |x_k - \xi| < |x_k - \xi|$$

ou seja, a distância entre x_{k+1} e ξ é estritamente menor que a distância entre x_k e ξ e, como I está centrado em ξ , temos que se $x_k \in I$, então $x_{k+1} \in I$.

Por hipótese, $x_0 \in I$, então $x_k \in I$, $\forall k$.

2) Provar que $\lim_{k \rightarrow \infty} x_k = \xi$.

De (1), segue que:

$$|x_1 - \xi| = |\varphi(x_0) - \varphi(\xi)| = \underbrace{|\varphi'(c_0)|}_{\leq M} |x_0 - \xi| \leq M |x_0 - \xi| \quad (c_0 \text{ está entre } x_0 \text{ e } \xi)$$

$$|x_2 - \xi| = |\varphi(x_1) - \varphi(\xi)| = \underbrace{|\varphi'(c_1)|}_{\leq M} |x_1 - \xi| \leq M |x_1 - \xi| \leq M^2 |x_0 - \xi| \quad (c_1 \text{ está entre } x_1 \text{ e } \xi)$$

⋮
⋮
⋮

$$|x_k - \xi| = |\varphi(x_{k-1}) - \varphi(\xi)| = \underbrace{|\varphi'(c_{k-1})|}_{\leq M} |x_{k-1} - \xi| \leq M |x_{k-1} - \xi| \leq \dots \leq M^k |x_0 - \xi| \quad (c_k \text{ está entre } x_k \text{ e } \xi)$$

Então, $0 \leq \lim_{k \rightarrow \infty} |x_k - \xi| \leq \lim_{k \rightarrow \infty} M^k |x_0 - \xi| = 0$ pois $0 < M < 1$.

Assim, $\lim_{k \rightarrow \infty} |x_k - \xi| = 0 \Rightarrow \lim_{k \rightarrow \infty} x_k = \xi$.

Exemplo 9

No Exemplo 8, verificamos que $\varphi_1(x)$ gera uma seqüência divergente de $\xi_2 = 2$ enquanto $\varphi_2(x)$ gera uma seqüência convergente para esta raiz.

A seguir, analisaremos as condições do Teorema 2 para estas funções:

a) $\varphi_1(x) = 6 - x^2$ e $\varphi_1'(x) = -2x$

$\varphi_1(x)$ e $\varphi_1'(x)$ são contínuas em \mathbb{R} .

$$|\varphi_1'(x)| < 1 \Leftrightarrow |2x| < 1 \Leftrightarrow -\frac{1}{2} < x < \frac{1}{2}. \text{ Então, não existe um intervalo } I \text{ centrado}$$

em $\xi_2 = 2$, tal que $|\varphi_1'(x)| < 1, \forall x \in I$. Portanto, $\varphi_1(x)$ não satisfaz a condição (ii) do Teorema 2 com relação a $\xi_2 = 2$. Esta é a justificativa teórica da divergência da seqüência $\{x_k\}$ gerada por $\varphi_1(x)$ para $x_0 = 1.5$.

b) $\varphi_2(x) = \sqrt{6 - x}$ e $\varphi_2'(x) = \frac{-1}{2\sqrt{6 - x}}$

$$\varphi_2(x) \text{ é contínua em } S = \{x \in \mathbb{R} \mid x \leq 6\} \tag{3}$$

$$\varphi_2'(x) \text{ é contínua em } S' = \{x \in \mathbb{R} \mid x < 6\} \tag{4}$$

$$|\varphi_2'(x)| < 1 \Leftrightarrow \left| \frac{1}{2\sqrt{6 - x}} \right| < 1 \Leftrightarrow x < 5.75$$

De (3) e (4) temos que é possível obter um intervalo I centrado em $\xi_2 = 2$ tal que as condições do Teorema 2 sejam satisfeitas.

Exemplo 10

Analisaremos aqui a função $\varphi_3(x) = \frac{6}{x} - 1$ e a convergência da seqüência $\{x_k\}$ para $\xi_1 = -3$; usando $x_0 = -2.5$:

$$\varphi_3'(x) = \frac{-6}{x^2} < 0, \quad \forall x \in \mathbb{R}, \quad x \neq 0$$

$$|\varphi_3'(x)| = \left| \frac{-6}{x^2} \right| = \frac{6}{x^2} \quad \forall x \in \mathbb{R}, \quad x \neq 0$$

$$|\varphi_3'(x)| < 1 \Leftrightarrow \frac{-6}{x^2} < 1 \Leftrightarrow x^2 > 6 \Leftrightarrow x < -\sqrt{6} \text{ ou } x > \sqrt{6}$$

Assim, como o objetivo é obter a raiz negativa, temos que

I_1 tal que $|\varphi'(x)| < 1, \forall x \in I_1$, será: $I_1 = (-\infty; \sqrt{6})$.

$$(\sqrt{6} \approx 2.4494897)$$

Podemos, pois, trabalhar no intervalo $I = [-3.5, -2.5]$ que o processo convergirá, visto que $I \subset I_1$ está centrado na raiz $\xi_1 = -3$.

Tomando $x_0 = -2.5$, temos:

$$x_1 = -3.4$$

$$x_2 = -2.764706$$

$$x_3 = -3.170213$$

$$x_4 = -2.892617$$

.

.

.

Como a raiz $\xi_1 = -3$ é conhecida, é possível escolher um intervalo I centrado em ξ_1 , tal que em I as condições do teorema são satisfeitas. Contudo, ao se aplicar o MPF na resolução de uma equação $f(x) = 0$, escolhe-se I “aproximadamente” centrado em ξ . Quanto mais preciso for o processo de isolamento de ξ , maior exatidão será obtida na escolha de I .

CRITÉRIOS DE PARADA

No algoritmo do método do ponto fixo, escolhe-se x_k como raiz aproximada de ξ se $|x_k - x_{k-1}| = |\varphi(x_{k-1}) - x_{k-1}| < \varepsilon$ ou se $|f(x_k)| < \varepsilon$.

Devemos observar que $|x_k - x_{k-1}| < \varepsilon$ não implica necessariamente que $|x_k - \xi| < \varepsilon$ conforme mostra a Figura 2.19:

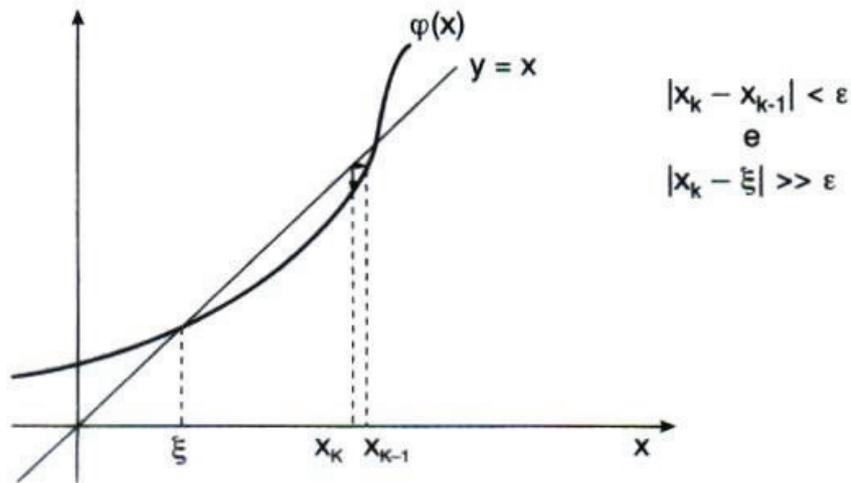


Figura 2.19

Contudo, se $\varphi'(x) < 0$ em I (intervalo centrado em ξ), a seqüência $\{x_k\}$ será oscilante em torno de ξ e, neste caso, se $|x_k - x_{k-1}| < \varepsilon \Rightarrow |x_k - \xi| < \varepsilon$, pois $|x_k - \xi| < |x_k - x_{k-1}|$.

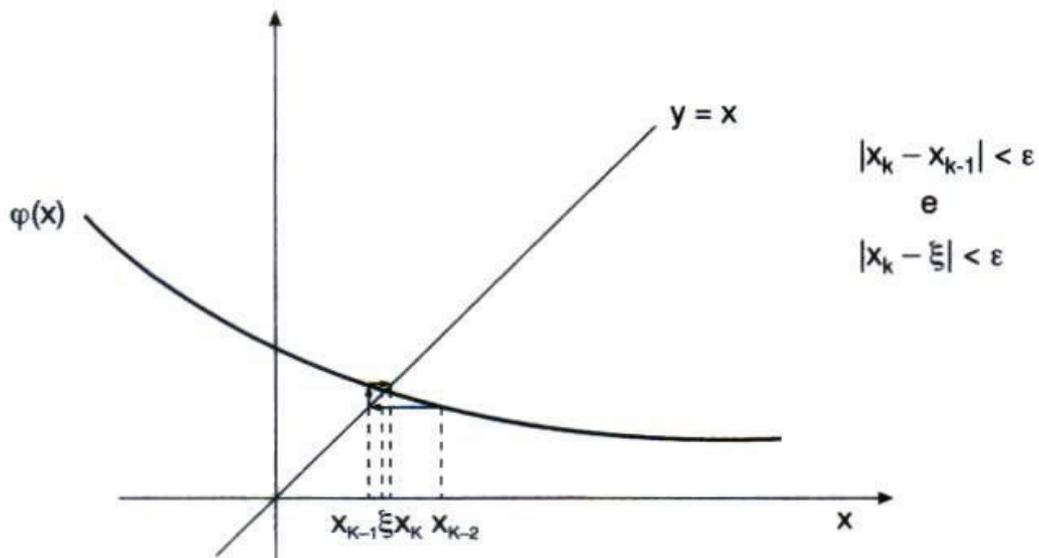


Figura 2.20

ALGORITMO 3

Considere a equação $f(x) = 0$ e a equação equivalente $x = \varphi(x)$.

Supor que as hipóteses do Teorema 2 estão satisfeitas.

1) Dados iniciais:

a) x_0 : aproximação inicial;

b) ε_1 e ε_2 : precisões.

2) Se $|f(x_0)| < \varepsilon_1$, faça $\bar{x} = x_0$. FIM.

3) $k = 1$

4) $x_1 = \varphi(x_0)$

5) Se $|f(x_1)| < \varepsilon_1$
ou se $|x_1 - x_0| < \varepsilon_2$] então faça $\bar{x} = x_1$. FIM.

6) $x_0 = x_1$

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

Exemplo 11

$$f(x) = x^3 - 9x + 3; \quad \varphi(x) = \frac{x^3}{9} + \frac{1}{3}; \quad x_0 = 0.5; \quad \varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}; \quad \xi \in (0,1)$$

Iteração	x	f(x)
1	.3472222	$-0.8313799 \times 10^{-1}$
2	.3379847	$-0.3253222 \times 10^{-2}$
3	.3376233	$-0.1239777 \times 10^{-3}$

assim, $\bar{x} = 0.3376233$ e $f(\bar{x}) = -0.12 \times 10^{-3}$.

Deixamos como exercício a verificação de que $\varphi(x) = \frac{x^3}{9} + \frac{1}{3}$ satisfaz as hipóteses do Teorema 2 considerando a raiz de $f(x) = 0$ que se encontra no intervalo $(0,1)$.

ORDEM DE CONVERGÊNCIA DO MÉTODO DO PONTO FIXO

Definição: “Seja $\{x_k\}$ uma seqüência que converge para um número ξ e seja $e_k = x_k - \xi$ o erro na iteração k .

Se existir um número $p > 1$ e uma constante $C > 0$, tais que

$$\lim_{k \rightarrow \infty} \frac{|e_{k+1}|}{|e_k|^p} = C \quad (5)$$

então p é chamada de *ordem de convergência* da seqüência $\{x_k\}$ e C é a *constante assintótica de erro*.

Se $\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k} = C$, $0 \leq |C| < 1$, então a convergência é pelo menos linear.”

Uma vez obtida a ordem de convergência p de um método iterativo, ela nos dá uma informação sobre a rapidez de convergência do processo, pois de (5) podemos escrever a seguinte relação:

$$|e_{k+1}| \approx C |e_k|^p \text{ para } k \rightarrow \infty.$$

Considerando que a seqüência $\{x_k\}$ é convergente, temos que $e_k \rightarrow 0$ quando $k \rightarrow \infty$, portanto quanto maior for p , mais próximo de zero estará o valor $C |e_k|^p$ (independentemente do valor de C), o que implica uma convergência mais rápida da seqüência $\{x_k\}$. Assim, se dois processos iterativos geram seqüências $\{x_k^1\}$ e $\{x_k^2\}$, ambas convergentes para ξ , com ordem de convergência p_1 e p_2 , respectivamente, e se $p_1 > p_2 \geq 1$, o processo que gera a seqüência $\{x_k^1\}$ converge mais rapidamente que o outro.

A seguir, provaremos que o MPF, em geral, tem convergência apenas linear. Da demonstração do Teorema 2 temos a relação:

$$x_{k+1} - \xi = \varphi(x_k) - \varphi(\xi) = \varphi'(c_k) (x_k - \xi) \text{ com } c_k \text{ entre } x_k \text{ e } \xi$$

$$\Rightarrow \frac{x_{k+1} - \xi}{x_k - \xi} = \varphi'(c_k).$$

Tomando o limite quando $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} \frac{x_{k+1} - \xi}{x_k - \xi} = \lim_{k \rightarrow \infty} \varphi'(c_k) = \varphi'(\lim_{k \rightarrow \infty} (c_k)) = \varphi'(\xi).$$

Portanto, $\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k} = \varphi'(\xi) = C$ e $|C| < 1$ pois $\varphi'(x)$ satisfaz as hipóteses do

Teorema 2.

A relação acima afirma que para grandes valores de k o erro em qualquer iteração é proporcional ao erro na iteração anterior, sendo que o fator de proporcionalidade é $\varphi'(\xi)$. Observamos que a convergência será mais rápida quanto menor for $|\varphi'(\xi)|$.

IV. MÉTODO DE NEWTON-RAPHSON

No estudo do método do ponto fixo, vimos que:

- i) uma das condições de convergência é que $|\varphi'(x)| \leq M < 1$, $\forall x \in I$, onde I é um intervalo centrado na raiz;
- ii) a convergência do método será mais rápida quanto menor for $|\varphi'(\xi)|$.

O que o método de Newton faz, na tentativa de garantir e acelerar a convergência do MPF, é escolher para função de iteração a função $\varphi(x)$ tal que $\varphi'(\xi) = 0$.

Então, dada a equação $f(x) = 0$ e partindo da forma geral para $\varphi(x)$, queremos obter a função $A(x)$ tal que $\varphi'(\xi) = 0$.

$$\varphi(x) = x + A(x)f(x) \Rightarrow$$

$$\Rightarrow \varphi'(x) = 1 + A'(x)f(x) + A(x)f'(x)$$

$$\Rightarrow \varphi'(\xi) = 1 + A'(\xi)f(\xi) + A(\xi)f'(\xi) \Rightarrow \varphi'(\xi) = 1 + A(\xi)f'(\xi).$$

Assim, $\varphi'(\xi) = 0 \Leftrightarrow 1 + A(\xi)f'(\xi) = 0 \Rightarrow A(\xi) = \frac{-1}{f'(\xi)}$, donde tomamos $A(x) = \frac{-1}{f'(x)}$.

Então, dada $f(x)$, a função de iteração $\varphi(x) = x - \frac{f(x)}{f'(x)}$ será tal que $\varphi'(\xi) = 0$, pois como podemos verificar:

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

e, como $f(\xi) = 0$, $\varphi'(\xi) = 0$ (desde que $f'(\xi) \neq 0$).

Assim, escolhido x_0 , a seqüência $\{x_k\}$ será determinada por $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$, $k = 0, 1, 2, \dots$.

MOTIVAÇÃO GEOMÉTRICA

O método de Newton é obtido geometricamente da seguinte forma:

dado o ponto $(x_k, f(x_k))$ traçamos a reta $L_k(x)$ tangente à curva neste ponto:

$$L_k(x) = f(x_k) + f'(x_k)(x - x_k).$$

$L_k(x)$ é um modelo linear que aproxima a função $f(x)$ numa vizinhança de x_k .

Encontrando o zero deste modelo, obtemos:

$$L_k(x) = 0 \Leftrightarrow x = x_k - \frac{f(x_k)}{f'(x_k)}$$

Fazemos então $x_{k+1} = x$.

GRAFICAMENTE

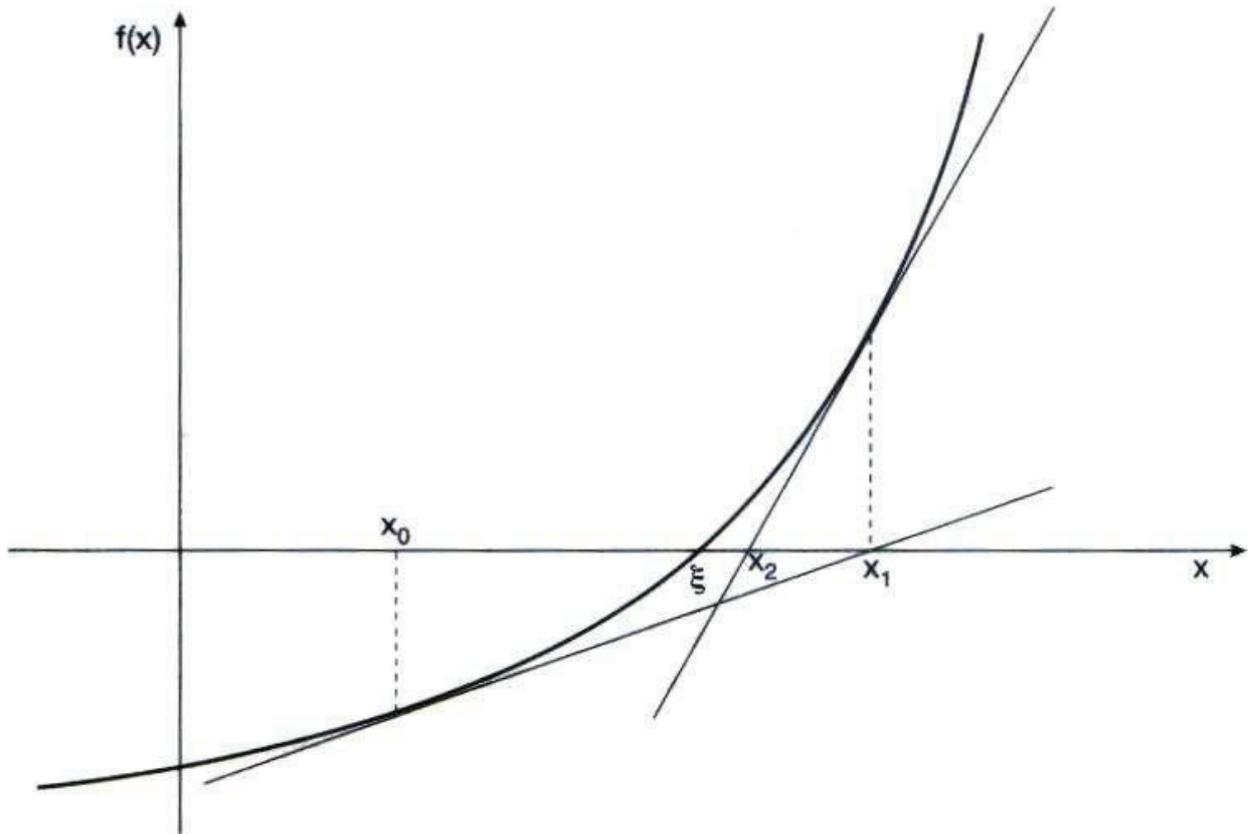


Figura 2.21

Exemplo 12

Consideremos $f(x) = x^2 + x - 6$, $\xi_2 = 2$ e $x_0 = 1.5$

$$\varphi(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^2 + x - 6}{2x + 1}$$

Temos, pois,

$$x_0 = 1.5$$

$$x_1 = \varphi(x_0) = 2.0625$$

$$x_2 = \varphi(x_1) = 2.00076$$

$$x_3 = \varphi(x_2) = 2.00000.$$

Assim, trabalhando com cinco casas decimais, $\bar{x} = x_3 = \xi$. Observamos que no MPF com $\varphi(x) = \sqrt{6 - x}$ (Exemplo 8) obtivemos $x_5 = 2.00048$ com cinco casas decimais.

ESTUDO DA CONVERGÊNCIA DO MÉTODO DE NEWTON

TEOREMA 3

Sejam $f(x)$, $f'(x)$ e $f''(x)$ contínuas num intervalo I que contém a raiz $x = \xi$ de $f(x) = 0$. Supor que $f'(\xi) \neq 0$.

Então, existe um intervalo $\bar{I} \subset I$, contendo a raiz ξ , tal que se $x_0 \in \bar{I}$, a seqüência $\{x_k\}$ gerada pela fórmula recursiva $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$ convergirá para a raiz.

DEMONSTRAÇÃO

Vimos que o método de Newton-Raphson é um MPF com função de iteração $\varphi(x)$ dada por $\varphi(x) = x - \frac{f(x)}{f'(x)}$.

Portanto, para provar a convergência do método, basta verificar que, sob as hipóteses acima, as hipóteses do Teorema 2 estão satisfeitas para $\varphi(x)$.

Ou seja, é preciso provar que existe $\bar{I} \subset I$ centrado em ξ , tal que:

- i) $\varphi(x)$ e $\varphi'(x)$ são contínuas em \bar{I} ;
- ii) $|\varphi'(x)| \leq M < 1, \forall x \in \bar{I}$.

Temos que

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

Por hipótese, $f'(\xi) \neq 0$ e, como $f'(x)$ é contínua em I , é possível obter $I_1 \subset I$ tal que $f'(x) \neq 0, \forall x \in I_1$.

Assim, no intervalo $I_1 \subset I$, tem-se que $f(x)$, $f'(x)$ e $f''(x)$ são contínuas e $f'(x) \neq 0$.

Portanto, $\varphi(x)$ e $\varphi'(x)$ são contínuas em I_1 .

Agora, $\varphi'(x) = \frac{f(x)f''(x)}{[f'(x)]^2}$. Como $\varphi'(x)$ é contínua em I_1 e $\varphi'(\xi) = 0$, é possível

escolher $I_2 \subset I_1$ tal que $|\varphi'(x)| < 1, \forall x \in I_2$ e, ainda mais, I_2 pode ser escolhido de forma que ξ seja seu centro.

Concluindo, conseguimos obter um intervalo $I_2 \subset I$, centrado em ξ , tal que $\varphi(x)$ e $\varphi'(x)$ sejam contínuas em I_2 e $|\varphi'(x)| < 1, \forall x \in I_2$. Assim, $\bar{I} = I_2$.

Portanto, se $x_0 \in \bar{I}$, a seqüência $\{x_k\}$ gerada pelo processo iterativo $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$ converge para a raiz ξ .

Em geral, afirma-se que o método de Newton converge desde que x_0 seja escolhido “suficientemente próximo” da raiz ξ .

A razão desta afirmação está na demonstração acima, onde se verificou que, para pontos suficientemente próximos de ξ , as hipóteses do teorema da convergência do MPF estão satisfeitas.

Exemplo 13

Comprovaremos neste exemplo que uma escolha cuidadosa da aproximação inicial é, em geral, essencial para o bom desempenho do método de Newton.

Consideremos a função $f(x) = x^3 - 9x + 3$ que possui três zeros: $\xi_1 \in I_1 = (-4, -3)$, $\xi_2 \in I_2 = (0, 1)$ e $\xi_3 \in I_3 = (2, 3)$ e seja $x_0 = 1.5$. A seqüência gerada pelo método é

Iteração	x	f(x)
1	-1.6666667	0.1337037×10^2
2	18.3888889	0.6055725×10^4
3	12.3660104	0.1782694×10^4
4	8.4023067	0.5205716×10^3
5	5.83533816	0.1491821×10^3
6	4.23387355	0.4079022×10^2
7	3.32291096	0.9784511×10
8	2.91733893	0.1573032×10
9	2.82219167	0.7837065×10^{-1}
10	2.81692988	0.2342695×10^{-3}

Podemos observar que de início há uma divergência da região onde estão as raízes, mas, a partir de x_7 , os valores aproximam-se cada vez mais de ξ_3 . A causa da divergência inicial é que x_0 está próximo de $\sqrt{3}$ que é um zero de $f'(x)$ e esta aproximação inicial gera $x_1 = -1.66667 \approx -\sqrt{3}$ que é o outro zero de $f'(x)$ pois

$$f'(x) = 3x^2 - 9 \Rightarrow f'(x) = 0 \Leftrightarrow x = \pm\sqrt{3}.$$

ALGORITMO 4

Seja a equação $f(x) = 0$.

Supor que estão satisfeitas as hipóteses do Teorema 3.

1) Dados iniciais:

- a) x_0 : aproximação inicial;
- b) ε_1 e ε_2 : precisões

2) Se $|f(x_0)| < \varepsilon_1$, faça $\bar{x} = x_0$. FIM.

3) $k = 1$

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

5) Se $|f(x_1)| < \varepsilon_1$
ou se $|x_1 - x_0| < \varepsilon_2$ } faça $\bar{x} = x_1$. FIM.

6) $x_0 = x_1$

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

Exemplo 14

$$f(x) = x^3 - 9x + 3; \quad x_0 = 0.5; \quad \varepsilon_1 = \varepsilon_2 = 1 \times 10^{-4}; \quad \xi \in (0,1).$$

Os resultados obtidos ao aplicar o método de Newton são:

Iteração	x	f(x)
0	0.5	-0.1375×10
1	.333333333	0.3703703×10^{-1}
2	.337606838	0.1834054×10^{-4}

Assim, $\bar{x} = 0.337606838$ e $f(\bar{x}) = 1.8 \times 10^{-5}$.

ORDEM DE CONVERGÊNCIA

Inicialmente supomos que o método de Newton gera uma seqüência $\{x_k\}$ que converge para ξ .

Ao observá-lo como um MPF, diríamos que ele tem ordem de convergência linear. Contudo, o fato de sua função de iteração ser tal que $\varphi'(\xi) = 0$ nos levará a demonstrar que a ordem de convergência é quadrática, ou seja, $p = 2$.

Vamos supor que estão satisfeitas aqui todas as hipóteses do Teorema 3.

$$\text{Temos que } x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{k+1} - \xi = x_k - \xi - \frac{f(x_k)}{f'(x_k)} \Rightarrow e_k - \frac{f(x_k)}{f'(x_k)} = e_{k+1}.$$

O desenvolvimento de Taylor de $f(x)$ em torno de x_k nos dá

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f''(c_k)}{2}(x - x_k)^2, \quad c_k \text{ entre } x \text{ e } x_k.$$

$$\text{Assim, } 0 = f(\xi) = f(x_k) - f'(x_k)(x_k - \xi) + \frac{f''(c_k)}{2}(x_k - \xi)^2$$

$$\Rightarrow f(x_k) = f'(x_k)(x_k - \xi) - \frac{f''(c_k)}{2}(x_k - \xi)^2 \quad (+ f'(x_k))$$

$$\Rightarrow \frac{f''(c_k)}{2f'(x_k)} e_k^2 = -\frac{f(x_k)}{f'(x_k)} + e_k = e_{k+1}$$

$$\Rightarrow \frac{e_{k+1}}{e_k^2} = \frac{1}{2} \frac{f''(c_k)}{f'(x_k)}$$

$$\text{Assim, } \lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k^2} = \frac{1}{2} \lim_{k \rightarrow \infty} \frac{f''(c_k)}{f'(x_k)} =$$

$$= \frac{1}{2} \frac{f''[\lim_{k \rightarrow \infty} (c_k)]}{f'[\lim_{k \rightarrow \infty} (x_k)]} = \frac{1}{2} \frac{f''(\xi)}{f'(\xi)} = \frac{1}{2} \varphi''(\xi) = C$$

Portanto, o método de Newton tem convergência quadrática.

Exemplo 15

Seja obter a raiz quadrada de um número positivo A , usando o método de Newton. Temos de resolver a equação $f(x) = x^2 - A = 0$. Tomando $A = 7$ e $x_0 = 2$, a seqüência gerada é:

$$x_0 = 2$$

$$x_1 = 2.75$$

$$x_2 = 2.\underline{647727273}$$

$$x_3 = 2.\underline{645752048}$$

$$x_4 = 2.\underline{645751311}$$

$$x_5 = 2.\underline{645751311}.$$

Portanto, trabalhando com nove casas decimais, $\bar{x} = 2.645751311$.

Os dígitos sublinhados são os dígitos decimais corretos de cada valor x_k obtido.

Podemos observar que estes dígitos corretos começam a surgir após x_2 e, a partir dele, a quantidade de dígitos corretos praticamente duplica. A duplicação de dígitos corretos ocorre à medida que os valores x_k se aproximam da raiz exata, e isto se deve ao fato do método de Newton ter convergência quadrática; como esta é uma propriedade assintótica, não se deve esperar a duplicação de dígitos corretos nas iterações iniciais.

V. MÉTODO DA SECANTE

Uma grande desvantagem do método de Newton é a necessidade de se obter $f'(x)$ e calcular seu valor numérico a cada iteração.

Uma forma de se contornar este problema é substituir a derivada $f'(x_k)$ pelo quociente das diferenças:

$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

onde x_k e x_{k-1} são duas aproximações para a raiz.

Neste caso, a função de iteração fica

$$\varphi(x_k) = x_k - \frac{f(x_k)}{\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}} =$$

$$x_k - \frac{f(x_k)}{f(x_k) - f(x_{k-1})} (x_k - x_{k-1})$$

$$\text{Ou ainda, } \varphi(x_k) = \frac{x_{k-1} f(x_k) - x_k f(x_{k-1})}{f(x_k) - f(x_{k-1})}$$

Observamos que são necessárias duas aproximações para se iniciar o método.

INTERPRETAÇÃO GEOMÉTRICA

A partir de duas aproximações x_{k-1} e x_k , o ponto x_{k+1} é obtido como sendo a abscissa do ponto de intersecção do eixo \vec{ox} e da reta secante que passa por $(x_{k-1}, f(x_{k-1}))$ e $(x_k, f(x_k))$:

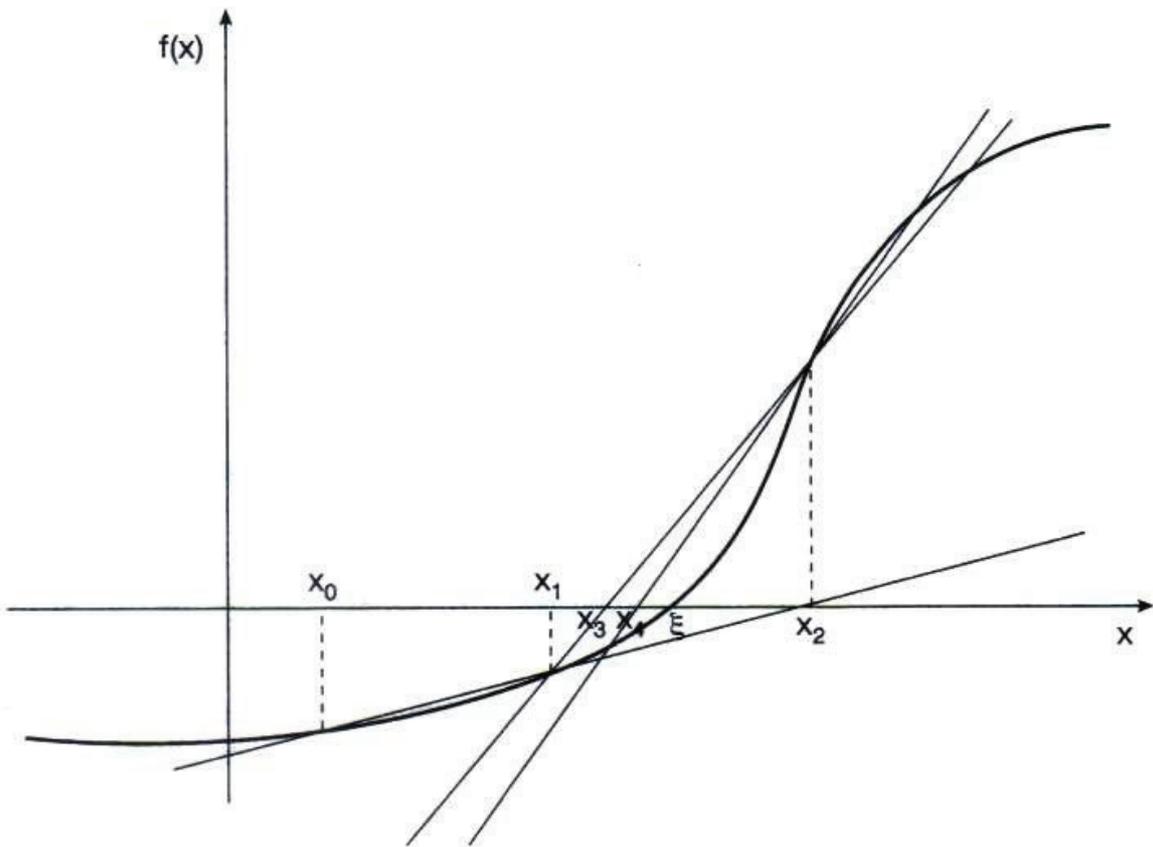


Figura 2.22

Exemplo 16

Consideremos $f(x) = x^2 + x - 6$; $\xi_2 = 2$; $x_0 = 1.5$ e $x_1 = 1.7$. Então,

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_2) - f(x_1)} = \frac{1.5(-1.41) - 1.7(-2.25)}{-1.41 + 2.25} = 2.03571$$

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)} = \frac{1.7(0.17983) - (2.03571)(-1.41)}{0.17983 + 1.41} = 1.99774$$

$$x_4 = \frac{x_2 f(x_3) - x_3 f(x_2)}{f(x_3) - f(x_2)} = \frac{(2.03571)(-0.01131) - (1.99774)(0.17983)}{-0.01131 - 0.17983} =$$

$$\begin{array}{ccc} \cdot & \cdot & = 1.99999 \\ \cdot & \cdot & \\ \cdot & \cdot & \end{array}$$

ALGORITMO 5

Seja a equação $f(x) = 0$.

- 1) Dados iniciais:
 - a) x_0 e x_1 : aproximações iniciais;
 - b) ε_1 e ε_2 : precisões.
- 2) Se $|f(x_0)| < \varepsilon_1$, faça $\bar{x} = x_0$. FIM.
- 3) Se $|f(x_1)| < \varepsilon_1$
ou se $|x_1 - x_0| < \varepsilon_2$] então faça $\bar{x} = x_1$. FIM.
- 4) $k = 1$
- 5) $x_2 = x_1 - \frac{f(x_1)}{f(x_1) - f(x_0)} (x_1 - x_0)$
- 6) Se $|f(x_2)| < \varepsilon_1$
ou se $|x_2 - x_1| < \varepsilon_2$] então faça $\bar{x} = x_2$. FIM.
- 7) $x_0 = x_1$
 $x_1 = x_2$
- 8) $k = k + 1$
Volte ao passo 5.

Exemplo 17

$$f(x) = x^3 - 9x + 3, \quad x_0 = 0, \quad x_1 = 1, \quad \varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$$

Os resultados obtidos ao aplicarmos o método da secante são:

Iteração	x	f(x)
1	.375	-.322265625
2	.331941545	.0491011376
3	.337634621	$-0.2222052 \times 10^{-3}$

Assim, $\bar{x} = 0.337634621$ e $f(\bar{x}) = -2.2 \times 10^{-4}$

COMENTÁRIOS FINAIS

Visto que o método da secante é uma aproximação para o método de Newton, as condições para a convergência do método são praticamente as mesmas; acrescenta-se ainda que o método pode divergir se $f(x_k) \approx f(x_{k-1})$.

A ordem de convergência do método da secante não é quadrática como a do método de Newton, mas também não é apenas linear. Na referência [5] Capítulo 3, § 5, está provado que para o método da secante $p = 1.618...$

2.4 COMPARAÇÃO ENTRE OS MÉTODOS

Finalizando este capítulo realizaremos alguns testes com o objetivo de comparar os vários métodos.

Esta comparação deve levar em conta vários critérios entre os quais: garantias de convergência, rapidez de convergência, esforço computacional.

Observamos que o único dado que os exemplos fornecem para se medir a rapidez de convergência é o número de iterações efetuadas, o que não nos permite tirar conclusões sobre o tempo de execução do programa, pois o tempo gasto na execução de uma iteração varia de método para método.

Conforme constatamos no estudo teórico, os métodos da bissecção e da posição falsa têm convergência garantida desde que a função seja contínua num intervalo $[a, b]$ tal que $f(a)f(b) < 0$. Já o MPF e os métodos de Newton e secante têm condições mais restritivas de convergência. Porém, uma vez que as condições de convergência sejam satisfeitas, os dois últimos são mais rápidos que os três primeiros.

O esforço computacional é medido através do número de operações efetuadas a cada iteração, da complexidade destas operações, do número de decisões lógicas, do número de avaliações de função a cada iteração e do número total de iterações.

Tendo isto em mente, percebe-se que é difícil tirar conclusões gerais sobre a eficiência computacional de um método, pois, por exemplo, o método da bissecção é o que efetua cálculos mais simples por iteração enquanto que o de Newton requer cálculos mais elaborados, porque requer o cálculo da função e de sua derivada a cada iteração. No entanto, o número de iterações efetuadas pela bissecção pode ser muito maior que o número de iterações efetuadas por Newton.

Considerando que o método ideal seria aquele em que a convergência estivesse assegurada, a ordem de convergência fosse alta e os cálculos por iteração fossem simples, o método de Newton é o mais indicado sempre que for fácil verificar as condições de convergência e que o cálculo de $f'(x)$ não seja muito elaborado. Nos casos em que é trabalhoso obter e/ou avaliar $f'(x)$, é aconselhável usar o método da secante, uma vez que este é o método que converge mais rapidamente entre as outras opções.

Outro detalhe importante na escolha é o critério de parada, pois, por exemplo, se o objetivo for reduzir o intervalo que contém a raiz, não se deve usar métodos como o da posição falsa que, apesar de trabalhar com intervalo, pode não atingir a precisão requerida, nem secante, MPF ou Newton que trabalham exclusivamente com aproximações x_k para a raiz exata.

Após estas considerações, podemos concluir que a escolha do método está diretamente relacionada com a equação que se quer resolver, no que diz respeito ao comportamento da função na região da raiz exata, às dificuldades com o cálculo de $f'(x)$, ao critério de parada etc.

Exemplo 18

$$f(x) = e^{-x^2} - \cos(x); \quad \xi \in (1, 2); \quad \varepsilon_1 = \varepsilon_2 = 10^{-4}$$

	Bisseccção	Posição Falsa	MPF $\varphi(x) = \cos(x) - e^{-x^2} + x$	Newton	Secante
Dados Iniciais	[1, 2]	[1, 2]	$x_0 = 1.5$	$x_0 = 1.5$	$x_0 = 1; x_1 = 2$
\bar{x}	1.44741821	1.44735707	1.44752471	1.44741635	1.44741345
$f(\bar{x})$	2.1921×10^{-5}	-3.6387×10^{-5}	7.0258×10^{-5}	1.3205×10^{-6}	-5.2395×10^{-7}
Erro em x	6.1035×10^{-5}	.552885221	1.9319×10^{-4}	1.7072×10^{-3}	1.8553×10^{-4}
Número de Iterações	14	6	6	2	5

Exemplo 19

$$f(x) = x^3 - x - 1; \quad \xi \in (1, 2); \quad \varepsilon_1 = \varepsilon_2 = 10^{-6}$$

	Bisseccção	Posição Falsa	MPF $\varphi(x) = (x + 1)^{1/3}$	Newton	Secante
Dados Iniciais	[1, 2]	[1, 2]	$x_0 = 1$	$x_0 = 0$	$x_0 = 0; x_1 = 0.5$
\bar{x}	0.1324718×10^1	0.1324718×10^1	0.1324717×10^1	0.1324718×10^1	0.1324718×10^1
$f(\bar{x})$	$-0.1847744 \times 10^{-5}$	$-0.7897615 \times 10^{-6}$	$-0.52154406 \times 10^{-6}$	0.1821000×10^{-6}	$-0.8940697 \times 10^{-7}$
Erro em x	0.9536743×10^{-6}	0.6752825	0.3599538×10^{-6}	0.6299186×10^{-6}	0.8998843×10^{-5}
Número de Iterações	20	17	9	21	27

No método de Newton, o valor inicial $x_0 = 0$, além de estar muito distante da raiz $\xi(\approx 1.3)$, gera para x_1 o valor $x_1 = 0.5$ que está próximo de um zero da derivada de $f(x)$; $f'(x) = 3x^2 - 1 \Rightarrow f'(x) = 0 \Leftrightarrow x = \pm \sqrt{3}/3 \approx 0.5773502$. Isto é uma justificativa para o método ter efetuado 21 iterações.

Argumentos semelhantes podem ser usados para justificar as 27 iterações do método da secante.

Exemplo 20

$$f(x) = 4\text{sen}(x) - e^x; \quad \xi \in (0, 1); \quad \varepsilon_1 = \varepsilon_2 = 10^{-5}$$

	Bisseccção	Posição Falsa	MPF $\varphi(x) = x - 2 \text{sen}(x) + 0.5e^x$	Newton	Secante
Dados Iniciais	[0, 1]	[0, 1]	$x_0 = 0.5$	$x_0 = 0.5$	$x_0 = 0; x_1 = 1$
\bar{x}	0.370555878	0.370558828	.370556114	.370558084	.370558098
$f(\bar{x})$	-1.3755×10^{-5}	1.6695×10^{-6}	-4.5191×10^{-6}	-2.7632×10^{-8}	5.8100×10^{-9}
Erro em x	7.6294×10^{-6}	.370562817	1.1528×10^{-4}	$+1.3863 \times 10^{-4}$	5.7404×10^{-6}
Número de Iterações	17	8	5	3	7

Exemplo 21

$$f(x) = x \log(x) - 1; \quad \xi \in (2, 3); \quad \varepsilon_1 = \varepsilon_2 = 10^{-7}$$

	Bisseccção	Posição Falsa	MPF $\varphi(x) = x - 1.3(x \log x - 1)$	Newton	Secante
Dados Iniciais	[2, 3]	[2, 3]	$x_0 = 2.5$	$x_0 = 2.5$	$x_0 = 2.3; x_1 = 2.7$
\bar{x}	2.506184413	2.50618403	2.50618417	2.50618415	2.50618418
$f(\bar{x})$	1.2573×10^{-8}	-9.9419×10^{-8}	2.0489×10^{-8}	4.6566×10^{-10}	2.9337×10^{-8}
Erro em x	5.9605×10^{-8}	.49381442	3.8426×10^{-6}	3.9879×10^{-6}	8.0561×10^{-5}
Número de Iterações	24	5	5	2	3

Exemplo 22

Métodos mais simples como o da bissecção podem ser usados para fornecer uma aproximação inicial para métodos mais elaborados como o de Newton que exigem um bom “chute inicial”.

$$\text{Consideremos } f(x) = x^3 - 3.5x^2 + 4x - 1.5 = (x - 1)^2 (x - 1.5).$$

Como vemos, $\xi_1 = 1$ é raiz dupla de $f(x) = 0$.

Nos testes a seguir, $\varepsilon = 10^{-2}$ para o método da bissecção e $\varepsilon = 10^{-7}$ para o método de Newton.

Nos testes 1, 2 e 3, executamos apenas o método de Newton. No teste 4, usamos o método conjugado bissecção-Newton no qual o valor que o método da bissecção encontra para \bar{x} é tomado como x_0 para o método de Newton.

	Teste 1	Teste 2	Teste 3
x_0	0.5	1.33333	1.33334
\bar{x}	.999778284	.999708915	1.50000001
$f(\bar{x})$	-2.4214×10^{-8}	-4.1910×10^{-8}	1.3970×10^{-8}
erro em x	2.2491×10^{-4}	2.9079×10^{-4}	3.5082×10^{-5}
nº de iterações	12	35	27

Observamos que nos testes 1 e 2 a raiz encontrada foi a raiz dupla $\xi_1 = 1$. Era de se esperar que o número de iterações fosse grande, pois $\xi_1 = 1$ é zero de $f'(x)$. No entanto, o método conseguiu encontrar a raiz (pois, para as seqüências $\{x_k\}$ geradas, o valor de $f(x_k)$ tendeu a zero mais rapidamente que o valor de $f'(x_k)$).

Temos que $f'(x) = 3x^2 - 7x + 4 \Rightarrow f'(x) = 0 \Leftrightarrow x_1 = 1$ e $x_2 = 4/3 = 1.33333\dots$ Observe que nos testes 2 e 3 tomamos propositalmente x_0 bem próximo de $4/3$; no teste 2, $x_0 < 4/3$ e o método encontrou $\xi_1 = 1$ e, no teste 3, $x_0 > 4/3$ e a raiz encontrada foi $\xi_2 = 1.5$. Uma análise do gráfico de $f(x)$ (Figura 2.23) nos ajuda a entender este fato.

No teste 4, aplicamos o método da bissecção até reduzir o intervalo $[0.5, 2]$ a um intervalo de amplitude 0.01 e tomamos como aproximação inicial para o método de Newton o ponto médio desse intervalo: $x_0 = 1.50194313$. A partir desse ponto inicial foram executadas duas iterações do método de Newton e obtivemos os seguintes resultados:

$$\bar{x} = 1.5 \text{ e } f(\bar{x}) = 2.3 \times 10^{-10}.$$

Devemos observar que no intervalo inicial para o método da bissecção existem duas raízes distintas $\xi_1 = 1$ e $\xi_2 = 1.5$ e a raiz obtida foi $\bar{x} = 1.5$; isto ocorreu porque o método da bissecção ignora raízes com multiplicidade par, que é o caso de $\xi_1 = 1$.

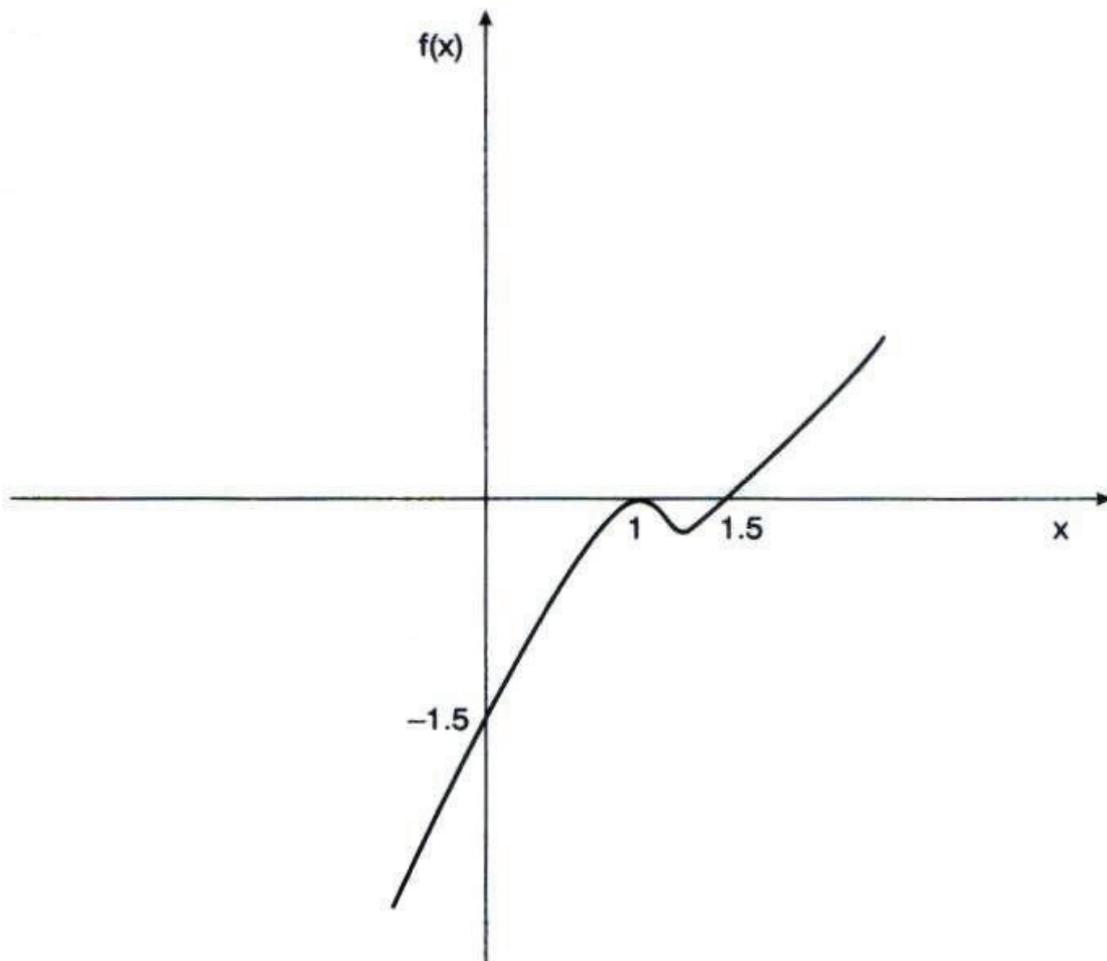


Figura 2.23

2.5 ESTUDO ESPECIAL DE EQUAÇÕES POLINOMIAIS

2.5.1 INTRODUÇÃO

Embora possamos usar qualquer um dos métodos vistos anteriormente para encontrar um zero de um polinômio, o fato de os polinômios aparecerem com tanta frequência em aplicações faz com que lhes dediquemos especial atenção.

Normalmente, um polinômio de grau n é escrito na forma

$$p_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n, \quad a_n \neq 0 \quad (6)$$

Se $n = 2$, sabemos da álgebra elementar como achar os zeros de $p_2(x)$. Existem fórmulas fechadas, semelhantes à fórmula para polinômios de grau 2, mas bem mais complicadas, para zeros de polinômios de grau 3 e 4. Agora, para $n \geq 5$, em geral, não existem fórmulas explícitas e somos forçados a usar métodos iterativos para encontrar zeros de polinômios.

Vários teoremas da álgebra são úteis na localização e classificação dos tipos de zeros de um polinômio.

Faremos nosso estudo dividido em duas partes:

- 1) localização de raízes,
- 2) determinação das raízes reais.

2.5.2 LOCALIZAÇÃO DE RAÍZES

Alguns teoremas são úteis ao nosso estudo:

TEOREMA 4 (Teorema Fundamental da Álgebra) (4)

“Se $p_n(x)$ é um polinômio de grau $n \geq 1$, ou seja, $p_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$, a_0, a_1, \dots, a_n reais ou complexos, com $a_n \neq 0$, então $p_n(x)$ tem pelo menos um zero, ou seja, existe um número complexo ξ tal que $p_n(\xi) = 0$.”

Para determinarmos o número de zeros reais de um polinômio com coeficientes reais, podemos fazer uso da *regra de sinal de Descartes*:

“Dado um polinômio com coeficientes reais, o número de zeros reais positivos, p , desse polinômio não excede o número v de variações de sinal dos coeficientes. Ainda mais, $v - p$ é inteiro, par, não negativo”.

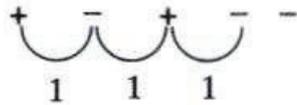
Exemplo 23

a) $p_5(x) = 2x^5 - 3x^4 - 4x^3 + x + 1$



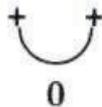
$$\Rightarrow v = 2 \Rightarrow p: \begin{cases} \text{se } v - p = 0, p = 2 \\ \text{se } v - p = 2, p = 0 \end{cases} \quad \text{ou}$$

$$b) p_5(x) = 4x^5 - x^3 + 4x^2 - x - 1$$



$$\Rightarrow v = 3 \text{ e } p: \begin{cases} \text{se } v - p = 0, p = 3 \\ \text{se } v - p = 2, p = 1 \end{cases} \quad \text{ou}$$

$$c) p_7(x) = x^7 + 1$$



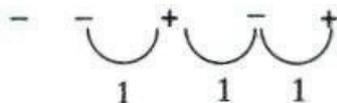
$$\Rightarrow v = 0 \text{ e } p: (v - p \geq 0) \Rightarrow p = 0.$$

Para determinar o número de raízes reais negativas, neg, tomamos $p_n(-x)$ e usamos a regra para raízes positivas:

Exemplo 24

$$a) p_5(x) = 2x^5 - 3x^4 - 4x^3 + x + 1$$

$$p_5(-x) = -2x^5 - 3x^4 + 4x^3 - x + 1$$



$$\Rightarrow v = 3 \text{ e } \text{neg}: \begin{cases} \text{se } v - \text{neg} = 0, \text{neg} = 3 \\ \text{se } v - \text{neg} = 2, \text{neg} = 1 \end{cases} \quad \text{ou}$$

b) $p_5(x) = 4x^5 - x^3 + 4x^2 - x - 1$

$p_5(-x) = -4x^5 + x^3 + 4x^2 + x - 1$



$\Rightarrow v = 2$ e neg: $\begin{cases} \text{se } v - \text{neg} = 0, \text{ neg} = 2 \\ \text{se } v - \text{neg} = 2, \text{ neg} = 0 \end{cases}$ ou

c) No caso do exemplo $p_7(x) = x^7 + 1$, vimos que não existe zero positivo. Temos ainda $p_7(0) = 1 \neq 0$. Como

$p_7(-x) = -x^7 + 1$



$\Rightarrow v = 1$ e neg: $\{v - \text{neg} = 0 \Rightarrow \text{neg} = 1\}$, ou seja, $p_n(x) = 0$, não tem raiz real positiva, o zero não é raiz e tem apenas uma raiz real negativa donde tem três raízes complexas conjugadas.

TEOREMA 5

Dado o polinômio $p_n(x)$ de grau n , se o desenvolvermos por Taylor em torno do ponto $x = \alpha$, temos

$$p_n(x) \approx p_n(\alpha) + p_n'(\alpha)(x - \alpha) + \frac{p_n''(\alpha)}{2!} (x - \alpha)^2 + \dots + \frac{p_n^{(n)}(\alpha)}{n!} (x - \alpha)^n.$$

Se chamarmos $x - \alpha = y$, ao acharmos o número de raízes reais de $p_n(y) = 0$ que são maiores que zero estaremos encontrando o número de raízes de $p_n(x) = 0$ que são maiores que α .

Podemos usar este resultado juntamente com a regra de sinal de Descartes para analisar as raízes de um determinado polinômio.

Se estamos interessados em estimar o número de zeros que um polinômio possui num intervalo $[\alpha, \beta]$ podemos também usar as *seqüências de Sturm*, que são construídas da seguinte maneira:

Dado o polinômio $p_n(x)$ e um número real α , vamos definir $\tilde{v}(\alpha)$ como sendo o número de variações de sinal em $\{g_i(\alpha)\}$ onde construímos a seqüência $g_0(\alpha), g_1(\alpha), \dots, g_n(\alpha)$, ignorando os zeros, assim:

$$\begin{cases} g_0(x) = p_n(x) \\ g_1(x) = p_n'(x) \end{cases}$$

e, para $k \geq 2$, $g_k(x)$ é o resto da divisão de g_{k-2} por g_{k-1} , com sinal trocado.

Exemplo 25

$$p_3(x) = x^3 + x^2 - x + 1$$

$$\begin{cases} g_0(x) = p_3(x) = x^3 + x^2 - x + 1 \\ g_1(x) = p_3'(x) = 3x^2 + 2x - 1 \end{cases}$$

$$g_2(x) = ?$$

$$x^3 + x^2 - x + 1$$

$$-x^3 - \frac{2}{3}x^2 + \frac{1}{3}x$$

$$\frac{1}{3}x^2 - \frac{2}{3}x + 1$$

$$-\frac{1}{3}x^2 - \frac{2}{9}x + \frac{1}{9}$$

$$-\frac{8}{9}x + \frac{10}{9}$$

$$3x^2 + 2x - 1$$

$$\frac{1}{3}x + \frac{1}{9}$$

$$\Rightarrow g_2(x) = \frac{8}{9}x - \frac{10}{9}$$

$$g_3(x) = ?$$

$$3x^2 + 2x - 1$$

$$\frac{8}{9}x - \frac{10}{9}$$

$$\frac{27}{8}x + \frac{207}{32}$$

$$\frac{99}{16}$$

$$\Rightarrow g_3(x) = -\frac{99}{16}$$

Assim, se $\alpha = 2$, por exemplo, temos

$$g_0(\alpha) = 11 > 0$$

$$g_1(\alpha) = 15 > 0$$

$$\left. \begin{array}{l} g_2(\alpha) = \frac{2}{3} > 0 \\ g_3(\alpha) = -\frac{99}{16} < 0 \end{array} \right\} 1$$

$$\Rightarrow \tilde{v}(\alpha) = \tilde{v}(2) = 1.$$

TEOREMA 6 (de Sturm) (17,30)

Se $p_n(\alpha) \neq 0$ e $p_n(\beta) \neq 0$, então o número de raízes distintas $p_n(x) = 0$ no intervalo $\alpha \leq x \leq \beta$ é exatamente $\tilde{v}(\alpha) - \tilde{v}(\beta)$.

Tomando $\beta = 3$, por exemplo, no polinômio do exemplo anterior:

$$g_0(\beta) = 34 > 0$$

$$g_1(\beta) = 32 > 0$$

$$\left. \begin{array}{l} g_2(\beta) = \frac{14}{9} > 0 \\ g_3(\beta) = -\frac{99}{16} < 0 \end{array} \right\} 1$$

$$\Rightarrow \tilde{v}(\beta) = \tilde{v}(3) = 1.$$

Então $x^3 + x^2 - x + 1 = 0$ não possui raízes reais no intervalo $[2, 3]$ pois $\tilde{v}(2) - \tilde{v}(3) = 1 - 1 = 0$.

Os teoremas a seguir fornecem regiões do plano que contém zeros de polinômios.

TEOREMA 7 (30)

Se $p_n(x)$ é um polinômio com coeficientes a_k , $k = 0, 1, \dots, n$ como em (6), então $p_n(x)$ tem pelo menos um zero no interior do círculo centrado na origem e de raio igual a $\min\{\rho_1, \rho_n\}$ onde

$$\rho_1 = n \frac{|a_0|}{|a_1|} \quad \rho_n = \sqrt[n]{\frac{|a_0|}{|a_n|}}$$

Exemplo 26

Se $p_5(x) = x^5 - 3.7x^4 + 7.4x^3 - 10.8x^2 + 10.8x - 6.8$; $n = 5$, $a_5 = 1$, $a_1 = 10.8$, $a_0 = -6.8$

Assim,

$$\rho_1 = 5 \left(\frac{6.8}{10.8} \right) = 3.14 \dots \quad \rho_5 = \sqrt[5]{\frac{6.8}{1}} = 1.46 \dots$$

Então $p_5(x)$ tem pelo menos um zero (real ou complexo) no círculo de raio 1.46... ou seja, $|x| \leq 1.46 \dots$

TEOREMA 8

Se $p_n(x)$ é o polinômio (6) e se

$$r \approx 1 + \max_{0 \leq k \leq n-1} \frac{|a_k|}{|a_n|}$$

então cada zero de $p_n(x)$ se encontra na região circular definida por $|x| \leq r$.

Exemplo 27

Seja $p_3(x) = x^3 - x^2 + x - 1$.

Então $n = 3$, $a_0 = -1$, $a_1 = +1$, $a_2 = -1$, $a_3 = 1$

$$\frac{|a_0|}{|a_3|} = \frac{1}{1} = 1 \quad \frac{|a_1|}{|a_3|} = \frac{1}{1} = 1 \quad \frac{|a_2|}{|a_3|} = \frac{1}{1} = 1$$

$$\max_{0 \leq k \leq 2} \frac{|a_k|}{|a_3|} = \max \{1, 1, 1\} = 1.$$

Assim, $r = 1 + 1 = 2$. Então, todos os zeros de $p_3(x)$ se encontram num disco centrado na origem e com raio 2.

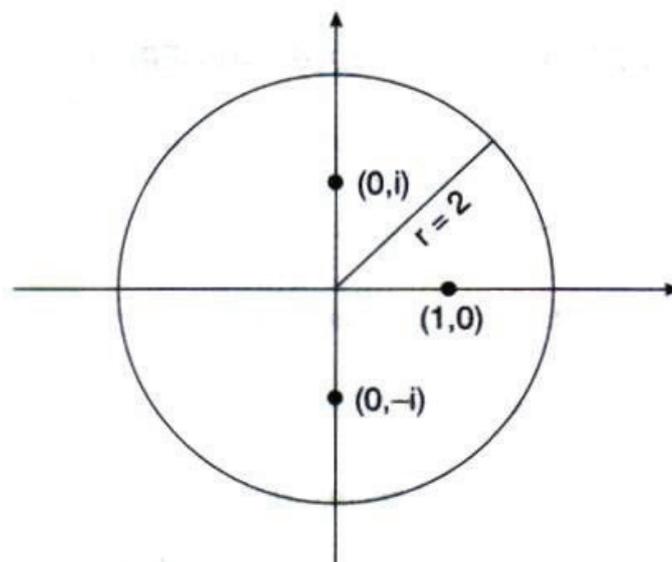


Figura 2.24

De fato, os zeros de $p_3(x)$ são:

$$x_1 = 1$$

$$x_2 = i$$

$$x_3 = -i.$$

2.5.3 DETERMINAÇÃO DAS RAÍZES REAIS

Para se obter raízes reais de equações polinomiais, pode-se aplicar qualquer um dos métodos numéricos estudados anteriormente.

Contudo, estas equações surgem tão frequentemente que merecem um estudo especial, conforme comentamos no início desta seção.

Conforme vimos, um polinômio de grau n com coeficientes reais será representado na forma (6) onde $a_i \in \mathbb{R}$, $i = 0, 1, 2, \dots, n$, ou seja:

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 \quad (a_n \neq 0)$$

Estudaremos um processo para se calcular o valor numérico de um polinômio, isto porque em qualquer dos métodos este cálculo deve ser feito uma ou mais vezes por iteração.

Por exemplo, no método de Newton, a cada iteração deve-se fazer uma avaliação do polinômio e uma de sua derivada.

MÉTODO PARA SE CALCULAR O VALOR NUMÉRICO DE UM POLINÔMIO

Para simplificar, estudaremos o processo analisando um polinômio de grau 4:

$$p_4(x) = a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0. \quad (7)$$

Este polinômio pode ser escrito na forma:

$$p_4(x) = (((a_4 x + a_3)x + a_2)x + a_1)x + a_0, \quad (8)$$

conhecida como forma dos *parênteses encaixados*.

Deve-se observar que, se o valor numérico de $p_4(x)$ for calculado pelo processo (8), o número de operações será bem menor que pelo processo (7).

Para um polinômio genérico de grau n , vemos que, pelo processo (8), teremos de efetuar n multiplicações e n adições.

No entanto, pelo processo (7), o número de adições é também n mas o número de multiplicações é $n + (n - 1) + \dots + 2 + 1 = \frac{(1 + n)n}{2}$ desde que x^j seja calculado por $x \cdot x \cdot x \dots \cdot x$, j vezes, pois a potenciação calculada desta forma introduz erros menores de arredondamento.

Agora,

$$\frac{n + n^2}{2} = \frac{n}{2} + \frac{n^2}{2} > n \Leftrightarrow n \geq 2, \text{ ou seja,}$$

o processo (8) efetua realmente um número menor de operações que o processo (7).

Temos então, no caso de $n = 4$, que

$$p_4(x) = \underbrace{\underbrace{\underbrace{((a_4x + a_3)x + a_2)x + a_1}_{b_4}x + a_0}_{b_3}}_{b_2} \dots$$

Para se calcular o valor numérico de $p_4(x)$ em $x = c$, basta fazer sucessivamente:

$$\begin{aligned} b_4 &= a_4 \\ b_3 &= a_3 + b_4c \\ b_2 &= a_2 + b_3c \\ b_1 &= a_1 + b_2c \\ b_0 &= a_0 + b_1c \end{aligned}$$

$$\Rightarrow p(c) = b_0.$$

Portanto, para $p_n(x)$ de grau n qualquer, calculamos $p_n(c)$ calculando as constantes b_j , $j = n, n-1, \dots, 1, 0$ sucessivamente, sendo:

$$b_n = a_n$$

$$b_j = a_j + b_{j+1}c \quad j = n-1, n-2, \dots, 2, 1, 0$$

e b_0 será o valor de $p_n(x)$ para $x = c$.

Como calcular o valor de $p'_n(x)$ em $x = c$ usando os coeficientes b_j obtidos anteriormente? Tomando como exemplo o polinômio de grau 4, temos

$$p_4(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 \Rightarrow$$

$$\Rightarrow p'_4(x) = 4a_4x^3 + 3a_3x^2 + 2a_2x + a_1.$$

Para $x = c$, temos que

$$\begin{aligned} b_4 = a_4 & \Rightarrow a_4 = b_4 \\ b_3 = a_3 + b_4c & \Rightarrow a_3 = b_3 - b_4c \\ b_2 = a_2 + b_3c & \Rightarrow a_2 = b_2 - b_3c \\ b_1 = a_1 + b_2c & \Rightarrow a_1 = b_1 - b_2c \\ b_0 = a_0 + b_1c & \Rightarrow a_0 = b_0 - b_1c \end{aligned}$$

Dado que já conhecemos b_0, b_1, b_2, b_3, b_4 :

$$\begin{aligned} p'_4(c) &= 4a_4c^3 + 3a_3c^2 + 2a_2c + a_1 \\ &= 4b_4c^3 + 3(b_3 - b_4c)c^2 + 2(b_2 - b_3c)c + (b_1 - b_2c) \\ &= 4b_4c^3 - 3b_4c^3 + 3b_3c^2 - 2b_3c^2 + 2b_2c + b_1 - b_2c. \end{aligned}$$

$$\text{Assim } p'_4(c) = b_4c^3 + b_3c^2 + b_2c + b_1$$

Aplicando o mesmo esquema anterior, teremos

$$\begin{aligned} c_4 &= b_4 \\ c_3 &= b_3 + c_4c \\ c_2 &= b_2 + c_3c \\ c_1 &= b_1 + c_2c. \end{aligned}$$

Calculamos, pois, os coeficientes c_j , $j = n, n - 1, \dots, 1$ da seguinte forma:

$$\begin{aligned} c_n &= b_n \\ c_j &= b_j + c_{j+1} c \quad j = n - 1, \dots, 1. \end{aligned}$$

Teremos então $p'(c) = c_1$.

MÉTODO DE NEWTON PARA ZEROS DE POLINÔMIOS

Seja $p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$ e x_0 uma aproximação inicial para a raiz procurada.

Conforme vimos, o método de Newton consiste em desenvolver aproximações sucessivas para ξ a partir da iteração:

$$x_{k+1} = x_k - \frac{p(x_k)}{p'(x_k)}$$

Usando as observações anteriores sobre o cálculo de $p(x_k)$ e $p'(x_k)$, construímos o seguinte:

ALGORITMO 6

Dados a_0, a_1, \dots, a_n , coeficientes de $p_n(x)$, x a aproximação inicial, ε_1 e ε_2 precisões desejadas e fixado itmax, o número máximo de iterações que serão permitidas,

- 1) $\text{deltax} = x$
- 2) $\left[\begin{array}{l} \text{Para } k = 1, \dots, \text{itmax, faça:} \\ \quad b = a_n \\ \quad c = b \\ \quad \left[\begin{array}{l} \text{Para } i = (n - 1), \dots, 1, \text{ faça:} \\ \quad \quad b = a_i + bx \\ \quad \quad c = b + cx \\ \quad b = a_0 + bx \end{array} \right. \\ \quad \text{Se } |b| \leq \varepsilon_1 \text{ vá para o passo 4} \\ \quad \text{deltax} = b / c \\ \quad x = x - \text{deltax} \\ \quad \text{Se } |\text{deltax}| \leq \varepsilon_2 \text{ vá para o passo 4} \end{array} \right.$

- 3) Imprimir mensagem de que não houve convergência com “itmax” iterações.
- 4) FIM.

Exemplo 28

Dada a equação polinomial $x^5 - 3.7x^4 + 7.4x^3 - 10.8x^2 + 10.8x - 6.8 = 0$, temos que

$$p_5(1) = -2.1$$

$$p_5(2) = 3.6.$$

Então, existe uma raiz no intervalo (1,2).

Partindo de $x_0 = 1.5$ e considerando $\varepsilon_1 = \varepsilon_2 = 10^{-6}$, o método de Newton para polinômios fornece:

$$\bar{x} = x_5 = 1.7, \quad f(\bar{x}) = 1.91 \times 10^{-6} \quad \text{e} \quad |x_5 - x_4| = 2.62 \times 10^{-7}.$$

Exemplo 29

Consideremos agora $p_3(x) = x^3 - 3x + 3 = 0$ e $\varepsilon_1 = \varepsilon_2 = 10^{-6}$. A Figura 2.25 mostra o gráfico cartesiano de $p_3(x)$.

Vemos assim que $x^3 - 3x + 3 = 0$ tem uma única raiz no intervalo $(-3, -1.5)$.

Executamos o método de Newton para polinômios, para este polinômio, duas vezes:

$$i) \quad \text{com } x_0 = -0.8, \quad \varepsilon_1 = \varepsilon_2 = 10^{-6}, \quad \text{itmax} = 30$$

$$ii) \quad \text{com } x_0 = -2, \quad \varepsilon_1 = \varepsilon_2 = 10^{-6}, \quad \text{itmax} = 10$$

Veja o efeito de pegarmos x_0 próximo a um zero da derivada e depois x_0 próximo à raiz:

No caso (i) foi encontrada $\bar{x} = -2.103801$ com $|f(\bar{x})| = 2.4 \times 10^{-7}$ em 17 iterações.

No caso (ii) foi obtido exatamente o mesmo resultado em 3 iterações.

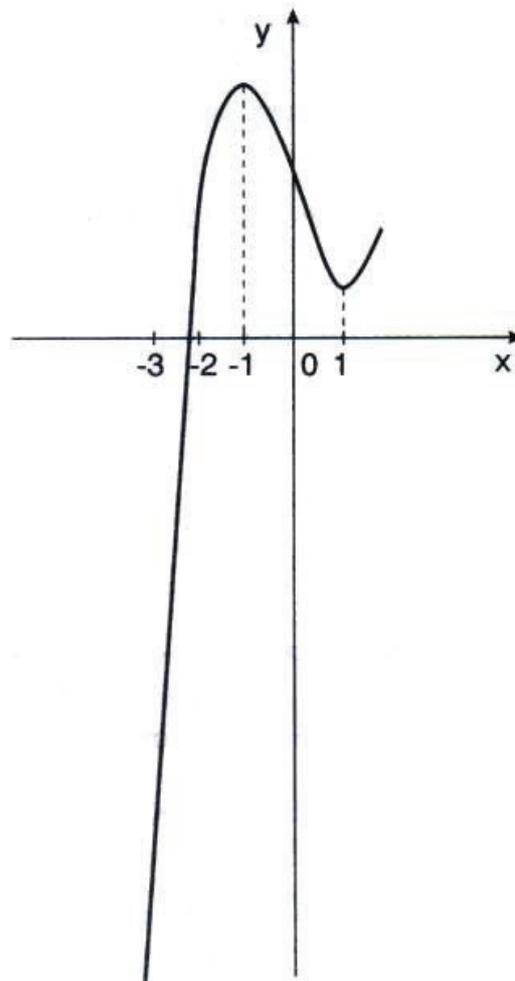


Figura 2.25

EXERCÍCIOS

1. Localize graficamente as raízes das equações a seguir:

a) $4 \cos(x) - e^{2x} = 0$

b) $\frac{x}{2} - \operatorname{tg}(x) = 0$

c) $1 - x \ln(x) = 0$

d) $2^x - 3x = 0$

e) $x^3 + x - 1000 = 0$

2. O método da bissecção pode ser aplicado sempre que $f(a)f(b) < 0$, mesmo que $f(x)$ tenha mais que um zero em (a, b) . Nos casos em que isto ocorre, verifique, com o auxílio de gráficos, se é possível determinar qual zero será obtido por este método.
3. Se no método da bissecção tomarmos sistematicamente $x = (a_k + b_k)/2$, teremos que $|\bar{x} - \xi| \leq (b_k - a_k)/2$.

Considerando este fato:

- a) estime o número de iterações que o método efetuará;
- b) escreva um novo algoritmo.
4. Seja $f(x) = x + \ln(x)$ que possui um zero no intervalo $I = [0.2, 2]$.

Se o objetivo for obter uma aproximação x_k para esta raiz de tal forma que $|x_k - \xi| < 10^{-5}$, é aconselhável usar o método da posição falsa tomando I como intervalo inicial? Justifique gráfica e analiticamente sem efetuar iterações numéricas.

Cite outros métodos nos quais este objetivo possa ser atingido.

5. Ao se aplicar o método do ponto fixo (MPF) à resolução de uma equação, obtivemos os seguintes resultados nas iterações indicadas:

$x_{10} = 1.5$	$x_{14} = 2.14128$
$x_{11} = 2.24702$	$x_{15} = 2.14151$
$x_{12} = 2.14120$	$x_{16} = 2.14133$
$x_{13} = 2.14159$	$x_{17} = 2.14147$

Escreva o que puder a respeito da raiz procurada.

6. a) Calcule b/a em uma calculadora que só soma, subtrai e multiplica.
- b) Calcule $3/13$ nessa calculadora.
7. A equação $x^2 - b = 0$ tem como raiz $\xi = \sqrt{b}$. Considere o MPF com $\varphi(x) = b/x$:
- a) comprove que $\varphi'(\xi) = -1$;
- b) o que acontece com a seqüência $\{x_k\}$ tal que $x_{k+1} = \varphi(x_k)$?

- c) sua conclusão do item (b) pode ser generalizada para qualquer equação $f(x) = 0$ que tenha $|\varphi'(\xi)| = 1$?
8. Verifique analiticamente que no MPF, se $\varphi'(x) < 0$ em I , intervalo centrado em ξ , então, dado $x_0 \in I$, a seqüência $\{x_k\}$, onde $x_{k+1} = \varphi(x_k)$, é oscilante em torno de ξ .
9. Se a função de iteração do MPF for tal que as condições do Teorema 2 estão satisfeitas:
- a) mostre que $|\xi - x_k| \leq \frac{M}{1 - M} |x_k - x_{k-1}|$;
- b) para que valores de M teremos então que $|\xi - x_k| < \varepsilon$ se $|x_k - x_{k-1}| < \varepsilon$?
10. Considere a função $f(x) = x^3 - x - 1$ (do Exemplo 19). Resolva-a pelo MPF com função de iteração $\varphi(x) = \frac{1}{x} + \frac{1}{x^2}$ e $x_0 = 1$. Justifique seus resultados.
11. Use o método de Newton-Raphson para obter a menor raiz positiva das equações a seguir com precisão $\varepsilon = 10^{-4}$
- a) $x/2 - \operatorname{tg}(x) = 0$
- b) $2 \cos(x) = e^x/2$
- c) $x^5 - 6 = 0$.
12. Aplique o método de Newton-Raphson à equação:
 $x^3 - 2x^2 - 3x + 10 = 0$ com $x_0 = 1.9$.
Justifique o que acontece.
13. Deduza o método de Newton a partir de sua interpretação geométrica.

14. Método de Newton Modificado:

Existe uma modificação no método de Newton na qual a função de iteração $\varphi(x)$ é dada por $\varphi(x) = x - \frac{f(x)}{f'(x_0)}$ onde x_0 é a aproximação inicial e é tal que $f'(x_0) \neq 0$.

- Com o auxílio de um gráfico, escreva a interpretação geométrica deste método.
- Cite algumas situações em que é conveniente usar este método em vez do método de Newton.

15. Seja $f(x) = e^x - 4x^2$ e ξ sua raiz no intervalo $(0, 1)$. Tomando $x_0 = 0.5$, encontre ξ com $\varepsilon = 10^{-4}$, usando:

- o MPF com $\varphi(x) = \frac{1}{2} e^{x/2}$;
- o método de Newton.
Compare a rapidez de convergência.

16. O valor de π pode ser obtido através da resolução das seguintes equações:

- $\text{sen}(x) = 0$
- $\text{cos}(x) + 1 = 0$

Aplice o método de Newton com $x_0 = 3$ e precisão 10^{-7} em cada caso e compare os resultados obtidos. Justifique.

17. Seja $f(x) = \text{sen}(x) - kx$.

- Encontre os valores positivos de k para que f tenha apenas uma raiz estritamente positiva.
- Encontre os valores positivos de k para que f tenha três raízes estritamente positivas.

18. Seja $f(x) = \frac{x^2}{2} + x(\ln(x) - 1)$. Obtenha seus pontos críticos com o auxílio de um método numérico.

19. O polinômio $p(x) = x^5 - \frac{10}{9}x^3 + \frac{5}{21}x$ tem seus cinco zeros reais, todos no intervalo $(-1, 1)$.

a) Verifique que $x_1 \in (-1, -0.75)$, $x_2 \in (-0.75, -0.25)$, $x_4 \in (0.3, 0.8)$ e $x_5 \in (0.8, 1)$.

b) Encontre, pelo respectivo método, usando $\epsilon = 10^{-5}$

x_1 : Newton ($x_0 = -0.8$)

x_2 : bissecção ($[a, b] = [-0.75, -0.25]$)

x_3 : posição falsa ($[a, b] = [-0.25, 0.25]$)

x_4 : MPF ($I = [0.2, 0.6]$, $x_0 = 0.4$)

x_5 : secante ($x_0 = 0.8$; $x_1 = 1$).

20. Seja a equação $f(x) = x - x \ln(x) = 0$.

Construa tabelas como as dos exemplos do final do capítulo para a raiz positiva desta equação. Use $\epsilon = 10^{-5}$.

Compare os diversos métodos considerando a garantia e rapidez de convergência e eficiência computacional em cada caso.

21. Seja $f(x) = xe^{-x} - e^{-3}$

a) verifique gráfica e analiticamente que $f(x)$ possui um zero no intervalo $(0, 1)$;

b) justifique teoricamente o comportamento da seqüência $\{x_k\}$ colocada a seguir, gerada pelo método de Newton para o cálculo do zero de $f(x)$ em $(0, 1)$, com $x_0 = 0.9$ e precisão $\epsilon = 5 \times 10^{-6}$.

$x_0 = +0.9$	$x_5 = -3.4962$	$x_{10} = -0.3041$
$x_1 = -6.8754$	$x_6 = -2.7182$	$x_{11} = 0.0427$
$x_2 = -6.0024$	$x_7 = -1.9863$	$x_{12} = 0.0440$
$x_3 = -5.1452$	$x_8 = -1.3189$	$x_{13} = 0.0480$
$x_4 = -4.3079$	$x_9 = -0.7444$	

22. Uma das dificuldades do método de Newton é o fato de uma aproximação x_k ser tal que $f'(x_k) = 0$. Uma modificação do algoritmo original para prever estes casos consiste em: dado λ um número positivo próximo de zero e supondo $|f'(x_n)| \geq \lambda$, a seqüência x_k é gerada através de:

$$x_{k+1} = x_k - f(x_k)/FL, \quad k = 0, 1, 2, \dots$$

onde

$$FL = \begin{cases} f'(x_k), & \text{se } |f'(x_k)| > \lambda \\ f'(x_w), & \text{caso contrário} \end{cases}$$

onde x_w é a última aproximação obtida tal que $|f'(x_w)| \geq \lambda$.

Pede-se:

- a) baseado no algoritmo de Newton, escreva um algoritmo para este método;
 - b) aplique este método à resolução da equação $x^3 - 9x + 3 = 0$, com $x_0 = -1.275$, $\lambda = 0.05$ e $\epsilon = 0.05$.
23. Usando a regra de sinal de Descartes e o Teorema 5, verifique que a equação $p(x) = 3x^5 - x^4 - x^3 + x + 1 = 0$ pode ter duas raízes reais no intervalo $[0, 1]$.
24. Resolva o Exercício 23 usando agora a seqüência de Sturm.
25. Encontre uma raiz da equação:
 $p(x) = x^4 - 6x^3 + 10x^2 - 6x + 9 = 0$, aplicando o método de Newton para polinômios.

PROJETOS

1. O PROBLEMA DE RAÍZES MÚLTIPLAS:

Se $f'(\xi) = 0$, o método de Newton perde suas características de convergência quadrática. O caso de $f'(\xi) = 0$ é um caso de raiz múltipla de $f(x) = 0$.

Definição: Dizemos que ξ é *raiz múltipla* de $f(x) = 0$ com multiplicidade p , quando $f(\xi) = f'(\xi) = \dots = f^{(p-1)}(\xi) = 0$ e $f^{(p)}(\xi) \neq 0$.

O método de Newton para raízes múltiplas é deduzido da seguinte forma: seja $f(x)$ uma função tal que ξ seja raiz de $f(x) = 0$ com multiplicidade p . A fórmula de Taylor para $f(x)$ numa vizinhança de ξ até o termo de ordem $(p-1)$, que é:

$$f(x) = f(\xi) + f'(\xi)(x - \xi) + \dots + f^{(p-1)}(\xi)(x - \xi)^{p-1} / (p-1)! + R_p(\xi_x)$$

onde $R_p(\xi) = f^{(p)}(\xi_p)(x - \xi)^p/p!$ com ξ_p entre x e ξ , fica:

$$f(x) = f^{(p)}(\xi_x)(x - \xi)^p / p!.$$

A dedução do método de Newton é baseada no modelo em que esta função $f(x)$ é tal que $f^{(p)}(x)$ é constante ($f^{(p)}(x) = b$), para x próximo a ξ . Para esta f ,

$$f(x) = b(x - \xi)^p/p! = c(x - \xi)^p, \quad f'(x) = cp(x - \xi)^{p-1}$$

e então

$$f(x)/f'(x) = (x - \xi)/p, \quad \text{donde } \xi = x - pf(x)/f'(x).$$

O MÉTODO

Se ξ é uma raiz de $f(x) = 0$ com multiplicidade p , dados x_0 uma aproximação inicial para ξ e $\varepsilon_1, \varepsilon_2$ precisões desejadas, para $k = 1, 2, \dots$, faça:

$$x_1 = x_0 - pf(x_0)/f'(x_0)$$

Se $|f(x_1)| < \varepsilon_1$ ou se $|x_1 - x_0| < \varepsilon_2$ então faça $\bar{x} = x_1$.

Caso contrário, $x_0 = x_1$ e recomece o processo.

De uma forma análoga, podemos introduzir um fator p no método da secante para trabalhar com raízes múltiplas, obtendo, dado x_0

$$x_{k+1} = x_k - (pf(x_k) (x_k - x_{k-1})) / (f(x_k) - f(x_{k-1})), \quad k = 0, 1, \dots$$

Temos os problemas de conseguir detectar computacionalmente a “proximidade” de uma raiz múltipla e também o de saber qual é essa multiplicidade, p .

Na referência [27] podem ser encontradas sugestões de como lidar com ambos.

- 1) Prove que, se ξ é raiz de multiplicidade p de $f(x) = 0$, a seqüência gerada pelo método de Newton converge quadraticamente, sob hipóteses adequadas de continuidade. Estabeleça essas hipóteses.
- 2) São dadas a seguir três funções com raízes múltiplas, a multiplicidade p das raízes, uma aproximação inicial x_0 e uma precisão ε .

i) $f(x) = |x - 9|^{4.5} / (1 + \text{sen}^2(x)); p = 4.5; x_0 = 6; \varepsilon = 10^{-6}.$

ii) $f(x) = (81 - y(108 - y(54 - y(12 - y)))) \text{sign}((y - 3)/(1 + x^2)),$
 $y = x + 1.11111; p = 3; x_0 = 1; \varepsilon = 10^{-6}.$

iii) $f(x) = |x - 8.3417|^{0.4} / (1 + x^2); p = 0.4; x_0 = 8.45; \varepsilon = 10^{-6}.$

- a)* Tente encontrar as raízes usando o método de Newton simples.
- b)* Idem com o método da secante.
- c)* Refaça agora os itens (*a*) e (*b*) com os dois métodos adaptados para raízes múltiplas.
- d)* Refaça o item (*c*) usando para p os valores:
 para a função (*i*), $p = 1$ e $p = 4.05$;
 para a função (*ii*), $p = 1$ e $p = 3.3$;
 para a função (*iii*), $p = 1$ e $p = 0.36$.

Em todos os casos imprima os valores:

NAF = número de avaliações de função que foram efetuadas

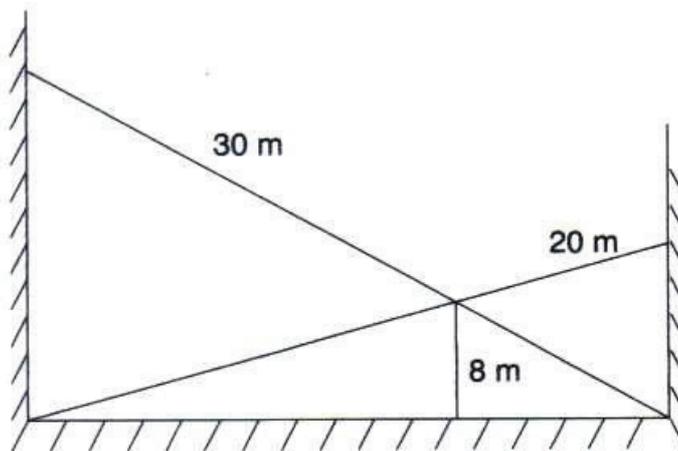
SOL = valor da “solução” encontrada

VAF = valor de $|f|$ na “solução” encontrada.

- 3) Use o método de Newton simples e o descrito anteriormente para encontrar os zeros de $f(x) = x^3 - 3.5x^2 + 4x - 1.5$. Localize-os, descubra p , escolha x_0 adequadamente e considere $\varepsilon = 10^{-6}$.

2. PROBLEMA DAS VIGAS

Duas vigas de madeira de 20 e 30 metros respectivamente se apóiam nas paredes de um galpão como mostra a figura. Se o ponto em que se cruzam está a 8 metros do solo, qual a largura deste galpão?



3. MÉTODO DE NEWTON GERANDO UMA SEQÜÊNCIA OSCILANTE

Analise algébrica e geometricamente e encontre justificativas para o comportamento do método de Newton quando aplicado à equação $p_3(x) = -0.5x^3 + 2.5x = 0$ nos seguintes casos:

a) $x_0 = 1$ e $x_0 = -1$

b) x_0 nos intervalos:

$$(-1, 1)$$

$$(1, 1.290994449)$$

$$(-1.290994449, -1)$$

$$(1.290994449, 2.236067977)$$

$$(-2.236067977, -1.290994449)$$

$$x_0 > 2.236067977$$

$$x_0 < -2.236067977.$$

RESOLUÇÃO DE SISTEMAS LINEARES

3.1 INTRODUÇÃO

A resolução de sistemas lineares é um problema que surge nas mais diversas áreas.

Exemplo 1

Considere o problema de determinar as componentes horizontal e vertical das forças que atuam nas junções da treliça abaixo:

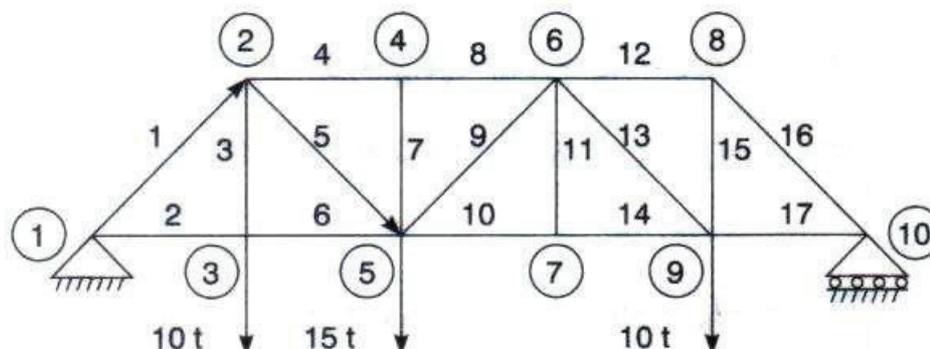


Figura 3.1

Para isto, temos de determinar as 17 forças desconhecidas que atuam nesta treliça. As componentes da treliça são supostamente presas nas junções por pinos, sem fricção.

Um teorema da mecânica elementar nos diz que, como o número de junções j está relacionado ao número de componentes m por $2j - 3 = m$, a treliça é estaticamente determinante; isto significa que as forças componentes são determinadas completamente pelas condições de equilíbrio estático nos nós.

Sejam F_x e F_y as componentes horizontal e vertical, respectivamente. Fazendo $\alpha = \sin(45^\circ) = \cos(45^\circ)$ e supondo pequenos deslocamentos, as condições de equilíbrio são:

$$\text{Junção 2} \begin{cases} \Sigma F_x = -\alpha f_1 + f_4 + \alpha f_5 = 0 \\ \Sigma F_y = -\alpha f_1 - f_3 - \alpha f_5 = 0 \end{cases}$$

$$\text{Junção 3} \begin{cases} \Sigma F_x = -f_2 + f_6 = 0 \\ \Sigma F_y = f_3 - 10 = 0 \end{cases}$$

$$\text{Junção 4} \begin{cases} \Sigma F_x = -f_4 + f_8 = 0 \\ \Sigma F_y = -f_7 = 0 \end{cases}$$

$$\text{Junção 5} \begin{cases} \Sigma F_x = -\alpha f_5 - f_6 + \alpha f_9 + f_{10} = 0 \\ \Sigma F_y = \alpha f_5 + f_7 + \alpha f_9 - 15 = 0 \end{cases}$$

$$\text{Junção 6} \begin{cases} \Sigma F_x = -f_8 - \alpha f_9 + f_{12} + \alpha f_{13} = 0 \\ \Sigma F_y = -\alpha f_9 - f_{11} - \alpha f_{13} = 0 \end{cases}$$

$$\text{Junção 7} \begin{cases} \Sigma F_x = -f_{10} + f_{14} = 0 \\ \Sigma F_y = f_{11} = 0 \end{cases}$$

$$\text{Junção 8} \begin{cases} \Sigma F_x = -f_{12} + \alpha f_{16} = 0 \\ \Sigma F_y = -f_{15} - \alpha f_{16} = 0 \end{cases}$$

$$\text{Junção 9} \begin{cases} \Sigma F_x = -\alpha f_{13} - f_{14} + f_{17} = 0 \\ \Sigma F_y = \alpha f_{13} + f_{15} - f_{10} = 0 \end{cases}$$

$$\text{Junção 10} \{ \Sigma F_x = -\alpha f_{16} - f_{17} = 0$$

Portanto, para obter as componentes pedidas é preciso resolver esse sistema linear, que tem 17 variáveis: f_1, f_2, \dots, f_{17} e 17 equações.

Um sistema linear com m equações e n variáveis é escrito usualmente na forma:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \cdot \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \quad \quad \quad \cdot \quad \quad \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

onde

$$a_{ij} : \text{coeficientes} \quad 1 \leq i \leq m, \quad 1 \leq j \leq n$$

$$x_j : \text{variáveis} \quad j = 1, \dots, n$$

$$b_i : \text{constantes} \quad i = 1, \dots, m$$

A resolução de um sistema linear consiste em calcular os valores de x_j , ($j = 1, \dots, n$), caso eles existam, que satisfaçam as m equações simultaneamente.

Usando notação matricial, o sistema linear pode ser assim representado:

$$Ax = b$$

onde

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \text{ é a matriz dos coeficientes,}$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \text{ é o vetor das variáveis}$$

e

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix} \text{ é o vetor constante.}$$

Chamaremos de x^* o vetor solução e de \bar{x} , uma solução aproximada do sistema linear $Ax = b$.

A formulação matricial do sistema $Ax = b$ do Exemplo 1, que será resolvido no final deste capítulo, é dada por:

$$A = \begin{bmatrix} -\alpha & 0 & 0 & 1 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\alpha & 0 & -1 & 0 & -\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\alpha & -1 & 0 & 0 & \alpha & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & 0 & 1 & 0 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -\alpha & 0 & 0 & 1 & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\alpha & 0 & -1 & 0 & -\alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\alpha & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\alpha & -1 \end{bmatrix}$$

$$b = [0 \ 0 \ 0 \ 10 \ 0 \ 0 \ 0 \ 15 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 10 \ 0]^T$$

$$x = [f_1 \ f_2 \ f_3 \ f_4 \ f_5 \ f_6 \ f_7 \ f_8 \ f_9 \ f_{10} \ f_{11} \ f_{12} \ f_{13} \ f_{14} \ f_{15} \ f_{16} \ f_{17}]^T$$

Analisaremos a seguir, através de exemplos com duas equações e duas variáveis, as situações que podem ocorrer com relação ao número de soluções de um sistema linear.

i) Solução única:

$$\begin{cases} 2x_1 + x_2 = 3 \\ x_1 - 3x_2 = -2 \end{cases} \quad \text{com } x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (1)$$

ii) Infinitas soluções:

$$\begin{cases} 2x_1 + x_2 = 3 \\ 4x_1 + 2x_2 = 6 \end{cases} \quad (2)$$

para o qual, qualquer $x^* = (\alpha, 3 - 2\alpha)^t$ com $\alpha \in \mathbb{R}$, é solução.

iii) Nenhuma solução:

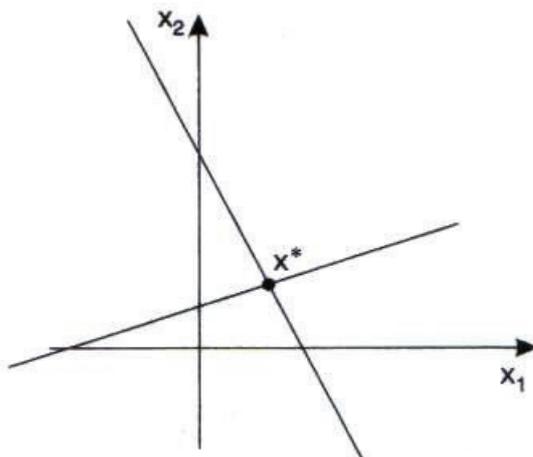
$$\begin{cases} 2x_1 + x_2 = 3 \\ 4x_1 + 2x_2 = 2 \end{cases} \quad (3)$$

Graficamente, cada um desses casos é representado respectivamente por:

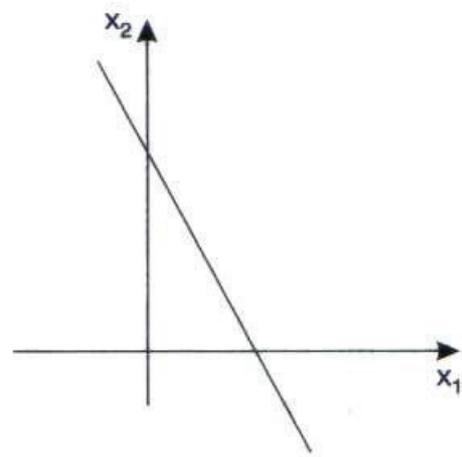
(1) retas concorrentes

(2) retas coincidentes

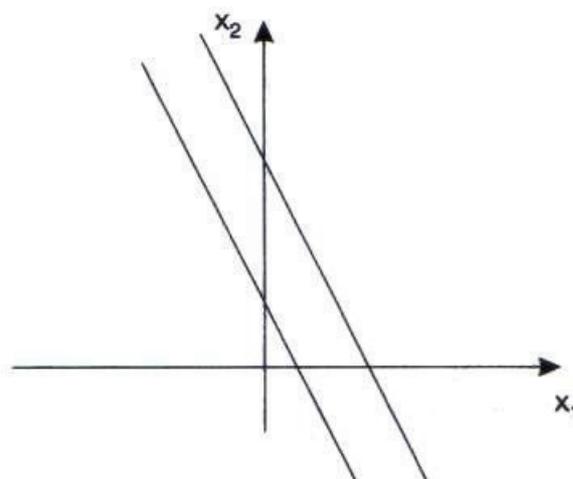
(3) retas paralelas



(1) retas concorrentes



(2) retas coincidentes



(3) retas paralelas

Figura 3.2

Mesmo no caso geral em que o sistema linear envolve m equações e n variáveis, apenas uma entre as situações abaixo irá ocorrer:

- i) o sistema linear tem solução única;
- ii) o sistema linear admite infinitas soluções;
- iii) o sistema linear não admite solução.

No caso em que $m = n = 2$ este fato foi facilmente verificado através dos gráficos das retas envolvidas no sistema, conforme mostra a Figura 3.2. Para analisar o caso geral, m equações e n variáveis, usaremos conceitos de Álgebra Linear.

Consideremos a matriz $A: m \times n$ como uma função que a cada vetor $x \in \mathbb{R}^n$ associa um vetor $b \in \mathbb{R}^m$, $b = Ax$:

$$A: \mathbb{R}^n \rightarrow \mathbb{R}^m \\ x \rightarrow b = Ax$$

Então, resolver o sistema linear $Ax = b$ consiste em:

“dado $b \in \mathbb{R}^m$ obter, caso exista, $x \in \mathbb{R}^n$, tal que $Ax = b$ ”.

A resolução de $Ax = b$ nos leva a encontrar respostas para as seguintes perguntas:

- existe $x^* \in \mathbb{R}^n$ tal que $Ax^* = b$?
- se existir, x^* é único?
- como obter x^* ?

Consideremos a matriz $A: 2 \times 2$,

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

Esta matriz associa a um vetor pertencente ao \mathbb{R}^2 um outro vetor do \mathbb{R}^2 .

Por exemplo:

$$\text{se } v = (1 \ 1)^T \quad \text{então: } u = Av = (3 \ -2)^T;$$

$$\text{se } w = (2 \ -1)^T \quad \text{então: } t = Aw = (3 \ 5)^T;$$

e, dado $b = (3 \ -2)^T$, existe um único $x^* = (1 \ 1)^T$ tal que $Ax^* = b$, conforme podemos comprovar graficamente através da Figura 3.2 (1).

Graficamente:

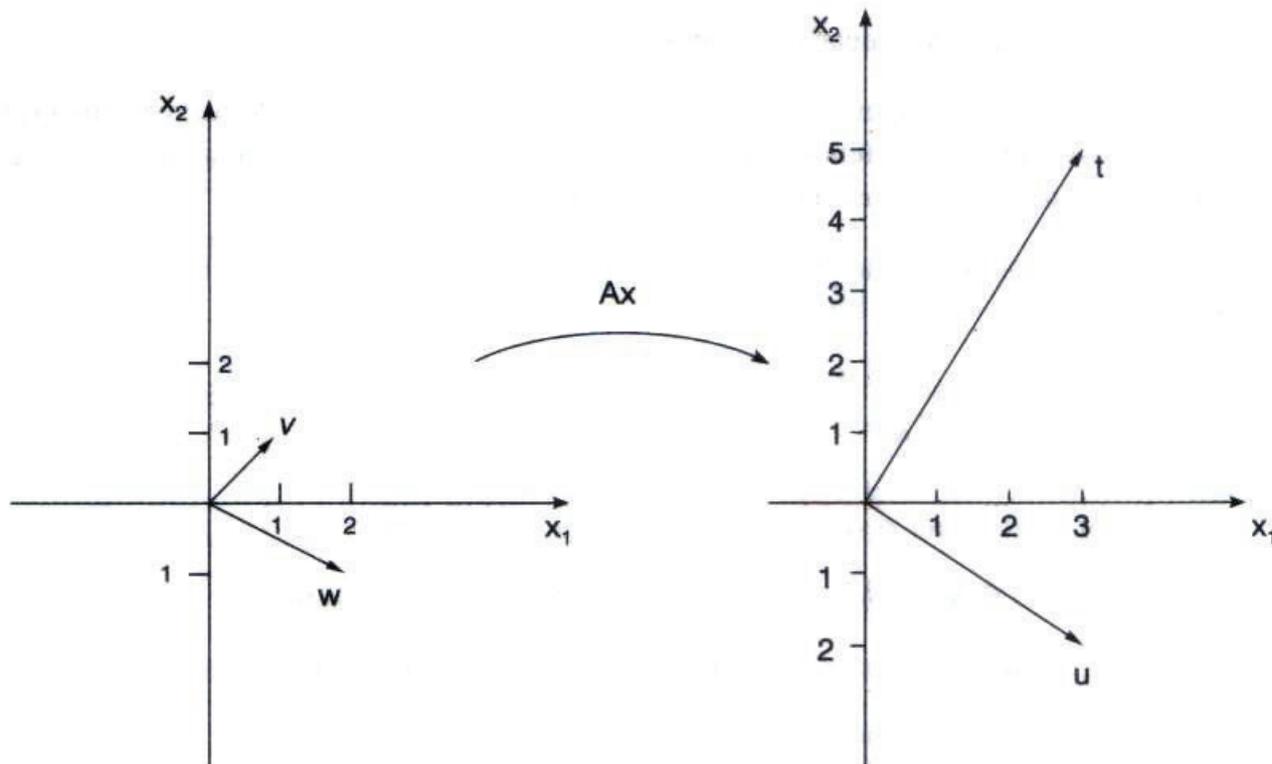


Figura 3.3

Dada uma matriz $A: m \times n$, definimos o conjunto Imagem de A (denotado por $\text{Im}(A)$) por:

$$\text{Im}(A) = \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n \mid y = Ax\}$$

O conjunto $\text{Im}(A)$ é um subespaço vetorial do \mathbb{R}^m .

Sob o ponto de vista das colunas de A , resolver o sistema linear $Ax = b$, $A: m \times n$ implica em se obter os escalares x_1, x_2, \dots, x_n que permitem escrever o vetor b de \mathbb{R}^m como combinação linear das n colunas de A .

$$b = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ \vdots \\ a_{m2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ \vdots \\ a_{mn} \end{pmatrix}$$

No sistema (1) as colunas da matriz $A = \begin{pmatrix} 2 & 1 \\ 1 & -3 \end{pmatrix}$ são linearmente independentes e portanto formam uma base para o \mathbb{R}^2 . Então, dado qualquer $u \in \mathbb{R}^2$, existem e são únicos os escalares $x_1 \in \mathbb{R}$ e $x_2 \in \mathbb{R}$ tais que

$$u = x_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ -3 \end{pmatrix}. \text{ Para este caso, temos } \text{Im}(A) = \mathbb{R}^2.$$

Na Figura 3.4 representamos os vetores coluna de A : $a^1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ e $a^2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ e o vetor $u = \begin{pmatrix} 5 \\ -1 \end{pmatrix} = 2a^1 + a^2$:

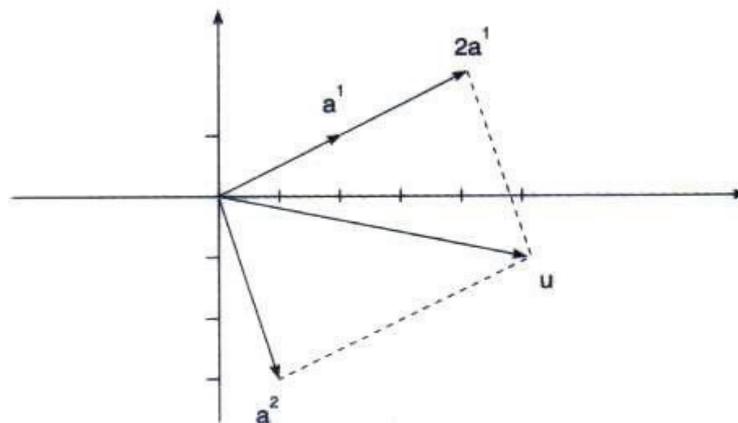


Figura 3.4

Definimos:

$\text{Posto}(A) = \text{dimensão}(\text{Im}(A)) = \text{dim}(\text{Im}(A))$.

Retomando os sistemas lineares (1), (2) e (3) do início desta secção:

$$\text{caso i): solução única} \begin{cases} 2x_1 + x_2 = 3 \\ x_1 - 3x_2 = -2 \end{cases}$$

neste caso, já vimos anteriormente que $\text{Im}(A) = \mathbb{R}^2$, portanto existe um único $x^* = (1 \ 1)^T$ tal que $b = 1a^1 + 1a^2$. Graficamente:

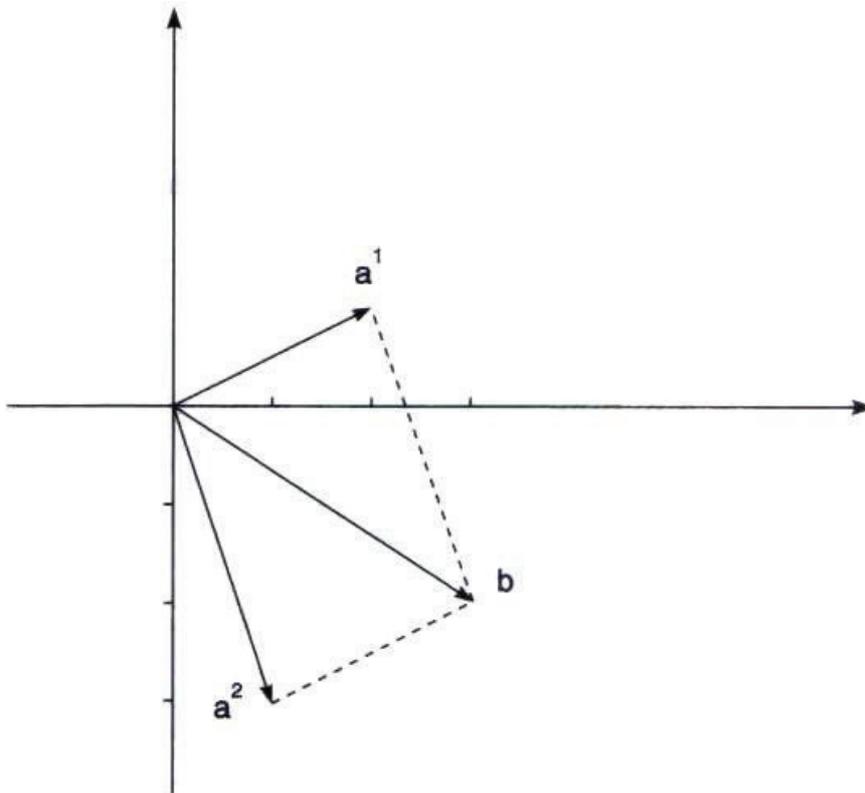


Figura 3.5

Assim, o sistema (1) é *compatível determinado*.

Nos casos (ii) e (iii) a matriz A dos coeficientes é $A = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$ na qual:
 $a^1 = 2a^2$.

Estas colunas são pois linearmente dependentes e conseqüentemente não formam uma base para o \mathbb{R}^2 ; para esta matriz, $\text{posto}(A) = \dim(\text{Im}(A)) = 1$. Dado um vetor $b \in \mathbb{R}^2$, se $b \in \text{Im}(A)$ o sistema linear $Ax = b$ admitirá infinitas soluções e será *compatível indeterminado*. Se $b \notin \text{Im}(A)$, o sistema linear não admitirá solução e será *incompatível*.

No caso (ii) $b = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$ pertence a $\text{Im}(A)$:

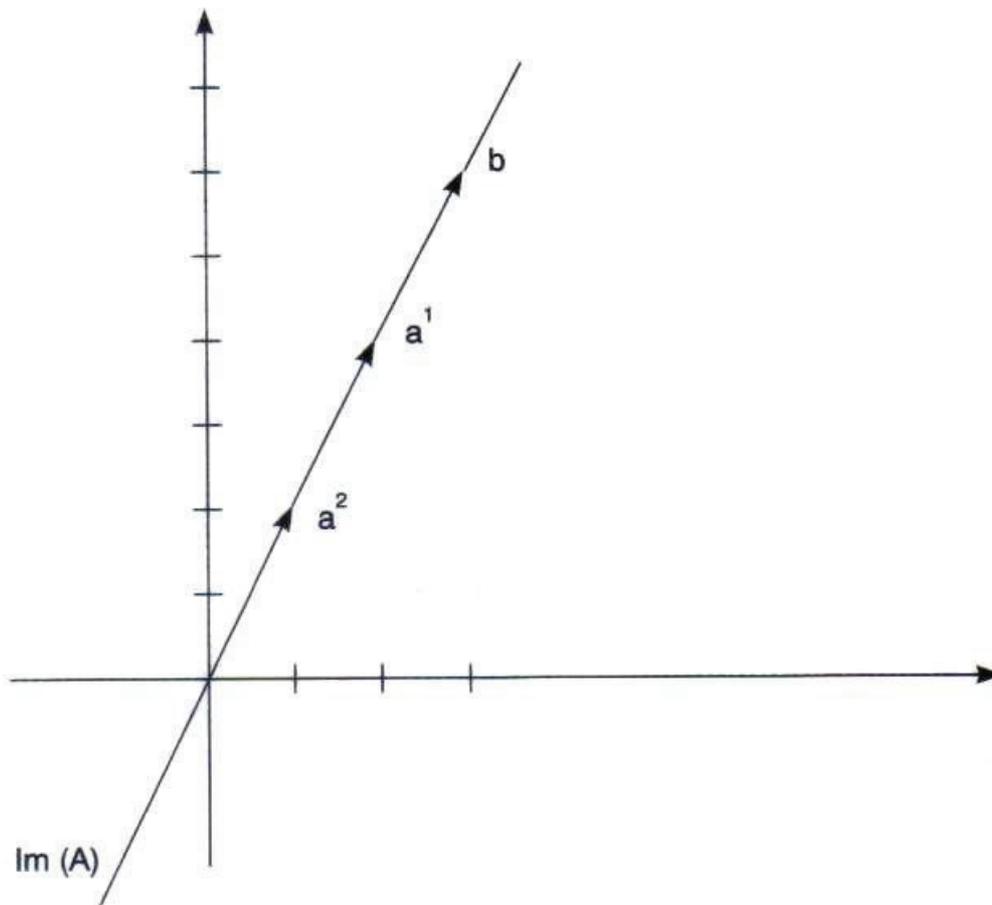


Figura 3.6

$$b = \alpha \begin{pmatrix} 2 \\ 4 \end{pmatrix} + (3 - 2\alpha) \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ para qualquer } \alpha \in \mathbb{R}.$$

No caso (iii), $b = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ não pertence a $\text{Im}(A)$:

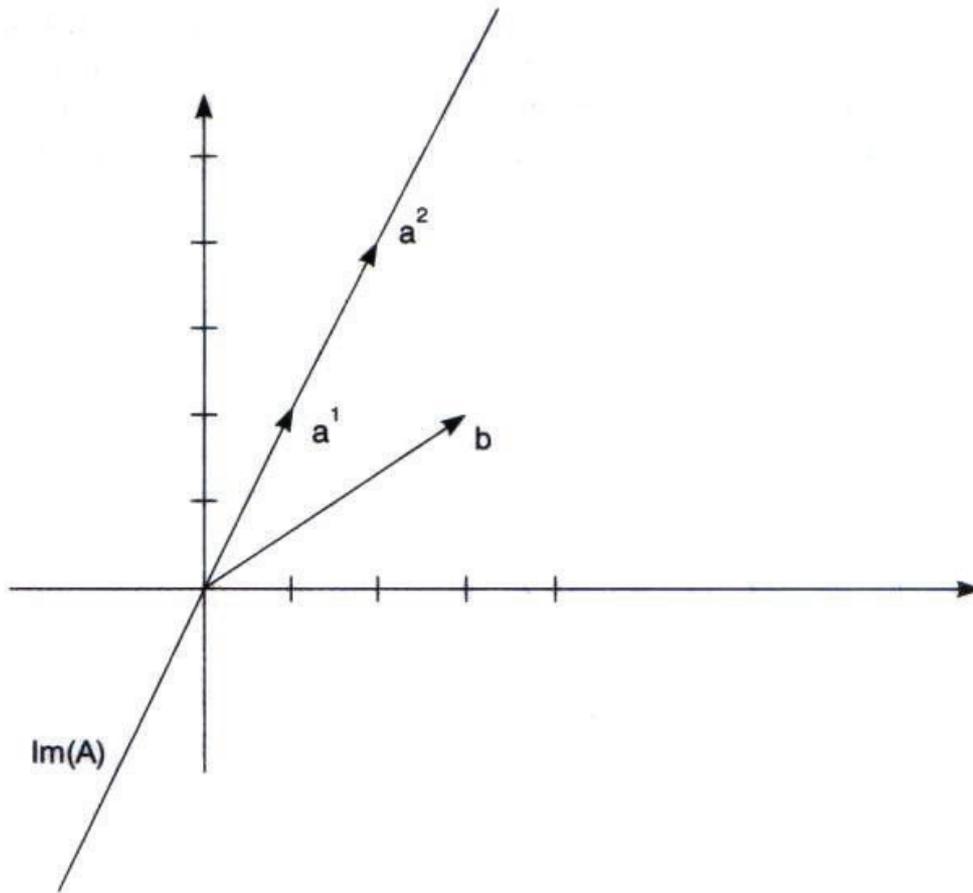


Figura 3.7

Nos casos em que $m \neq n$, embora tenhamos situações semelhantes, gostaríamos de observar que:

- i) $\text{posto}(A) \leq \min\{m,n\}$
- ii) se $m < n$ o sistema linear $Ax = b$ nunca poderá ter solução única pois $\text{posto}(A) < n$, sempre. Ilustrando,

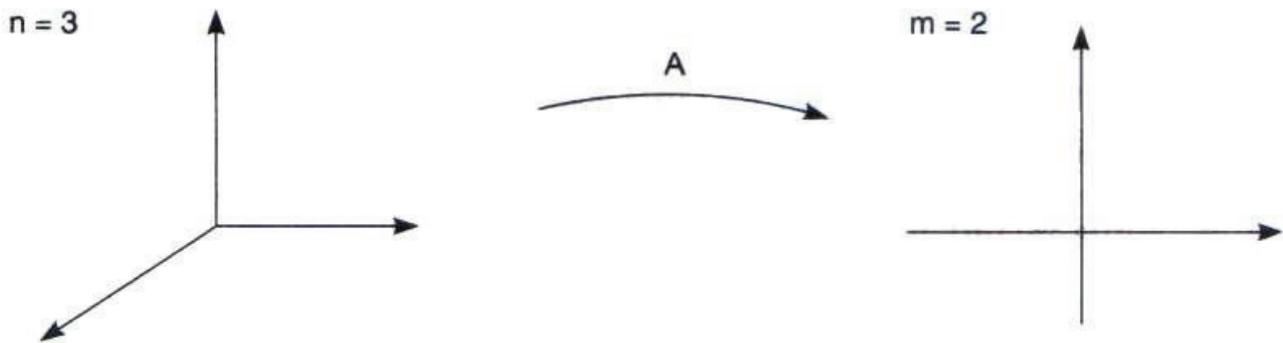


Figura 3.8

Por exemplo, consideremos o sistema linear:

$$\begin{cases} -x_1 + 2x_2 + 3x_3 = 6 \\ x_2 + x_3 = 9 \end{cases}$$

Eliminando x_2 da 2ª equação e substituindo na 1ª equação obtemos $x_1 = 12 + x_3$ e teremos o conjunto das infinitas soluções dado por:

$$\begin{aligned} S &= \{x \in \mathbb{R}^3 \text{ tais que } x = (12 + x_3 \quad 9 - x_3 \quad x_3)^T\} = \\ &= \{x \in \mathbb{R}^3 \text{ tais que } x = \begin{pmatrix} 12 \\ 9 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \forall x_3 \in \mathbb{R}\}. \end{aligned}$$

Neste exemplo, $\text{posto}(A) = m = 2 < n = 3$ e o sistema é compatível indeterminado.

iii) se $m > n$, mesmo que $\text{posto}(A) = n$ o sistema pode não ter solução pois a situação $b \notin \text{Im}(A)$ ocorre com frequência:

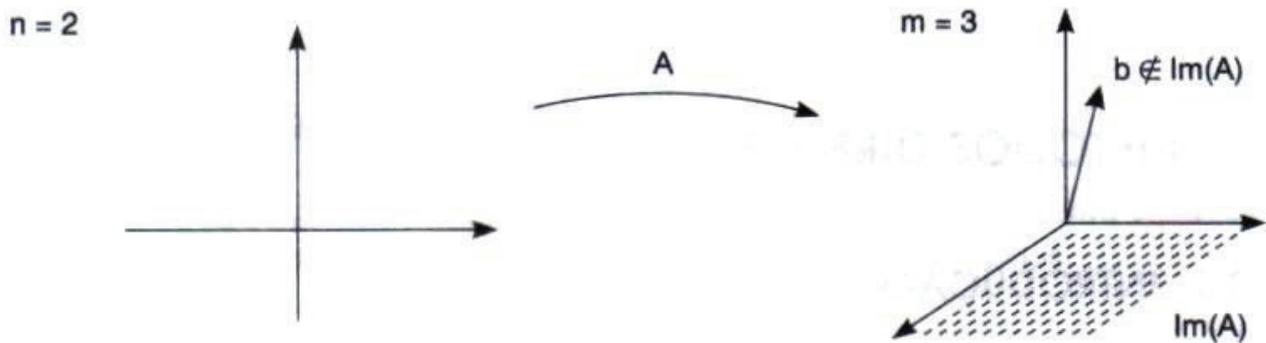


Figura 3.9

A tabela a seguir apresenta um resumo de todas as possibilidades para sistemas lineares:

Dada A , matriz $m \times n$ usaremos na tabela a seguinte definição:

Se $\text{posto}(A) = \min\{m, n\}$, então A é posto-completo.

Se $\text{posto}(A) < \min\{m, n\}$, então A é posto-deficiente.

Matriz A		$m = n$	$m < n$	$m > n$
Posto Completo		(posto(A) = n) Compatível determinado	(posto(A) = m) Infinitas soluções	(posto(A) = n) $b \in \text{Im}(A)$, solução única $b \notin \text{Im}(A)$, incompatível
Posto Deficiente	$b \in \text{Im}(A)$	Infinitas soluções	Infinitas soluções	Infinitas soluções
	$b \notin \text{Im}(A)$	Incompatível	Incompatível	Incompatível

Neste capítulo apresentaremos métodos numéricos para a resolução de sistemas lineares $n \times n$.

Os métodos numéricos para resolução de um sistema linear podem ser divididos em dois grupos: métodos diretos e métodos iterativos.

Métodos diretos são aqueles que, a menos de erros de arredondamento, fornecem a solução exata do sistema linear, caso ela exista, após um número finito de operações.

Os *métodos iterativos* geram uma seqüência de vetores $\{x^{(k)}\}$, a partir de uma aproximação inicial $x^{(0)}$. Sob certas condições esta seqüência converge para a solução x^* , caso ela exista.

3.2 MÉTODOS DIRETOS

3.2.1 INTRODUÇÃO

Pertencem a esta classe todos os métodos estudados nos cursos de 1º e 2º graus, destacando-se a regra de Cramer. Este método, aplicado à resolução de um sistema $n \times n$ envolve o cálculo de $(n + 1)$ determinantes de ordem n . Se n for igual a 20 podemos mostrar que o número total de operações efetuadas será $21 \times 20! \times 19$ multiplicações mais um número semelhante de adições. Assim, um computador que efetue cerca de cem milhões de multiplicações por segundo levaria 3×10^5 anos para efetuar as operações necessárias.

Desta forma, o estudo de métodos mais eficientes é necessário, pois, em geral, os problemas práticos exigem a resolução de sistemas lineares de grande porte, isto é, sistemas que envolvem um grande número de equações e variáveis.

Da última equação, temos

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

x_{n-1} pode então ser obtido da penúltima equação:

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

e assim sucessivamente obtém-se x_{n-2} , ..., x_2 e finalmente x_1 :

$$x_1 = \frac{b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n}{a_{11}}$$

ALGORITMO 1: Resolução de um Sistema Triangular Superior

Dado um sistema triangular superior $n \times n$ com elementos da diagonal da matriz A não nulos, as variáveis $x_n, x_{n-1}, x_{n-2}, \dots, x_2, x_1$ são assim obtidas:

$$x_n = b_n / a_{nn}$$

Para $k = (n - 1), \dots, 1$

$$\left[\begin{array}{l} s = 0 \\ \text{Para } j = (k + 1), \dots, n \\ s = s + a_{kj}x_j \\ x_k = (b_k - s) / a_{kk} \end{array} \right.$$

DESCRIÇÃO DO MÉTODO DA ELIMINAÇÃO DE GAUSS

Conforme dissemos anteriormente, o método consiste em transformar convenientemente o sistema linear original para obter um sistema linear equivalente com matriz dos coeficientes triangular superior.

Para modificar convenientemente o sistema linear dado de forma a obter um sistema equivalente, faremos uso do teorema, cuja demonstração pode ser encontrada em [2].

TEOREMA 1

Seja $Ax = b$ um sistema linear. Aplicando sobre as equações deste sistema uma seqüência de operações elementares escolhidas entre:

- i) trocar duas equações;
- ii) multiplicar uma equação por uma constante não nula;
- iii) adicionar um múltiplo de uma equação a uma outra equação;

obtemos um novo sistema $\tilde{A}x = \tilde{b}$ e os sistemas $Ax = b$ e $\tilde{A}x = \tilde{b}$ são equivalentes.

Descreveremos a seguir como o método da Eliminação de Gauss usa este teorema para triangularizar a matriz A . Vamos supor que $\det(A) \neq 0$.

A eliminação é efetuada por colunas e chamaremos de etapa k do processo a fase em que se elimina a variável x_k das equações $k + 1, k + 2, \dots, n$.

Usaremos a notação $a_{ij}^{(k)}$ para denotar o coeficiente da linha i e coluna j no final da k -ésima etapa, bem como $b_i^{(k)}$ será o i -ésimo elemento do vetor constante no final da etapa k .

Considerando que $\det(A) \neq 0$, é sempre possível reescrever o sistema linear de forma que o elemento da posição a_{11} seja diferente de zero, usando apenas a operação elementar (i):

$$\text{Seja } A^{(0)} | b^{(0)} = A | b = \left(\begin{array}{cccc|c} a_{11}^{(0)} & a_{12}^{(0)} & \dots & a_{1n}^{(0)} & b_1^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} & \dots & a_{2n}^{(0)} & b_2^{(0)} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ a_{n1}^{(0)} & a_{n2}^{(0)} & \dots & a_{nn}^{(0)} & b_n^{(0)} \end{array} \right)$$

onde $a_{ij}^{(0)} = a_{ij}$, $b_i^{(0)} = b_i$ e $a_{11}^{(0)} \neq 0$.

Etapa 1:

A eliminação da variável x_1 das equações $i = 2, \dots, n$ é feita da seguinte forma: da equação i subtraímos a 1ª equação multiplicada por m_{i1} . Observamos que para que esta eliminação seja efetuada, a única escolha possível é $m_{i1} = \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$, $i = 2, \dots, n$.

Os elementos $m_{i1} = \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$, $i = 2, \dots, n$ são os *multiplicadores* e o elemento $a_{11}^{(0)}$ é denominado *pivô* da 1ª etapa.

Ao final desta etapa teremos a matriz:

$$A^{(1)} | b^{(1)} = \left(\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} & b_2^{(1)} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} & b_n^{(1)} \end{array} \right)$$

onde

$$a_{ij}^{(1)} = a_{ij}^{(0)} \quad \text{para } j = 1, \dots, n$$

$$b_i^{(1)} = b_i^{(0)}$$

e

$$a_{ij}^{(1)} = a_{ij}^{(0)} - m_{i1} a_{1j}^{(0)} \quad i = 2, \dots, n \quad \text{e } j = 1, \dots, n$$

$$b_i^{(1)} = b_i^{(0)} - m_{i1} b_1^{(0)} \quad i = 2, \dots, n$$

Etapa 2:

Deve-se ter pelo menos um elemento $a_{i2}^{(1)} \neq 0$, para $i = 2, \dots, n$, caso contrário, $\det(A^{(1)}) = 0$, o que implica que $\det(A) = 0$; mas $\det(A) \neq 0$, por hipótese.

Então, é sempre possível reescrever a matriz $A^{(1)}$, sem alterar a posição da linha 1, de forma que o pivô, $a_{22}^{(1)}$, seja não nulo.

Os multiplicadores desta etapa serão os elementos $m_{i2} = \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}$ para $i = 3, \dots, n$.

A variável x_2 é eliminada das equações $i = 3, \dots, n$ da seguinte forma: da equação i subtraímos a segunda equação multiplicada por m_{i2} .

Ao final, teremos a matriz $A^{(2)} \mid b^{(2)}$:

$$A^{(2)} \mid b^{(2)} = \left(\begin{array}{cccccc|c} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} & \dots & a_{1n}^{(2)} & b_1^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2n}^{(2)} & b_2^{(2)} \\ 0 & 0 & a_{33}^{(2)} & \dots & a_{3n}^{(2)} & b_3^{(2)} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & a_{n3}^{(2)} & \dots & a_{nn}^{(2)} & b_n^{(2)} \end{array} \right)$$

onde $a_{ij}^{(2)} = a_{ij}^{(1)}$ para $i = 1, 2$ e $j = i, i + 1, \dots, n$

$$b_i^{(2)} = b_i^{(1)} \text{ para } i = 1, 2$$

e

$$a_{ij}^{(2)} = a_{ij}^{(1)} - m_{i2} a_{2j}^{(1)} \text{ para } i = 3, \dots, n \text{ e } j = 2, \dots, n$$

$$b_i^{(2)} = b_i^{(1)} - m_{i2} b_2^{(1)} \text{ para } i = 3, \dots, n$$

Seguindo raciocínio análogo, procede-se até a etapa $(n - 1)$ e a matriz, ao final desta etapa, será:

$$A^{(n-1)} | b^{(n-1)} = \left(\begin{array}{cccc|c} a_{11}^{(n-1)} & a_{12}^{(n-1)} & a_{13}^{(n-1)} & \dots & a_{1n}^{(n-1)} & b_1^{(n-1)} \\ 0 & a_{22}^{(n-1)} & a_{23}^{(n-1)} & \dots & a_{2n}^{(n-1)} & b_2^{(n-1)} \\ 0 & 0 & a_{33}^{(n-1)} & \dots & a_{3n}^{(n-1)} & b_3^{(n-1)} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \dots & a_{nn}^{(n-1)} & b_n^{(n-1)} \end{array} \right)$$

e o sistema linear $A^{(n-1)}x = b^{(n-1)}$ é triangular superior e equivalente ao sistema linear original.

Exemplo 2

Seja o sistema linear:

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

Etapa 1:

Eliminar x_1 das equações 2 e 3:

Para facilitar o entendimento do processo, de agora em diante usaremos a notação L_i para indicar o vetor linha formado pelos elementos da linha i da matriz $A^{(k)} | b^{(k)}$. Assim, nesta etapa, $L_1 = (3 \ 2 \ 4 \ 1)$.

$$A^{(0)} | b^{(0)} = \left(\begin{array}{ccc|c} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} & b_1^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} & a_{23}^{(0)} & b_2^{(0)} \\ a_{31}^{(0)} & a_{32}^{(0)} & a_{33}^{(0)} & b_3^{(0)} \end{array} \right) = \left(\begin{array}{ccc|c} 3 & 2 & 4 & 1 \\ 1 & 1 & 2 & 2 \\ 4 & 3 & -2 & 3 \end{array} \right)$$

$$\text{Pivô: } a_{11}^{(0)} = 3$$

$$m_{21} = 1/3$$

$$m_{31} = 4/3$$

$$L_2 \leftarrow L_2 - m_{21} L_1$$

$$L_3 \leftarrow L_3 - m_{31} L_1$$

$$\Rightarrow A^{(1)} | b^{(1)} = \left(\begin{array}{ccc|c} 3 & 2 & 4 & 1 \\ 0 & 1/3 & 2/3 & 5/3 \\ 0 & 1/3 & -22/3 & 5/3 \end{array} \right) = \left(\begin{array}{ccc|c} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & b_2^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} & b_3^{(1)} \end{array} \right)$$

Etapa 2:

Eliminar x_2 da equação 3:

$$\text{Pivô: } a_{22}^{(1)} = 1/3$$

$$m_{32} = \frac{1/3}{1/3} = 1$$

$$L_3 \leftarrow L_3 - m_{32} L_2$$

$$\Rightarrow A^{(2)} | b^{(2)} = \left(\begin{array}{ccc|c} 3 & 2 & 4 & 1 \\ 0 & 1/3 & 2/3 & 5/3 \\ 0 & 0 & -8 & 0 \end{array} \right)$$

Assim, resolver $Ax = b$ é equivalente a resolver $A^{(2)}x = b^{(2)}$:

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

A solução deste sistema é o vetor $x^* = \begin{pmatrix} -3 \\ 5 \\ 0 \end{pmatrix}$.

ALGORITMO 2: Resolução de $Ax = b$ através da Eliminação de Gauss.

Seja o sistema linear $Ax = b$, $A: n \times n$, $x: n \times 1$, $b: n \times 1$.

Supor que o elemento que está na posição a_{kk} é diferente de zero no início da etapa k .

$$\text{Eliminação} \left[\begin{array}{l} \text{Para } k = 1, \dots, n-1 \\ \left[\begin{array}{l} \text{Para } i = k+1, \dots, n \\ m = \frac{a_{ik}}{a_{kk}} \\ a_{ik} = 0 \\ \text{Para } j = k+1, \dots, n \\ a_{ij} = a_{ij} - ma_{kj} \\ b_i = b_i - mb_k \end{array} \right. \end{array} \right.$$

$$\text{Resolução do sistema:} \left[\begin{array}{l} x_n = b_n / a_{nn} \\ \text{Para } k = (n-1), \dots, 2, 1 \\ \left[\begin{array}{l} s = 0 \\ \text{Para } j = (k+1), \dots, n \\ [s = s + a_{kj} x_j \\ x_k = (b_k - s) / a_{kk} \end{array} \right. \end{array} \right.$$

O algoritmo acima efetua, na fase da eliminação, $(4n^3 + 3n^2 - 7n) / 6$ operações e, para resolver o sistema triangular superior, o número de operações efetuadas é n^2 .

Assim, o total de operações para se resolver um sistema linear pelo método da Eliminação de Gauss é $(4n^3 + 9n^2 - 7n)/6$.

ESTRATÉGIAS DE PIVOTEAMENTO

Vimos que o algoritmo para o método da Eliminação de Gauss requer o cálculo dos multiplicadores:

$$m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \quad i = k + 1, \dots, n$$

em cada etapa k do processo.

O que acontece se o pivô for nulo? E se o pivô estiver próximo de zero?

Estes dois casos merecem atenção especial pois é impossível trabalhar com um pivô nulo. E trabalhar com um pivô próximo de zero pode conduzir a resultados totalmente imprecisos. Isto porque em qualquer calculadora ou computador os cálculos são efetuados com aritmética de precisão finita, e pivôs próximos de zero dão origem a multiplicadores bem maiores que a unidade que, por sua vez, origina uma ampliação dos erros de arredondamento.

Para se contornar estes problemas deve-se adotar uma *estratégia de pivoteamento*, ou seja, adotar um processo de escolha da linha e/ou coluna pivotal.

ESTRATÉGIA DE PIVOTEAMENTO PARCIAL

Esta estratégia consiste em:

- i) no início da etapa k da fase de eliminação, escolher para pivô o elemento de maior módulo entre os coeficientes: $a_{ik}^{(k-1)}$, $i = k, k + 1, \dots, n$;
- ii) trocar as linhas k e i se for necessário.

Exemplo 3

$n = 4$ e $k = 2$

$$A^{(1)} \mid b^{(1)} = \left(\begin{array}{cccc|c} 3 & 2 & 1 & -1 & 5 \\ 0 & 1 & 0 & 3 & 6 \\ 0 & -3 & -5 & 7 & 7 \\ 0 & 2 & 4 & 0 & 15 \end{array} \right)$$

Início da etapa 2:

i) escolher pivô

$$\max_{j=2,3,4} |a_{j2}^{(1)}| = |a_{32}^{(1)}| = 3 \Rightarrow \text{pivô} = -3$$

ii) trocar linhas 2 e 3.

Assim,

$$A^{(1)} \mid b^{(1)} = \left(\begin{array}{cccc|c} 3 & 2 & 1 & -1 & 5 \\ 0 & -3 & -5 & 7 & 7 \\ 0 & 1 & 0 & 3 & 6 \\ 0 & 2 & 4 & 0 & 15 \end{array} \right)$$

e os multiplicadores desta etapa serão:

$$m_{32} = \frac{1}{-3} = -1/3$$

$$m_{42} = \frac{2}{-3} = -2/3$$

Observamos que a escolha do maior elemento em módulo entre os candidatos a pivô faz com que os multiplicadores, em módulo, estejam entre zero e um, o que evita a ampliação dos erros de arredondamento.

ESTRATÉGIA DE PIVOTEAMENTO COMPLETO

Nesta estratégia, no início da etapa k é escolhido para pivô o elemento de maior módulo, entre todos os elementos que ainda atuam no processo de eliminação:

$$\begin{aligned} \max_{i,j \geq k} |a_{ij}^{(k-1)}| = |a_{rs}^{(k-1)}| \Rightarrow \text{pivô} = a_{rs}^{(k-1)} \\ \forall i, j \geq k \end{aligned}$$

Observamos que, no Exemplo 3, se fosse adotada esta estratégia, o pivô da etapa 2 seria $a_{34}^{(1)} = 7$, o que acarretaria a troca das colunas 2 e 4 e, em seguida, das linhas 2 e 3, donde:

$$A^{(1)} | b^{(1)} = \left(\begin{array}{cccc|c} 3 & -1 & 1 & 2 & 5 \\ 0 & 7 & -5 & -3 & 7 \\ 0 & 3 & 0 & 1 & 6 \\ 0 & 0 & 4 & 2 & 15 \end{array} \right)$$

Esta estratégia não é muito empregada, pois envolve uma comparação extensa entre os elementos $a_{ij}^{(k-1)}$, $i, j \geq k$ e troca de linhas e colunas, conforme vimos no exemplo anterior; é evidente que todo este processo acarreta um esforço computacional maior que a estratégia de pivoteamento parcial.

Exemplo 4

Consideremos o sistema linear

$$\begin{cases} 0.0002x_1 + 2x_2 = 5 \\ 2x_1 + 2x_2 = 6 \end{cases}$$

Inicialmente vamos resolvê-lo sem a estratégia de pivoteamento parcial e vamos supor que temos de trabalhar com aritmética de três dígitos. Nosso sistema é:

$$\begin{cases} 0.2 \times 10^{-3}x_1 + 0.2 \times 10^1x_2 = 0.5 \times 10^1 \\ 0.2 \times 10^1x_1 + 0.2 \times 10^1x_2 = 0.6 \times 10^1 \end{cases}$$

Então,

$$A^{(0)} \mid b^{(0)} = \left(\begin{array}{cc|c} 0.2 \times 10^{-3} & 0.2 \times 10^1 & 0.5 \times 10^1 \\ 0.2 \times 10^1 & 0.2 \times 10^1 & 0.6 \times 10^1 \end{array} \right)$$

Etapa 1:

Pivô: 0.2×10^{-3}

$$m_{21} = (0.2 \times 10^1)/(0.2 \times 10^{-3}) = 1 \times 10^4 = 0.1 \times 10^5 \text{ e } a_{21}^{(1)} = 0$$

$$\begin{aligned} a_{22}^{(1)} &= a_{22}^{(0)} - a_{12}^{(0)} \times m_{21} = 0.2 \times 10^1 - (0.2 \times 10^1) \times (0.1 \times 10^5) = \\ &= 0.2 \times 10^1 - 0.2 \times 10^5 = -0.2 \times 10^5 \end{aligned}$$

$$\begin{aligned} b_2^{(1)} &= b_2^{(0)} - b_1^{(0)} \times m_{21} = 0.6 \times 10^1 - (0.5 \times 10^1) \times (0.1 \times 10^5) = \\ &= 0.6 \times 10^1 - 0.5 \times 10^5 = -0.5 \times 10^5 \end{aligned}$$

$$\Rightarrow A^{(1)} \mid b^{(1)} = \left(\begin{array}{cc|c} 0.2 \times 10^{-3} & 0.2 \times 10^1 & 0.5 \times 10^1 \\ 0 & -0.2 \times 10^5 & -0.5 \times 10^5 \end{array} \right)$$

E a solução do sistema $A^{(1)}x = b^{(1)}$ resultante é

$$-0.2 \times 10^5x_2 = -0.5 \times 10^5 \Rightarrow x_2 = (0.5) / (0.2) = 2.5 = 0.25 \times 10$$

$$\Rightarrow 0.2 \times 10^{-3}x_1 + 0.2 \times 10^1 \times 0.25 \times 10^1 = 0.5 \times 10^1$$

$$\Rightarrow 0.2 \times 10^{-3}x_1 = 0.5 \times 10^1 - 0.05 \times 10^2 = 0.5 \times 10^1 - 0.5 \times 10^1 = 0$$

e, portanto, $\bar{x} = (0 \ 2.5)^T$.

É fácil verificar que \bar{x} não satisfaz a segunda equação, pois

$$2 \times 0 + 2 \times 2.5 = 5 \neq 6.$$

Usando agora a estratégia de pivoteamento parcial (e ainda aritmética de três dígitos), temos

$$A^{(0)} \mid b^{(0)} = \left(\begin{array}{cc|c} 0.2 \times 10^1 & 0.2 \times 10^1 & 0.6 \times 10^1 \\ 0.2 \times 10^{-3} & 0.2 \times 10^1 & 0.5 \times 10^1 \end{array} \right)$$

Assim o pivô é 0.2×10^1 e $m_{21} = (0.2 \times 10^{-3}) / (0.2 \times 10^1) = 0.1 \times 10^{-3}$. De forma análoga ao que fizemos acima, obtemos o novo sistema

$$A^{(1)} \mid b^{(1)} = \left(\begin{array}{cc|c} 0.2 \times 10^1 & 0.2 \times 10^1 & 0.6 \times 10^1 \\ 0 & 0.2 \times 10^1 & 0.5 \times 10^1 \end{array} \right)$$

cuja solução é $\bar{x} = \begin{pmatrix} 0.5 \\ 0.25 \times 10^1 \end{pmatrix}$

E o vetor \bar{x} é realmente a solução do nosso sistema, pois

$$0.2 \times 10^{-3} \times 0.5 + 0.2 \times 10^1 \times 0.25 \times 10^1 = 0.1 \times 10^{-3} + 0.05 \times 10^2 = 0.5 \times 10^1 = 5$$

e

$$\begin{aligned} 0.2 \times 10^1 \times 0.5 + 0.2 \times 10^1 \times 0.25 \times 10^1 &= 0.1 \times 10^1 + 0.05 \times 10^2 = \\ &= 0.01 \times 10^2 + 0.05 \times 10^2 = 0.06 \times 10^2 = 0.6 \times 10^1 = 6. \end{aligned}$$

3.2.3 FATORAÇÃO LU

Seja o sistema linear $Ax = b$.

O *processo de fatoração* para resolução deste sistema consiste em decompor a matriz A dos coeficientes em um produto de dois ou mais fatores e, em seguida, resolver uma seqüência de sistemas lineares que nos conduzirá à solução do sistema linear original.

Por exemplo, se pudermos realizar a fatoração: $A = CD$, o sistema linear $Ax = b$ pode ser escrito:

$$(CD)x = b$$

Se $y = Dx$, então resolver o sistema linear $Ax = b$ é equivalente a resolver o sistema linear $Cy = b$ e, em seguida, o sistema linear $Dx = y$.

A vantagem dos processos de fatoração é que podemos resolver qualquer sistema linear que tenha A como matriz dos coeficientes. Se o vetor b for alterado, a resolução do novo sistema linear será quase que imediata.

A fatoração LU é um dos processos de fatoração mais empregados. Nesta fatoração a matriz L é triangular inferior com diagonal unitária e a matriz U é triangular superior.

CÁLCULO DOS FATORES L e U

Os fatores L e U podem ser obtidos através de fórmulas para os elementos l_{ij} e u_{ij} , ou então, podem ser construídos usando a idéia básica do método da Eliminação de Gauss.

A obtenção dos fatores L e U pelas fórmulas dificulta o uso de estratégias de pivoteamento e, por esta razão, veremos como obter L e U através do processo de Gauss.

Usaremos um exemplo teórico de dimensão 3:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}$$

Trabalharemos somente com a matriz dos coeficientes. Seja então:

$$A^{(0)} = \begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} & a_{23}^{(0)} \\ a_{31}^{(0)} & a_{32}^{(0)} & a_{33}^{(0)} \end{pmatrix} = A$$

Os multiplicadores da etapa 1 do processo de Gauss são:

$$m_{21} = \frac{a_{21}^{(0)}}{a_{11}^{(0)}} \quad \text{e} \quad m_{31} = \frac{a_{31}^{(0)}}{a_{11}^{(0)}} \quad (\text{supondo que } a_{11}^{(0)} \neq 0)$$

Para eliminar x_1 da linha i , $i = 2, 3$, multiplicamos a linha 1 por m_{i1} e subtraímos o resultado da linha i .

Os coeficientes $a_{ij}^{(0)}$ serão alterados para $a_{ij}^{(1)}$, onde:

$$a_{1j}^{(1)} = a_{1j}^{(0)} \quad \text{para } j = 1, 2, 3$$

$$a_{ij}^{(1)} = a_{ij}^{(0)} - m_{i1} a_{1j}^{(0)} \quad \text{para } i = 2, 3 \text{ e } j = 1, 2, 3$$

Estas operações correspondem a se pré-multiplicar a matriz $A^{(0)}$ pela matriz $M^{(0)}$, onde

$$M^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{pmatrix}, \text{ pois:}$$

$$\begin{aligned}
 \mathbf{M}^{(0)}\mathbf{A}^{(0)} &= \begin{pmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} & a_{23}^{(0)} \\ a_{31}^{(0)} & a_{32}^{(0)} & a_{33}^{(0)} \end{pmatrix} = \\
 &= \begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} \\ a_{21}^{(0)} - m_{21}a_{11}^{(0)} & a_{22}^{(0)} - m_{21}a_{12}^{(0)} & a_{23}^{(0)} - m_{21}a_{13}^{(0)} \\ a_{31}^{(0)} - m_{31}a_{11}^{(0)} & a_{32}^{(0)} - m_{31}a_{12}^{(0)} & a_{33}^{(0)} - m_{31}a_{13}^{(0)} \end{pmatrix} = \\
 &= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} \end{pmatrix} = \mathbf{A}^{(1)}
 \end{aligned}$$

Portanto, $\mathbf{M}^{(0)}\mathbf{A}^{(0)} = \mathbf{A}^{(1)}$ onde $\mathbf{A}^{(1)}$ é a mesma matriz obtida no final da etapa 1 do processo de Gauss.

Supondo agora que $a_{22}^{(1)} \neq 0$, o multiplicador da etapa 2 será: $m_{32} = \frac{a_{32}^{(1)}}{a_{22}^{(1)}}$

Para eliminar x_2 da linha 3, multiplicamos a linha 2 por m_{32} e subtraímos o resultado da linha 3.

Os coeficientes $a_{ij}^{(1)}$ serão alterados para:

$$a_{1j}^{(2)} = a_{1j}^{(1)} \quad \text{para } j = 1, 2, 3$$

$$a_{2j}^{(2)} = a_{2j}^{(1)} \quad \text{para } j = 2, 3$$

$$a_{3j}^{(2)} = a_{3j}^{(1)} - m_{32} a_{2j}^{(1)} \quad \text{para } j = 2, 3$$

onde As operações efetuadas em $A^{(1)}$ são equivalentes a pré-multiplicar $A^{(1)}$ por $M^{(1)}$,

$$M^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -m_{32} & 1 \end{pmatrix}, \text{ pois:}$$

$$\begin{aligned} M^{(1)}A^{(1)} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -m_{32} & 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} \end{pmatrix} = \\ &= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} \\ 0 & a_{32}^{(1)} - m_{32}a_{22}^{(1)} & a_{33}^{(1)} - m_{32}a_{23}^{(1)} \end{pmatrix} = \end{aligned}$$

$$= \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} \\ 0 & 0 & a_{33}^{(2)} \end{pmatrix}$$

Portanto, $M^{(1)}A^{(1)} = A^{(2)}$ onde $A^{(2)}$ é a mesma matriz obtida no final da etapa 2 do método da Eliminação de Gauss.

Temos então que:

$$A = A^{(0)}$$

$$A^{(1)} = M^{(0)}A^{(0)} = M^{(0)}A$$

$$A^{(2)} = M^{(1)}A^{(1)} = M^{(1)}M^{(0)}A^{(0)} = M^{(1)}M^{(0)}A$$

onde $A^{(2)}$ é triangular superior.

É fácil verificar que:

$$(M^{(0)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & 0 & 1 \end{pmatrix} \quad \text{e} \quad (M^{(1)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{32} & 1 \end{pmatrix}$$

Assim,

$$(M^{(0)})^{-1}(M^{(1)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{pmatrix}$$

$$\text{Então, } A = (M^{(1)} M^{(0)})^{-1} A^{(2)} = (M^{(0)})^{-1} (M^{(1)})^{-1} A^{(2)}$$

$$A = \begin{pmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} \\ 0 & 0 & a_{33}^{(2)} \end{pmatrix} = LU$$

$$\text{Ou seja: } L = (M^{(0)})^{-1}(M^{(1)})^{-1} \text{ e } U = A^{(2)}.$$

Isto é, fatoramos a matriz A em duas matrizes triangulares L e U , sendo que o fator L é triangular inferior com diagonal unitária e seus elementos l_{ij} para $i > j$ são os multiplicadores m_{ij} obtidos no processo da Eliminação de Gauss; o fator U é triangular superior e é a matriz triangular superior obtida no final da fase da triangularização do método da Eliminação de Gauss.

TEOREMA 2: (Fatoração LU)

Dada uma matriz quadrada A de ordem n , seja A_k a matriz constituída das primeiras k linhas e colunas de A . Suponha que $\det(A_k) \neq 0$ para $k = 1, 2, \dots, (n - 1)$. Então, existe uma única matriz triangular inferior $L = (m_{ij})$, com $m_{ii} = 1$, $1 \leq i \leq n$ e uma única matriz triangular superior $U = (u_{ij})$ tais que $LU = A$. Ainda mais, $\det(A) = u_{11}u_{22} \dots u_{nn}$.

Demonstração: ver [13]

RESOLUÇÃO DO SISTEMA LINEAR $Ax = b$ USANDO A FATORAÇÃO LU DE A

Dados o sistema linear $Ax = b$ e a fatoração LU da matriz A , temos:

$$Ax = b \Leftrightarrow (LU)x = b$$

Seja $y = Ux$. A solução do sistema linear pode ser obtida da resolução dos sistemas lineares triangulares:

$$i) \quad Ly = b$$

$$ii) \quad Ux = y$$

Verifiquemos teoricamente que o vetor y é o vetor constante do lado direito obtido ao final do processo da Eliminação de Gauss.

Considerando o sistema linear $Ly = b$, temos que $y = L^{-1}b$.

Mas, $L = (M^{(0)})^{-1}(M^{(1)})^{-1} \Rightarrow L^{-1} = M^{(1)}M^{(0)}$.

Então, $y = M^{(1)}M^{(0)}b^{(0)}$, onde $b^{(0)} = b$

Temos que

$$\begin{aligned} M^{(0)} b^{(0)} &= \begin{pmatrix} 1 & 0 & 0 \\ -m_{21} & 1 & 0 \\ -m_{31} & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1^{(0)} \\ b_2^{(0)} \\ b_3^{(0)} \end{pmatrix} = \begin{pmatrix} b_1^{(0)} \\ b_2^{(0)} - m_{21}b_1^{(0)} \\ b_3^{(0)} - m_{31}b_1^{(0)} \end{pmatrix} = \\ &= \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \end{pmatrix} = b^{(1)}. \end{aligned}$$

Isto é, o vetor obtido após o produto de $M^{(0)}$ por $b^{(0)}$ é o mesmo vetor do lado direito obtido após a etapa 1 do processo da Eliminação de Gauss.

Obtido $b^{(1)}$, temos que $y = M^{(1)}b^{(1)} =$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -m_{32} & 1 \end{pmatrix} \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} - m_{32}b_2^{(1)} \end{pmatrix} = \begin{pmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \end{pmatrix} = b^{(2)}.$$

Exemplo 5

Resolver o sistema linear a seguir usando a fatoração LU:

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

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

Usando o processo de Gauss, sem estratégia de pivoteamento parcial, para triangularizar A , temos:

Etapa 1:

$$\text{Pivô} = a_{11}^{(0)} = 3$$

$$\text{Multiplicadores: } m_{21} = \frac{a_{21}^{(0)}}{a_{11}^{(0)}} = \frac{1}{3} \text{ e } m_{31} = \frac{a_{31}^{(0)}}{a_{11}^{(0)}} = \frac{4}{3}.$$

Então,

$$\begin{aligned} L_1 &\leftarrow L_1 \\ L_2 &\leftarrow L_2 - m_{21} L_1 \\ L_3 &\leftarrow L_3 - m_{31} L_1 \end{aligned} \quad \text{e} \quad A^{(1)} = \begin{pmatrix} 3 & 2 & 4 \\ 0 & 1/3 & 2/3 \\ 0 & 1/3 & -10/3 \end{pmatrix}.$$

Uma vez que os elementos $a_{21}^{(1)}$ e $a_{31}^{(1)}$ são nulos, podemos guardar os multiplicadores nestas posições, então:

$$A^{(1)} = \begin{pmatrix} 3 & 2 & 4 \\ \hline 1/3 & 1/3 & 2/3 \\ 4/3 & 1/3 & -10/3 \end{pmatrix}.$$

Etapas 2:

Pivô: $a_{22}^{(1)} = 1/3$

$$\text{Multiplicadores: } m_{32} = \frac{a_{32}^{(1)}}{a_{22}^{(1)}} = \frac{1/3}{1/3} = 1$$

Teremos:

$$\begin{aligned} L_1 &\leftarrow L_1 \\ L_2 &\leftarrow L_2 \\ L_3 &\leftarrow L_3 - m_{32} L_2 \end{aligned} \quad \text{e} \quad A^{(2)} = \begin{pmatrix} 3 & 2 & 4 \\ 1/3 & 1/3 & 2/3 \\ 4/3 & 1 & -4 \end{pmatrix}$$

Os fatores L e U são

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

Resolvendo $L(Ux) = b$:

i) $Ly = b$

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

$$y = (1 \quad 5/3 \quad 0)^T$$

ii) $Ux = y$:

$$Ux = y \Rightarrow \begin{cases} 3x_1 + 2x_2 + 4x_3 = 1 \\ 1/3x_2 + 2/3x_3 = 5/3 \\ -4x_3 = 0 \end{cases}$$

$$x = (-3 \ 5 \ 0)^T.$$

FATORAÇÃO LU COM ESTRATÉGIA DE PIVOTEAMENTO PARCIAL

Estudaremos a aplicação da estratégia de pivoteamento parcial à fatoração LU. Esta estratégia requer permutação de linhas na matriz $A^{(k)}$, quando necessário. Por este motivo, veremos inicialmente o que é uma matriz de permutação e, em seguida, como se usa a estratégia de pivoteamento parcial no cálculo dos fatores L e U e quais os efeitos das permutações realizadas na resolução dos sistemas lineares $Ly = b$ e $Ux = y$.

Uma matriz quadrada de ordem n é uma *matriz de permutação* se pode ser obtida da matriz identidade de ordem n permutando-se suas linhas (ou colunas).

Pré-multiplicando-se uma matriz A por uma matriz de permutação P obtém-se a matriz PA com as linhas permutadas e esta permutação de linhas é a mesma efetuada na matriz identidade para se obter P .

Exemplo 6

Sejam

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{e} \quad A = \begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix}.$$

$$PA = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 5 \\ 3 & 1 & 4 \end{pmatrix}.$$

Seja o sistema linear $Ax = b$ e sejam os fatores L e U obtidos pelo processo da Eliminação de Gauss com estratégia de pivoteamento parcial.

L e U são fatores da matriz A' , onde A' é a matriz A com as linhas permutadas, isto é, $A' = PA$

Mas as mesmas permutações efetuadas nas linhas de A devem ser efetuadas sobre o vetor b , uma vez que permutar as linhas de A implica permutar as equações de $Ax = b$.

Seja então $b' = Pb$

O sistema linear $A'x = b'$ é equivalente ao original e, se $A' = LU$, teremos $A'x = b' \Rightarrow PAx = Pb \Rightarrow LUx = Pb$

Resolvemos então os sistemas triangulares:

i) $Ly = Pb$

ii) $Ux = y$ e obtemos a solução do sistema linear original.

Exemplo 7

Seja o sistema linear:

$$\begin{cases} 3x_1 - 4x_2 + x_3 = 9 \\ x_1 + 2x_2 + 2x_3 = 3 \\ 4x_1 - 3x_3 = -2 \end{cases}$$

$$A^{(0)} = \begin{pmatrix} 3 & -4 & 1 \\ 1 & 2 & 2 \\ 4 & 0 & -3 \end{pmatrix}.$$

Etapa 1:

Pivô: $4 = a_{31}^{(0)}$; então devemos permutar as linhas 1 e 3:

$$A'^{(0)} = \begin{pmatrix} 4 & 0 & -3 \\ 1 & 2 & 2 \\ 3 & -4 & 1 \end{pmatrix}, \quad P^{(0)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ e } A'^{(0)} = P^{(0)}A^{(0)}$$

Efetuada a eliminação em $A'^{(0)}$:

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

Etapa 2:

Pivô: $-4 = a_{32}^{(1)}$; então devemos permutar as linhas 2 e 3:

$$A'^{(1)} = \begin{pmatrix} 4 & 0 & -3 \\ \boxed{3/4} & -4 & 13/4 \\ 1/4 & 2 & 11/4 \end{pmatrix}, \quad P^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ e } A'^{(1)} = P^{(1)}A^{(1)}$$

Efetuada a eliminação temos:

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

Os fatores L e U são

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 3/4 & 1 & 0 \\ 1/4 & -1/2 & 1 \end{pmatrix} \quad \text{e} \quad U = \begin{pmatrix} 4 & 0 & -3 \\ 0 & -4 & 13/4 \\ 0 & 0 & 35/8 \end{pmatrix}$$

e estes são os fatores da matriz $A' = PA$ onde $P = P^{(1)} P^{(0)}$, isto é:

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

Resolução dos sistemas lineares triangulares:

i) $Ly = Pb$ onde

$$Pb = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 9 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -2 \\ 9 \\ 3 \end{pmatrix}$$

$$\begin{cases} y_1 & = -2 \\ 3/4y_1 + y_2 & = 9 \\ 1/4y_1 - 1/2y_2 + y_3 & = 3 \end{cases} \Rightarrow y = \begin{pmatrix} -2 \\ 21/2 \\ 35/4 \end{pmatrix}$$

ii) $Ux = y$

$$\begin{cases} 4x_1 + 0x_2 - 3x_3 = -2 \\ -4x_2 + 13/4x_3 = 21/2 \\ 35/8x_3 = 35/4 \end{cases} \Rightarrow x = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Considerando uma matriz geral, A : $n \times n$. Se A é não singular, então no início da etapa k da fase de eliminação existe pelo menos um elemento não nulo entre os elementos $a_{kk}^{(k-1)}, \dots, a_{nk}^{(k-1)}$ de modo que através de uma troca de linhas sobre $A^{(k-1)}$ é sempre possível obter a matriz $A^{(k-1)}$ com elemento não nulo na posição (k, k) . Desta forma, os cálculos necessários em cada etapa da eliminação podem ser realizados e os fatores L e U da matriz PA serão unicamente determinados, onde $P = P^{(n-1)} P^{(n-2)} \dots P^{(0)}$ e $P^{(k)}$ representa a troca de linhas efetuada na etapa k .

As permutações de linha realizadas durante a fatoração podem ser representadas através de um vetor $n \times 1$, que denotaremos por p , definido por $p(k) = i$ se na etapa k a linha i da matriz original $A^{(0)}$ for a linha pivotada.

Considerando o Exemplo 7, teríamos inicialmente: $p = (1 \ 2 \ 3)$. No início da etapa 1, a linha 3 é a pivotada, então $p = (3 \ 2 \ 1)$. No início da etapa 2, a linha 3 da matriz $A^{(1)}$ é a linha pivotada, então $p = (3 \ 1 \ 2)$.

ALGORITMO 3: Resolução de $Ax = b$ através da fatoração LU com pivoteamento parcial

Considere o sistema linear $Ax = b$, A : $n \times n$; o vetor p representará as permutações realizadas durante a fatoração.

(Cálculo dos fatores:)

Para $i = 1, \dots, n$

$$\left[\begin{array}{l} p(i) = i \end{array} \right.$$

Para $k = 1, \dots, (n - 1)$

$$\left[\begin{array}{l} pv = |a(k, k)| \\ r = k \\ \text{Para } i = (k + 1), \dots, n \\ \left[\begin{array}{l} \text{se } (|a(i, k)| > pv), \text{ faça:} \\ \left[\begin{array}{l} pv = |a(i, k)| \\ r = i \end{array} \right. \end{array} \right. \end{array} \right.$$

se $pv = 0$, parar; a matriz A é singular

se $r \neq k$, faça:

$$\left[\begin{array}{l} aux = p(k) \\ p(k) = p(r) \\ p(r) = aux \\ \text{Para } j = 1, \dots, n \\ \left[\begin{array}{l} aux = a(k, j) \\ a(k, j) = a(r, j) \\ a(r, j) = aux \end{array} \right. \end{array} \right.$$

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

$$\left[\begin{array}{l} m = a(i, k)/a(k, k) \\ a(i, k) = m \\ \text{para } j = (k + 1), \dots, n \\ a(i, j) = a(i, j) - ma(k, j) \end{array} \right.$$

(Resolução dos sistemas triangulares)

$$c = Pb \quad \left[\begin{array}{l} \text{Para } i = 1, \dots, n \\ r = p(i) \\ c(i) = b(r) \end{array} \right.$$

$$Ly = c \begin{cases} \text{Para } i = 1, \dots, n \\ \text{soma} = 0 \\ \text{Para } j = 1, \dots, (i - 1) \\ [\text{soma} = \text{soma} + a(i, j)y(j) \\ y(i) = c(i) - \text{soma} \end{cases}$$

$$Ux = y \begin{cases} \text{Para } i = n, (n - 1), \dots, 1 \\ \text{soma} = 0 \\ \text{Para } j = (i + 1), \dots, n \\ [\text{soma} = \text{soma} + a(i, j)x(j) \\ x(i) = (y(i) - \text{soma})/a(i, i) \end{cases}$$

3.2.4 FATORAÇÃO DE CHOLESKY

Uma matriz $A: n \times n$ é *definida positiva* se $x^T Ax > 0$ para todo $x \in \mathbb{R}^n, x \neq 0$.

A resolução de sistemas lineares em que a matriz A é simétrica, definida positiva, é freqüente em problemas práticos e tais matrizes podem ser fatoradas na forma:

$$A = GG^T$$

onde $G: n \times n$ é uma matriz triangular inferior com elementos da diagonal estritamente positivos. Esta fatoração é conhecida como *fatoração de Cholesky*.

Seja $A: n \times n$ e vamos supor que A satisfaça as hipóteses do Teorema 2. Então, A pode ser fatorada, de forma única, como $LD\bar{U}$ (ver Exercício 12) com:

$L: n \times n$, triangular inferior com diagonal unitária;

$D: n \times n$, diagonal e

$\bar{U}: n \times n$, triangular superior com diagonal unitária.

Se, além das hipóteses do Teorema 2, a matriz for simétrica, demonstra-se [14] que $\bar{U} = L^T$, e, então, a fatoração fica: $A = LDL^T$.

Exemplo 8

Considere a matriz

$$A = \begin{pmatrix} 16 & -4 & 12 & -4 \\ -4 & 2 & -1 & 1 \\ 12 & -1 & 14 & -2 \\ -4 & 1 & -2 & 83 \end{pmatrix}$$

Calculando os fatores L e U de A e, em seguida, os fatores L, D e \bar{U} , teremos:

$$\begin{pmatrix} 16 & -4 & 12 & -4 \\ -4 & 2 & -1 & 1 \\ 12 & -1 & 14 & -2 \\ -4 & 1 & -2 & 83 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1/4 & 1 & 0 & 0 \\ 3/4 & 2 & 1 & 0 \\ -1/4 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 16 & -4 & 12 & -4 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 81 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1/4 & 1 & 0 & 0 \\ 3/4 & 2 & 1 & 0 \\ -1/4 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 16 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 81 \end{pmatrix} \begin{pmatrix} 1 & -1/4 & 3/4 & -1/4 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Observamos que:

- i) $\bar{u}_{ij} = u_{ij} / u_{ii}$;
- ii) como a matriz A é simétrica, $\bar{U} = L^T$.

Se A for definida positiva, os elementos da matriz D são estritamente positivos, conforme demonstramos a seguir: como A é definida positiva, temos que para qualquer $x \in \mathbb{R}^n$, $x \neq 0$, $x^T A x > 0$. Usando a fatoração LDL^T de A , temos:

$$0 < x^T A x = x^T (LDL^T) x = y^T D y.$$

Agora, $y = L^T x$ e L tem posto completo. Então, $y \neq 0$ pois x é não nulo e, para cada $y \in \mathbb{R}^n$, existe $x \in \mathbb{R}^n$, tal que $y = L^T x$.

Fazendo $y = e_i$, $i = 1, \dots, n$, teremos: $e_i^T D e_i = d_{ii}$, e, como $y^T D y > 0$, qualquer $y \neq 0$, obtemos: $d_{ii} > 0$, $i = 1, \dots, n$.

Concluindo, se A for simétrica definida positiva, então A pode ser fatorada na forma LDL^T com L triangular inferior com diagonal unitária e D matriz diagonal com elementos na diagonal estritamente positivos.

Podemos escrever então:

$$A = LDL^T = L\bar{D}\bar{D}L^T$$

onde $\bar{d}_{ii} = \sqrt{d_{ii}}$

e, se $G = L\bar{D}$, obtemos $A = GG^T$ com G triangular inferior com diagonal estritamente positiva.

Formalizamos este resultado no Teorema 3.

TEOREMA 3: (Fatoração de Cholesky)

Se $A: n \times n$ é simétrica e definida positiva, então existe uma única matriz triangular inferior $G: n \times n$ com diagonal positiva, tal que $A = GG^T$.

Exemplo 9

Retomando a matriz A do Exemplo 8 e sua fatoração LDL^T , observamos que o fator D é tal que $d_{ii} > 0$, $i = 1, \dots, 4$.

Fazendo $\bar{D} = D^{1/2}$, teremos:

$$A = LDL^T = L\bar{D}\bar{D}L^T = (L\bar{D})(\bar{D}L^T) = GG^T$$

$$\text{onde } \bar{D} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix} \text{ e}$$

$$G = \begin{pmatrix} 4 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ -1 & 0 & 1 & 9 \end{pmatrix}.$$

A matriz G , triangular inferior com diagonal positiva, é o fator de Cholesky da matriz A .

Neste exemplo, o fator de Cholesky foi obtido a partir da fatoração LDL^T , que por sua vez foi obtida a partir da fatoração LU . No entanto, o fator de Cholesky deve ser calculado através da equação matricial $A = GG^T$, uma vez que, assim, os cálculos envolvidos serão reduzidos pela metade.

Cálculo do fator de Cholesky:

É dada A : $n \times n$, matriz simétrica e definida positiva:

$$A = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{21} & a_{22} & \cdots & a_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

O fator G : $n \times n$ triangular inferior com diagonal positiva será obtido a partir da equação matricial:

$$A = GG^T$$

$$\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{21} & a_{22} & \cdots & a_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} g_{11} & & & \\ g_{21} & g_{22} & & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{pmatrix} \begin{pmatrix} g_{11} & g_{21} & \cdots & g_{n1} \\ & g_{22} & \cdots & g_{n2} \\ & & \cdot & \\ & & & \cdot \\ & & & & g_{nn} \end{pmatrix}.$$

O cálculo será realizado por colunas:

coluna 1:

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ a_{n1} \end{pmatrix} = G \begin{pmatrix} g_{11} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} = \begin{pmatrix} g_{11}^2 \\ g_{21}g_{11} \\ \cdot \\ \cdot \\ g_{n1}g_{11} \end{pmatrix};$$

então: $g_{11} = \sqrt{a_{11}}$

e $g_{j1} = a_{j1}/g_{11}$, $j = 2, \dots, n$;

coluna 2:

$$\begin{pmatrix} a_{21} \\ a_{22} \\ a_{32} \\ \cdot \\ a_{n2} \end{pmatrix} = G \begin{pmatrix} g_{21} \\ g_{22} \\ 0 \\ \cdot \\ 0 \end{pmatrix} = \begin{pmatrix} g_{11}g_{21} \\ g_{21}^2 + g_{22}^2 \\ g_{31}g_{21} + g_{32}g_{22} \\ \cdot \\ g_{n1}g_{21} + g_{n2}g_{22} \end{pmatrix};$$

$$\text{então: } g_{21}^2 + g_{22}^2 = a_{22} \Rightarrow g_{22} = \sqrt{a_{22} - g_{21}^2}$$

$$\text{e } g_{j1}g_{21} + g_{j2}g_{22} = a_{j2}, \quad j = 3, \dots, n.$$

Os elementos g_{j1} já estão calculados; assim,

$$g_{j2} = (a_{j2} - g_{j1}g_{21}) / g_{22}, \quad j = 3, \dots, n.$$

Coluna k :

Para obter os elementos da coluna k de G : $(0 \dots g_{kk} \ g_{k+1k} \dots g_{nk})^T$, $k = 3, \dots, n$, usamos a equação matricial:

$$\begin{pmatrix} a_{k1} \\ a_{k2} \\ \cdot \\ \cdot \\ a_{kk} \\ a_{k+1k} \\ \cdot \\ \cdot \\ a_{nk} \end{pmatrix} = G \begin{pmatrix} g_{k1} \\ g_{k2} \\ \cdot \\ \cdot \\ g_{kk} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

e teremos:

$$a_{kk} = g_{k1}^2 + g_{k2}^2 + \dots + g_{kk}^2 \text{ e daí}$$

$$g_{kk} = \left(a_{kk} - \sum_{i=1}^{k-1} g_{ki}^2 \right)^{1/2}$$

$$\text{e } a_{jk} = g_{j1}g_{k1} + g_{j2}g_{k2} + \dots + g_{jk}g_{kk}, \quad j = (k+1), \dots, n$$

Como todos os elementos g_{ik} , $i = 1, \dots, (k-1)$ já estão calculados, teremos:

$$g_{jk} = \left(a_{jk} - \sum_{i=1}^{k-1} g_{ji}g_{ki} \right) / g_{kk} \quad j = (k+1), \dots, n.$$

ALGORITMO 4: Fatoração de Cholesky

Seja A : $n \times n$, simétrica definida positiva:

```

Para  $k = 1, \dots, n$ 
  soma = 0
  Para  $j = 1, \dots, (k - 1)$ 
    [ soma = soma +  $g_{kj}^2$ 
   $r = a_{kk} - \text{soma}$ 
   $g_{kk} = (r)^{1/2}$ 
  Para  $i = (k + 1), \dots, n$ 
    [ soma = 0
    Para  $j = 1, \dots, (k - 1)$ 
      [ soma = soma +  $g_{ij}g_{kj}$ 
     $g_{ik} = (a_{ik} - \text{soma}) / g_{kk}$ 

```

Na prática, aplicamos a fatoração de Cholesky para verificar se uma determinada matriz A simétrica é definida positiva. Se o algoritmo falhar, isto é, se em alguma etapa tivermos $r \leq 0$, o processo será interrompido e, conseqüentemente, a matriz original não é definida positiva; caso contrário, ao final teremos $A = GG^T$ com o fator conforme descrito no Teorema 3. Demonstra-se (Exercício 21) que uma matriz na forma BB^T é definida positiva, se B tem posto completo.

A fatoração de Cholesky requer cerca de $n^3/3$ operações de multiplicação e adição no cálculo dos fatores, aproximadamente a metade do número de operações necessárias na fase da eliminação da fatoração LU.

Observamos que alguns autores contam uma adição e uma multiplicação como uma operação apenas; assim, para esses autores, a fatoração LU realiza cerca de $n^3/3$ operações e a fatoração de Cholesky, $n^3/6$.

Obtido o fator G , a resolução do sistema linear $Ax = b$ prossegue com a resolução dos sistemas triangulares:

$$Ax = b \Leftrightarrow (GG^T)x = b \Rightarrow \begin{cases} i) Gy = b \\ ii) G^T x = y \end{cases}$$

3.3 MÉTODOS ITERATIVOS

3.3.1 INTRODUÇÃO

A idéia central dos métodos iterativos é generalizar o método do ponto fixo utilizado na busca de raízes de uma equação que foi visto no Capítulo 2.

Seja o sistema linear $Ax = b$, onde:

A : matriz dos coeficientes, $n \times n$;

x : vetor das variáveis, $n \times 1$;

b : vetor dos termos constantes, $n \times 1$.

Este sistema é convertido, de alguma forma, num sistema do tipo $x = Cx + g$ onde C é matriz $n \times n$ e g vetor $n \times 1$. Observamos que $\varphi(x) = Cx + g$ é uma função de iteração dada na forma matricial.

É então proposto o esquema iterativo:

Partimos de $x^{(0)}$ (vetor aproximação inicial) e então construímos consecutivamente os vetores:

$$x^{(1)} = Cx^{(0)} + g = \varphi(x^{(0)}), \quad (\text{primeira aproximação}),$$

$$x^{(2)} = Cx^{(1)} + g = \varphi(x^{(1)}), \quad (\text{segunda aproximação}) \text{ etc.}$$

De um modo geral, a aproximação $x^{(k+1)}$ é calculada pela fórmula $x^{(k+1)} = Cx^{(k)} + g$, ou seja, $x^{(k+1)} = \varphi(x^{(k)})$, $k = 0, 1, \dots$

É importante observar que se a seqüência de aproximações $x^{(0)}, x^{(1)}, \dots, x^{(k)}, \dots$ é tal que, $\lim_{k \rightarrow \infty} x^{(k)} = \alpha$, então $\alpha = C\alpha + g$, ou seja, α é solução do sistema linear $Ax = b$.

3.3.2 TESTES DE PARADA

O processo iterativo é repetido até que o vetor $x^{(k)}$ esteja suficientemente próximo do vetor $x^{(k-1)}$.

$$\text{Medimos a distância entre } x^{(k)} \text{ e } x^{(k-1)} \text{ por } d^{(k)} = \max_{1 \leq i \leq n} |x_i^{(k)} - x_i^{(k-1)}|.$$

Assim, dada uma precisão ε , o vetor $x^{(k)}$ será escolhido como \bar{x} , solução aproximada da solução exata, se $d^{(k)} < \varepsilon$.

Da mesma maneira que no teste de parada dos métodos iterativos para zeros de funções, podemos efetuar aqui o teste do erro relativo:

$$d_r^{(k)} = \frac{d^{(k)}}{\max_{1 \leq i \leq n} |x_i^{(k)}|}.$$

Computacionalmente usamos também como teste de parada um número máximo de iterações.

3.3.3 MÉTODO ITERATIVO DE GAUSS-JACOBI

A forma como o método de Gauss-Jacobi transforma o sistema linear $Ax = b$ em $x = Cx + g$ é a seguinte:

Tomamos o sistema original:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \\ \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \quad \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases}$$

e supondo $a_{ii} \neq 0$, $i = 1, \dots, n$, isolamos o vetor x mediante a separação pela diagonal, assim:

$$\begin{cases} x_1 = \frac{1}{a_{11}} (b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n) \\ x_2 = \frac{1}{a_{22}} (b_2 - a_{21}x_1 - a_{23}x_3 - \dots - a_{2n}x_n) \\ \vdots \\ \vdots \\ \vdots \\ x_n = \frac{1}{a_{nn}} (b_n - a_{n1}x_1 - a_{n2}x_2 - \dots - a_{n,n-1}x_{n-1}). \end{cases}$$

Desta forma, temos $x = Cx + g$, onde

$$C = \begin{pmatrix} 0 & -a_{12}/a_{11} & -a_{13}/a_{11} & \dots & -a_{1n}/a_{11} \\ -a_{21}/a_{22} & 0 & -a_{23}/a_{22} & \dots & -a_{2n}/a_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_{n1}/a_{nn} & -a_{n2}/a_{nn} & -a_{n3}/a_{nn} & \dots & 0 \end{pmatrix}$$

e

$$g = \begin{pmatrix} b_1/a_{11} \\ b_2/a_{22} \\ \vdots \\ \vdots \\ b_n/a_{nn} \end{pmatrix}.$$

O método de Gauss-Jacobi consiste em, dado $x^{(0)}$, aproximação inicial, obter $x^{(1)}$..., $x^{(k)}$... através da relação recursiva $x^{(k+1)} = Cx^{(k)} + g$:

$$\begin{cases} x_1^{(k+1)} = \frac{1}{a_{11}} (b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \dots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} = \frac{1}{a_{22}} (b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \dots - a_{2n}x_n^{(k)}) \\ \vdots \\ x_n^{(k+1)} = \frac{1}{a_{nn}} (b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \dots - a_{n,n-1}x_{n-1}^{(k)}) \end{cases}$$

Exemplo 10

Resolva o sistema linear:

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

pelo método de Gauss-Jacobi com $x^{(0)} = \begin{pmatrix} 0.7 \\ -1.6 \\ 0.6 \end{pmatrix}$ e $\varepsilon = 0.05$.

O processo iterativo é

$$\begin{cases} x_1^{(k+1)} = \frac{1}{10} (7 - 2x_2^{(k)} - x_3^{(k)}) = 0x_1^{(k)} - \frac{2}{10} x_2^{(k)} - \frac{1}{10} x_3^{(k)} + \frac{7}{10} \\ x_2^{(k+1)} = \frac{1}{5} (-8 - x_1^{(k)} - x_3^{(k)}) = -\frac{1}{5} x_1^{(k)} + 0x_2^{(k)} - \frac{1}{5} x_3^{(k)} - \frac{8}{5} \\ x_3^{(k+1)} = \frac{1}{10} (6 - 2x_1^{(k)} - 3x_2^{(k)}) = -\frac{2}{10} x_1^{(k)} - \frac{3}{10} x_2^{(k)} + 0x_3^{(k)} + \frac{6}{10} \end{cases}$$

Na forma matricial $x^{(k+1)} = Cx^{(k)} + g$ temos

$$C = \begin{pmatrix} 0 & -2/10 & -1/10 \\ -1/5 & 0 & -1/5 \\ -1/5 & -3/10 & 0 \end{pmatrix} \text{ e } g = \begin{pmatrix} 7/10 \\ -8/5 \\ 6/10 \end{pmatrix}.$$

Assim ($k = 0$) temos

$$\begin{cases} x_1^{(1)} = -0.2x_2^{(0)} - 0.1x_3^{(0)} + 0.7 = -0.2(-1.6) - 0.1 \times 0.6 + 0.7 = 0.96 \\ x_2^{(1)} = -0.2x_1^{(0)} - 0.2x_3^{(0)} - 1.6 = -0.2 \times 0.7 - 0.2 \times 0.6 - 1.6 = -1.86 \\ x_3^{(1)} = -0.2x_1^{(0)} - 0.3x_2^{(0)} + 0.6 = -0.2 \times 0.7 - 0.3(-1.6) + 0.6 = 0.94 \end{cases}$$

ou

$$x^{(1)} = Cx^{(0)} + g = \begin{pmatrix} 0.96 \\ -1.86 \\ 0.94 \end{pmatrix}.$$

Calculando $d_r^{(1)}$, temos:

$$|x_1^{(1)} - x_1^{(0)}| = 0.26$$

$$|x_2^{(1)} - x_2^{(0)}| = 0.26 \quad \Rightarrow \quad d_r^{(1)} = \frac{0.34}{\max_{1 \leq i \leq 3} |x_i^{(1)}|} = \frac{0.34}{1.86} = 0.1828 > \epsilon$$

$$|x_3^{(1)} - x_3^{(0)}| = 0.34$$

Prosseguindo as iterações, temos:

para $k = 1$:

$$x^{(2)} = \begin{pmatrix} 0.978 \\ -1.98 \\ 0.966 \end{pmatrix} \Rightarrow d_r^{(2)} = \frac{0.12}{1.98} = 0.0606 > \varepsilon$$

e para $k = 2$:

$$x^{(3)} = \begin{pmatrix} 0.9994 \\ -1.9888 \\ 0.9984 \end{pmatrix} \Rightarrow d_r^{(3)} = \frac{0.0324}{1.9888} = 0.0163 < \varepsilon.$$

Então, a solução \bar{x} do sistema linear acima, com erro menor que 0.05, obtida pelo método de Gauss-Jacobi, é

$$\bar{x} = x^{(3)} = \begin{pmatrix} 0.9994 \\ -1.9888 \\ 0.9984 \end{pmatrix}.$$

Neste exemplo tomamos $x^{(0)} = \begin{pmatrix} 0.7 \\ -1.6 \\ 0.6 \end{pmatrix} = \begin{pmatrix} b_1/a_{11} \\ b_2/a_{22} \\ b_3/a_{33} \end{pmatrix}$. No entanto, o valor de

$x^{(0)}$ é arbitrário, pois veremos mais adiante que a convergência ou não de um método iterativo para a solução de um sistema linear de equações é independente da aproximação inicial escolhida.

UM CRITÉRIO DE CONVERGÊNCIA

Daremos aqui um teorema que estabelece uma condição suficiente para a convergência do método iterativo de Gauss-Jacobi.

TEOREMA 4: (Critério das linhas)

Seja o sistema linear $Ax = b$ e seja $\alpha_k = \left(\sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \right) / |a_{kk}|$. Se $\alpha = \max_{1 \leq k \leq n} \alpha_k < 1$, então o

método de Gauss-Jacobi gera uma seqüência $\{x^{(k)}\}$ convergente para a solução do sistema dado, independentemente da escolha da aproximação inicial, $x^{(0)}$.

A demonstração deste teorema pode ser encontrada na referência [30], Capítulo 9.

Exemplo 11

Analisando a matriz A do sistema linear do Exemplo 10,

$$A = \begin{pmatrix} 10 & 2 & 1 \\ 1 & 5 & 1 \\ 2 & 3 & 10 \end{pmatrix}, \text{ temos}$$

$$\alpha_1 = \frac{2+1}{10} = \frac{3}{10} = 0.3 < 1; \alpha_2 = \frac{1+1}{5} = 0.4 < 1; \alpha_3 = \frac{2+3}{10} = 0.5 < 1 \text{ e}$$

então $\max_{1 \leq k \leq 3} \alpha_k = 0.5 < 1$ donde, pelo critério das linhas, temos garantia de convergência

para o método de Gauss-Jacobi.

Exemplo 12

Para o sistema linear $\begin{cases} x_1 + x_2 = 3 \\ x_1 - 3x_2 = -3 \end{cases}$ o método de Gauss-Jacobi gera uma seqüência

convergente para a solução exata $x^* = \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix}$. (Verifique!) No entanto, o critério das

linhas não é satisfeito, visto que $\alpha_1 = \frac{1}{1} = 1$. Isto mostra que a condição do Teorema 4 é apenas suficiente.

Exemplo 13

A matriz A do sistema linear $\begin{cases} x_1 + 3x_2 + x_3 = -2 \\ 5x_1 + 2x_2 + 2x_3 = 3 \\ 6x_2 + 8x_3 = -6 \end{cases}$ não satisfaz o critério das linhas

pois $\alpha_1 = \frac{3 + 1}{1} = 4 > 1$. Contudo, se permutarmos a primeira equação com a segunda,

temos o sistema linear $\begin{cases} 5x_1 + 2x_2 + 2x_3 = 3 \\ x_1 + 3x_2 + x_3 = -2 \\ 6x_2 + 8x_3 = -6 \end{cases}$ que é equivalente ao sistema original e a

matriz $\begin{pmatrix} 5 & 2 & 2 \\ 1 & 3 & 1 \\ 0 & 6 & 8 \end{pmatrix}$ deste novo sistema satisfaz o critério das linhas.

Assim, é conveniente aplicarmos o método de Gauss-Jacobi a esta nova disposição do sistema, pois desta forma a convergência está assegurada.

Concluindo, sempre que o critério das linhas não for satisfeito, devemos tentar uma permutação de linhas e/ou colunas de forma a obtermos uma disposição para a qual a matriz dos coeficientes satisfaça o critério das linhas. No entanto, nem sempre é possível obter tal disposição, como facilmente verificamos com o sistema linear do Exemplo 12.

3.3.4 MÉTODO ITERATIVO DE GAUSS-SEIDEL

Da mesma forma que no método de Gauss-Jacobi, no método de Gauss-Seidel o sistema linear $Ax = b$ é escrito na forma equivalente $x = Cx + g$ por separação da diagonal.

O processo iterativo consiste em, sendo $x^{(0)}$ uma aproximação inicial, calcular $x^{(1)}, x^{(2)}, \dots, x^{(k)}, \dots$ por:

$$\left\{ \begin{array}{l} x_1^{(k+1)} = \frac{1}{a_{11}} (b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \dots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} = \frac{1}{a_{22}} (b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \dots - a_{2n}x_n^{(k)}) \\ x_3^{(k+1)} = \frac{1}{a_{33}} (b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} - a_{34}x_4^{(k)} - \dots - a_{3n}x_n^{(k)}) \\ \vdots \\ x_n^{(k+1)} = \frac{1}{a_{nn}} (b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - \dots - a_{n,n-1}x_{n-1}^{(k+1)}) \end{array} \right.$$

Portanto, no processo iterativo de Gauss-Seidel, no momento de se calcular $x_j^{(k+1)}$ usamos todos os valores $x_1^{(k+1)}, \dots, x_{j-1}^{(k+1)}$ que já foram calculados e os valores $x_{j+1}^{(k)}, \dots, x_n^{(k)}$ restantes.

Exemplo 14

Resolva o sistema linear:

$$\left\{ \begin{array}{l} 5x_1 + x_2 + x_3 = 5 \\ 3x_1 + 4x_2 + x_3 = 6 \\ 3x_1 + 3x_2 + 6x_3 = 0 \end{array} \right.$$

pelo método de Gauss-Seidel com $x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ e $\epsilon = 5 \times 10^{-2}$.

O processo iterativo é:

$$\left\{ \begin{array}{l} x_1^{(k+1)} = 1 - 0.2x_2^{(k)} - 0.2x_3^{(k)} \\ x_2^{(k+1)} = 1.5 - 0.75x_1^{(k+1)} - 0.25x_3^{(k)} \\ x_3^{(k+1)} = 0 - 0.5x_1^{(k+1)} - 0.5x_2^{(k+1)}. \end{array} \right.$$

$$\text{Como } \mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

($k = 0$):

$$\begin{cases} x_1^{(1)} = 1 - 0 - 0 = 1 \\ x_2^{(1)} = 1.5 - 0.75 \times 1 - 0 = 0.75 \\ x_3^{(1)} = -0.5 \times 1 - 0.5 \times 0.75 = -0.875 \end{cases} \Rightarrow \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 0.75 \\ -0.875 \end{pmatrix}, \text{ donde}$$

$$|x_1^{(1)} - x_1^{(0)}| = 1$$

$$|x_2^{(1)} - x_2^{(0)}| = 0.75 \quad \Rightarrow d_r^{(1)} = \frac{1}{\max_{1 \leq i \leq 3} |x_i^{(1)}|} = 1 > \varepsilon$$

$$|x_3^{(1)} - x_3^{(0)}| = 0.875.$$

Assim, ($k = 1$) e

$$\begin{cases} x_1^{(2)} = 1 - 0.2 \times 0.75 + 0.2 \times 0.875 = 1.025 \\ x_2^{(2)} = 1.5 - 0.75 \times 1.025 - 0.25 \times (-0.875) = 0.95 \\ x_3^{(2)} = -0.5 \times 1.025 - 0.5 \times 0.95 = -0.9875 \end{cases}$$

$$\Rightarrow \mathbf{x}^{(2)} = \begin{pmatrix} 1.025 \\ 0.95 \\ -0.9875 \end{pmatrix}, \text{ donde}$$

$$|x_1^{(2)} - x_1^{(1)}| = 0.025$$

$$|x_2^{(2)} - x_2^{(1)}| = 0.20 \quad \Rightarrow d_r^{(2)} = \frac{0.2}{\max_{1 \leq i \leq 3} |x_i^{(2)}|} = \frac{0.2}{1.025} = 0.1951 > \varepsilon$$

$$|x_3^{(2)} - x_3^{(1)}| = 0.1125$$

Continuando as iterações obtemos:

$$\mathbf{x}^{(3)} = \begin{pmatrix} 1.0075 \\ 0.9912 \\ -0.9993 \end{pmatrix} \Rightarrow d_r^{(3)} = 0.0409 < \epsilon.$$

Assim, a solução $\bar{\mathbf{x}}$ do sistema linear dado com erro menor que ϵ , pelo método de Gauss-Seidel, é

$$\bar{\mathbf{x}} = \mathbf{x}^{(3)} = \begin{pmatrix} 1.0075 \\ 0.9912 \\ -0.9993 \end{pmatrix}.$$

O esquema iterativo do método de Gauss-Seidel pode ser escrito na forma matricial da seguinte maneira:

Inicialmente escrevemos a matriz A , dos coeficientes, como $A = L + D + R$, onde:

L : matriz triangular inferior com diagonal nula;

D : matriz diagonal com $d_{ii} \neq 0$, $i = 1, \dots, n$;

R : matriz triangular superior com diagonal nula.

O modo mais simples de se escrever A nesta forma é

$$L = \begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ a_{31} & a_{32} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{pmatrix}, \quad D = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{pmatrix} \quad e$$

$$R = \begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & 0 & a_{23} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Portanto, $Ax = b \Leftrightarrow (L + D + R)x = b \Leftrightarrow Dx = b - Lx - Rx \Leftrightarrow$

$$\Leftrightarrow x = D^{-1}b - D^{-1}Lx - D^{-1}Rx.$$

No método de Gauss-Seidel o vetor $x^{(k+1)}$ é calculado por:

$$x^{(k+1)} = D^{-1}b - D^{-1}Lx^{(k+1)} - D^{-1}Rx^{(k)}.$$

Agora, podemos ainda escrever $x^{(k+1)} = Cx^{(k)} + g$, considerando que $A = D(L_1 + I + R_1)$ onde:

$$L_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \frac{a_{21}}{a_{22}} & 0 & 0 & \dots & 0 \\ \frac{a_{31}}{a_{33}} & \frac{a_{32}}{a_{33}} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{a_{n1}}{a_{nn}} & \frac{a_{n2}}{a_{nn}} & \frac{a_{n3}}{a_{nn}} & \dots & 0 \\ \frac{a_{nn}}{a_{nn}} & \frac{a_{nn}}{a_{nn}} & \frac{a_{nn}}{a_{nn}} & \dots & 0 \end{pmatrix}$$

$$R_1 = \begin{pmatrix} 0 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ 0 & 0 & \frac{a_{23}}{a_{22}} & \dots & \frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

então $Ax = b \Leftrightarrow$

$$D(L_1 + I + R_1) x = b \Leftrightarrow$$

$$(L_1 + I + R_1) x = D^{-1} b \Leftrightarrow$$

$x = -L_1 x - R_1 x + D^{-1} b$ e o método de Gauss-Seidel é

$$x^{(k+1)} = -L_1 x^{(k+1)} - R_1 x^{(k)} + D^{-1} b,$$

donde $(I + L_1) x^{(k+1)} = -R_1 x^{(k)} + D^{-1} b$

$$\text{ou } x^{(k+1)} = \underbrace{-(I + L_1)^{-1} R_1}_{C} x^{(k)} + \underbrace{(I + L_1)^{-1} D^{-1} b}_{g} = Cx^{(k)} + g$$

INTERPRETAÇÃO GEOMÉTRICA NO CASO 2 x 2

Consideremos a aplicação geométrica dos métodos de Gauss-Jacobi e Gauss-Seidel ao sistema linear:

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

Preparação:

$$\begin{cases} x_1 = 3 - x_2 \\ x_2 = \frac{1}{3} (3 + x_1). \end{cases}$$

O esquema iterativo para Gauss-Jacobi é:

$$\begin{cases} x_1^{(k+1)} = 3 - x_2^{(k)} \\ x_2^{(k+1)} = \frac{1}{3} (3 + x_1^{(k)}) \end{cases}$$

Teremos:

$$x^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad x^{(1)} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}; \quad x^{(2)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad x^{(3)} = \begin{pmatrix} 1 \\ 5/3 \end{pmatrix}; \quad x^{(4)} = \begin{pmatrix} 4/3 \\ 4/3 \end{pmatrix}$$

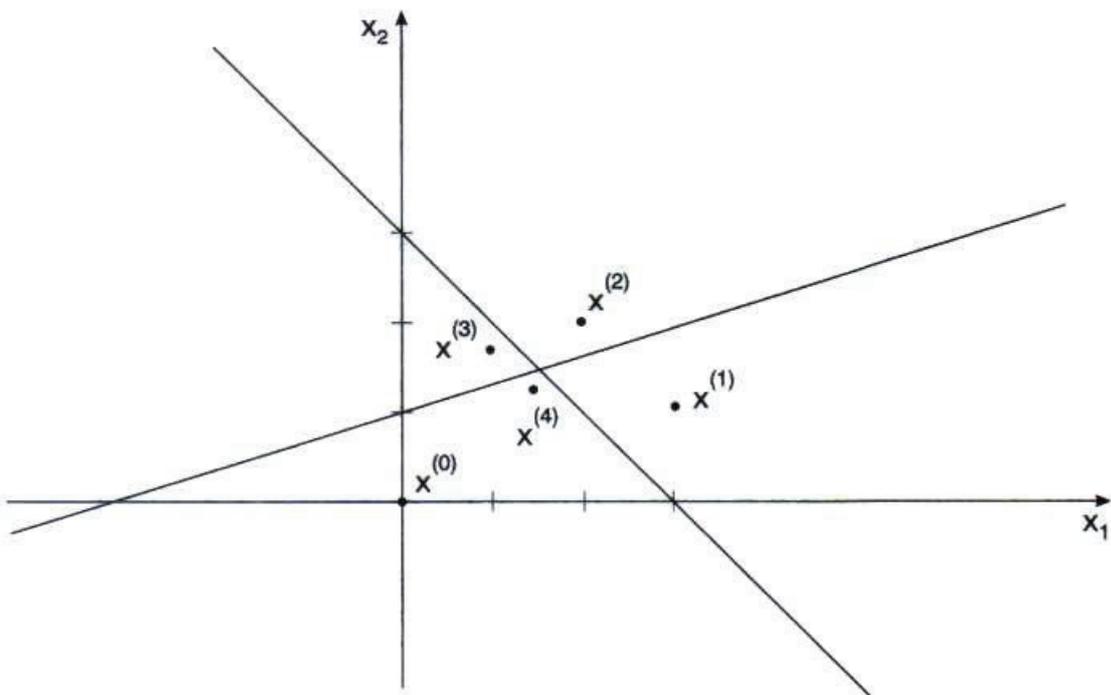


Figura 3.10

O esquema iterativo para Gauss-Seidel é:

$$\begin{cases} x_1^{(k+1)} = 3 - x_2^{(k)} \\ x_2^{(k+1)} = \frac{1}{3} (3 + x_1^{(k+1)}) \end{cases}$$

Para melhor visualização gráfica, marcaremos no gráfico os pontos $(x_1^{(k)}, x_2^{(k)}); (x_1^{(k+1)}, x_2^{(k+1)}), \dots$ para $k = 0, 1, 2, \dots$

$$\begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(1)} \\ x_2^{(0)} \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(2)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 1 \\ 4/3 \end{pmatrix}$$

$$\begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 1 \\ 4/3 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(3)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 5/3 \\ 4/3 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(3)} \\ x_2^{(3)} \end{pmatrix} = \begin{pmatrix} 5/3 \\ 14/9 \end{pmatrix}, \dots$$

Observamos que os pontos $(x_1^{(k+1)}, x_2^{(k)})$ satisfazem a primeira equação e os pontos $(x_1^{(k+1)}, x_2^{(k+1)})$ satisfazem a segunda equação.

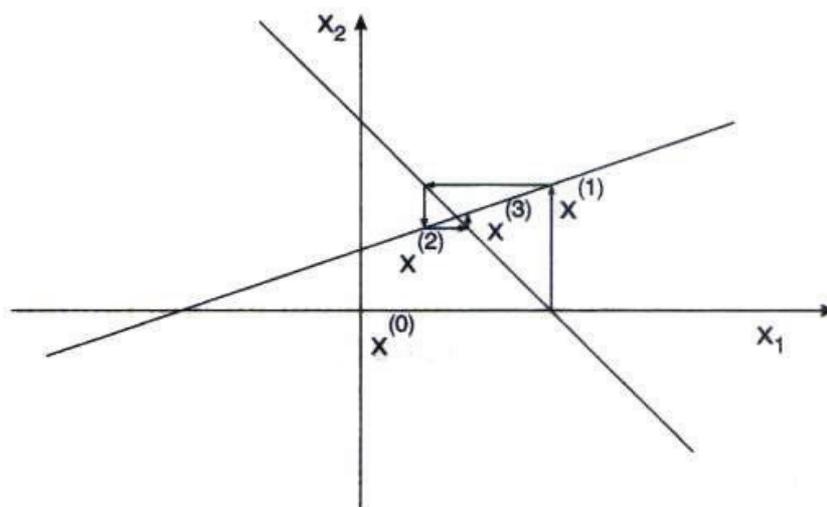


Figura 3.11

Embora a ordem das equações num sistema linear não mude a solução exata, as seqüências geradas pelos métodos de Gauss-Seidel e de Gauss-Jacobi dependem fundamentalmente da disposição das equações.

É fácil verificar que a seqüência $x^{(0)}, x^{(1)}, \dots, x^{(k)}, \dots$ está convergindo para a solução exata do sistema linear que é $x^* = (1.5, 1.5)$, tanto no método de Gauss-Jacobi quanto no de Gauss-Seidel.

No entanto, o método de Gauss-Seidel gera uma seqüência divergente para este mesmo sistema escrito da seguinte forma:

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

para a qual o esquema iterativo será:

$$\begin{cases} x_1^{(k+1)} = -3 + 3x_2^{(k)} \\ x_2^{(k+1)} = 3 - x_1^{(k+1)}. \end{cases}$$

Para $x^{(0)} = (0, 0)^T$ teremos:

$$\begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \end{pmatrix}$$

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 15 \\ 6 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 15 \\ -12 \end{pmatrix}, \dots$$

Graficamente, comprovamos a divergência de $x^* = (1.5, 1.5)^T$:

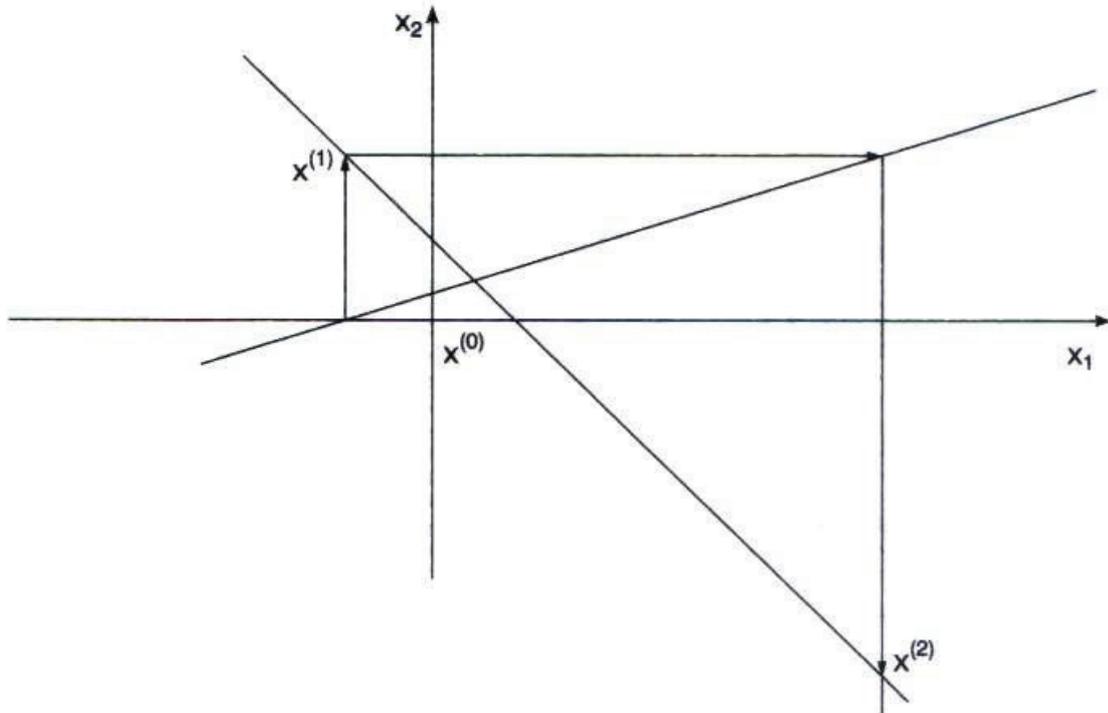


Figura 3.12

ESTUDO DA CONVERGÊNCIA DO MÉTODO DE GAUSS-SEIDEL

Como em todo processo iterativo, precisamos de critérios que nos forneçam garantia de convergência.

Para o método de Gauss-Seidel analisaremos os seguintes critérios, que estabelecem condições suficientes de convergência: o critério de Sassenfeld e o critério das linhas.

CRITÉRIO DE SASSENFELD

Seja $x^* = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ a solução exata do sistema $Ax = b$ e seja:

$$x^{(k)} = \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{pmatrix} \text{ a } k\text{-ésima aproximação de } x^*.$$

Queremos uma condição que nos garanta que $x^{(k)} \rightarrow x^*$ quando $k \rightarrow \infty$, ou seja, que $\lim_{k \rightarrow \infty} e_i^{(k)} = 0$ para $i = 1, \dots, n$ onde $e_i^{(k)} = x_i^{(k)} - x_i^*$.

Agora,

$$\begin{cases} e_1^{(k+1)} = -\frac{1}{a_{11}} (a_{12}e_2^{(k)} + a_{13}e_3^{(k)} + \dots + a_{1n}e_n^{(k)}) \\ e_2^{(k+1)} = -\frac{1}{a_{22}} (a_{21}e_1^{(k+1)} + a_{23}e_3^{(k)} + \dots + a_{2n}e_n^{(k)}) \\ \vdots \\ e_n^{(k+1)} = -\frac{1}{a_{nn}} (a_{n1}e_1^{(k+1)} + a_{n2}e_2^{(k+1)} + \dots + a_{n,n-1}e_{n-1}^{(k+1)}). \end{cases} \quad (4)$$

Chamemos de $E^{(k)} = \max_{1 \leq i \leq n} \{ |e_i^{(k)}| \}$ e sejam

$$\beta_1 = \sum_{j=2}^n |a_{1j}| / |a_{11}| \text{ e para } i = 2, 3, \dots, n$$

$$\beta_i = \left[\sum_{j=1}^{i-1} \beta_j |a_{ij}| + \sum_{j=i+1}^n |a_{ij}| \right] / |a_{ii}|.$$

Note que a condição $x^{(k)} \rightarrow x^*$ equivale a $E^{(k)} \rightarrow 0$ quando $k \rightarrow \infty$.

Mostremos por indução que $E^{(k+1)} \leq \beta E^{(k)}$ onde $\beta = \max_{1 \leq i \leq n} \beta_i$.

Para $i = 1$, temos

$$\begin{aligned} |e_1^{(k+1)}| &\leq \frac{1}{|a_{11}|} (|a_{12}| |e_2^{(k)}| + |a_{13}| |e_3^{(k)}| + \dots + |a_{1n}| |e_n^{(k)}|) \leq \\ &\leq \frac{1}{|a_{11}|} (|a_{12}| + |a_{13}| + \dots + |a_{1n}|) \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} \\ &\quad \underbrace{\hspace{10em}}_{= \beta_1} \end{aligned}$$

$$\text{Então, } |e_1^{(k+1)}| \leq \beta_1 \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} \leq \beta \max_{1 \leq j \leq n} \{|e_j^{(k)}|\}.$$

Suponhamos por indução que:

$$|e_2^{(k+1)}| \leq \beta_2 \max_{1 \leq j \leq n} \{|e_j^{(k)}|\}$$

$$|e_3^{(k+1)}| \leq \beta_3 \max_{1 \leq j \leq n} \{|e_j^{(k)}|\}$$

$$\vdots$$

$$|e_{i-1}^{(k+1)}| \leq \beta_{i-1} \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} \quad i \leq n$$

e mostraremos que $|e_i^{(k+1)}| \leq \beta_i \max_{1 \leq j \leq n} \{|e_j^{(k)}|\}$.

Mas,

$$\begin{aligned} |e_i^{(k+1)}| &\leq \frac{1}{|a_{ii}|} (|a_{i1}| |e_1^{(k+1)}| + |a_{i2}| |e_2^{(k+1)}| + \dots + |a_{i,i-1}| |e_{i-1}^{(k+1)}|) + \\ &\quad \frac{1}{|a_{ii}|} (|a_{i,i+1}| |e_{i+1}^{(k)}| + \dots + |a_{in}| |e_n^{(k)}|) \end{aligned}$$

e usando a hipótese de indução:

$$|e_i^{(k+1)}| \leq \frac{1}{|a_{ii}|} \underbrace{(|a_{i1}| \beta_1 + |a_{i2}| \beta_2 + \dots + |a_{ii-1}| \beta_{i-1} + |a_{ii+1}| + \dots + |a_{in}| \max_{1 \leq j \leq n} \{e_j^{(k)}\})}_{\beta_i}$$

ou seja,

$$|e_i^{(k+1)}| \leq \beta_i \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} \leq \beta \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} \quad \forall i, 1 \leq i \leq n.$$

Portanto,

$$\max_{1 \leq i \leq n} \{|e_i^{(k+1)}|\} = E^{(k+1)} \leq \beta \max_{1 \leq j \leq n} \{|e_j^{(k)}|\} = \beta E^{(k)}. \tag{5}$$

Assim, basta que $\beta < 1$ para que tenhamos $E^{(k+1)} < E^{(k)}$. Além disso, de (5) temos $E^{(k)} \leq \beta E^{(k-1)} \leq \beta(\beta E^{(k-2)}) \leq \dots \leq \beta^k E^{(0)}$ e desde que β seja menor que 1, então, $E^{(k)} \rightarrow 0$ quando $k \rightarrow \infty$ e, o que é importante, independentemente da aproximação inicial escolhida.

Com isto estabelecemos o *critério de Sassenfeld*:

$$\text{Sejam } \beta_1 = \frac{|a_{12}| + |a_{13}| + \dots + |a_{1n}|}{|a_{11}|}$$

$$\text{e } \beta_j = \frac{|a_{j1}| \beta_1 + |a_{j2}| \beta_2 + \dots + |a_{jj-1}| \beta_{j-1} + |a_{jj+1}| + \dots + |a_{jn}|}{|a_{jj}|}.$$

$$\text{Seja } \beta = \max_{1 \leq j \leq n} \{\beta_j\}.$$

Se $\beta < 1$, então o método de Gauss-Seidel gera uma seqüência convergente qualquer que seja $x^{(0)}$.

Além disto, quanto menor for β , mais rápida será a convergência.

Exemplo 15

a) Seja o sistema linear

$$\begin{cases} x_1 + 0.5x_2 - 0.1x_3 + 0.1x_4 = 0.2 \\ 0.2x_1 + x_2 - 0.2x_3 - 0.1x_4 = -2.6 \\ -0.1x_1 - 0.2x_2 + x_3 + 0.2x_4 = 1.0 \\ 0.1x_1 + 0.3x_2 + 0.2x_3 + x_4 = -2.5 \end{cases}$$

Para este sistema linear com esta disposição de linhas e colunas, temos

$$\beta_1 = [0.5 + 0.1 + 0.1]/1 = 0.7$$

$$\beta_2 = [(0.2)(0.7) + 0.2 + 0.1]/1 = 0.44$$

$$\beta_3 = [(0.1)(0.7) + (0.2)(0.44) + 0.2]/1 = 0.358$$

$$\beta_4 = [(0.1)(0.7) + (0.3)(0.44) + (0.2)(0.358)]/1 = 0.2736.$$

Portanto, $\beta = \max_{1 \leq i \leq n} \{\beta_i\} = 0.7 < 1$ e então temos a garantia de que o método de

Gauss-Seidel vai gerar uma seqüência convergente.

b) Seja agora o sistema linear

$$\begin{cases} 2x_1 + x_2 + 3x_3 = 9 \\ -x_2 + x_3 = 1 \\ x_1 + 3x_3 = 3 \end{cases}$$

com esta disposição de linhas e colunas, temos

$$\beta_1 = (1 + 3)/2 = 2 > 1!!$$

Trocando a 1ª equação pela 3ª, temos

$$\begin{cases} x_1 + 3x_3 = 3 \\ -x_2 + x_3 = 1 \\ 2x_1 + x_2 + 3x_3 = 9 \end{cases}$$

donde $\beta_1 = (0 + 3)/1 = 3 \gg 1!!$

A partir desta disposição, trocando a 1ª coluna pela 3ª, temos

$$\begin{cases} 3x_3 + x_1 = 3 \\ x_3 - x_2 = 1 \\ 3x_3 + x_2 + 2x_1 = 9. \end{cases}$$

Desta forma,

$$\beta_1 = 1/3$$

$$\beta_2 = [(1)(1/3) + 0]/1 = 1/3$$

$$\beta_3 = [(3)(1/3) + (1)(1/3)]/2 = 2/3.$$

Portanto, $\beta = \max_{1 \leq i \leq 3} \{\beta_i\} = 2/3 < 1$; então vale o critério de Sassenfeld e temos

garantia de convergência.

- c) Considerando agora o exemplo usado na interpretação geométrica do método de Gauss-Seidel, verificamos que o critério de Sassenfeld é apenas suficiente, pois para

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

vimos que o método de Gauss-Seidel gera uma seqüência convergente e, no entanto,

$$\beta_1 = 1/1 = 1 \quad \text{e}$$

$$\beta_2 = [1 \times 1]/3 = 1/3$$

e, portanto, o critério de Sassenfeld não é satisfeito.

CRITÉRIO DAS LINHAS

O critério das linhas estudado no método de Gauss-Jacobi pode ser aplicado no estudo da convergência do método de Gauss-Seidel.

O critério das linhas diz que se $\alpha = \max_{1 \leq k \leq n} \{\alpha_k\} < 1$, onde

$$\alpha_k = \left(\sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \right) / |a_{kk}|$$

então o método de Gauss-Seidel gera uma seqüência convergente.

A prova da convergência consiste em verificar que se o critério das linhas for satisfeito, automaticamente o critério de Sassenfeld é satisfeito:

$$\beta_1 = (|a_{12}| + |a_{13}| + \dots + |a_{1n}|) / |a_{11}| = \alpha_1 < 1$$

e, para $i = 2, \dots, k-1$, supor por indução que $\beta_i \leq \alpha_i < 1$.

Então,

$$\beta_k = (\beta_1 |a_{k1}| + \dots + \beta_{k-1} |a_{k,k-1}| + |a_{k,k+1}| + |a_{kn}|) / |a_{kk}| < (|a_{k1}| + \dots + |a_{k,k-1}| + |a_{k,k+1}| + \dots + |a_{kn}|) / |a_{kk}| = \alpha_k.$$

Assim, $\beta_i \leq \alpha_i$, $i = 1, \dots, n$.

Então, $\alpha_i < 1$ implica que $\beta_i < 1$, $i = 1, \dots, n$, ou seja, o critério de Sassenfeld é satisfeito.

Observamos, no entanto, que o critério de Sassenfeld pode ser satisfeito mesmo que o critério das linhas não o seja.

Exemplo 16

Seja o sistema linear:

$$\begin{cases} 3x_1 & + x_3 = 3 \\ x_1 - x_2 & = 1 \\ 3x_1 + x_2 + 2x_3 & = 9. \end{cases}$$

Temos

$$\alpha_1 = \beta_1 = \frac{1}{3} < 1 \quad \text{e}$$

$$\alpha_2 = \frac{1}{1} = 1; \text{ então o critério das linhas não é satisfeito.}$$

No entanto,

$$\beta_2 = \frac{1 \times \frac{1}{3}}{1} = \frac{1}{3} < 1$$

e

$$\beta_3 = \frac{3 \times \frac{1}{3} + \frac{1}{3}}{2} = \frac{2}{3} < 1.$$

Portanto, o critério de Sassenfeld é satisfeito.

3.4 COMPARAÇÃO ENTRE OS MÉTODOS

a) Convergência

Conforme vimos, os métodos diretos são processos finitos e, portanto, teoricamente, obtêm a solução de qualquer sistema não singular de equações. Já os métodos iterativos têm convergência assegurada apenas sob determinadas condições.

b) Esparsidade da matriz A

Inúmeros sistemas lineares, que surgem de problemas práticos como discretização de equações diferenciais por método dos elementos finitos ou método de diferenças finitas e descrição de redes de potência, são de grande porte com matriz dos coeficientes esparsa. Para estes casos, são adotados esquemas especiais para armazenamento da matriz A , que tiram proveito de sua esparsidade.

Os métodos diretos quando aplicados a sistemas esparsos provocam preenchimentos na matriz A , isto é, durante o processo de eliminação poderão surgir elementos não nulos em posições a_{ij} que originalmente eram nulas. Para exemplificar, considere a matriz A representada simbolicamente, sendo x a representação de um elemento não nulo:

$$\begin{pmatrix} x & x & 0 & x & x & 0 & x & x \\ x & 0 & x & 0 & 0 & x & 0 & x \\ 0 & x & x & x & 0 & x & x & 0 \\ 0 & 0 & x & x & x & 0 & x & 0 \\ x & 0 & 0 & 0 & x & x & 0 & 0 \\ 0 & x & 0 & 0 & 0 & x & x & 0 \\ x & 0 & x & 0 & 0 & x & x & x \\ x & 0 & 0 & x & 0 & 0 & x & 0 \end{pmatrix}.$$

Após a 1ª etapa do processo de eliminação teremos:

$$\begin{pmatrix} x & x & 0 & x & x & 0 & x & x \\ x & \bullet & x & \bullet & \bullet & x & \bullet & x \\ 0 & x & x & x & 0 & x & x & 0 \\ 0 & 0 & x & x & x & 0 & x & 0 \\ x & \bullet & 0 & \bullet & x & x & \bullet & \bullet \\ 0 & x & 0 & 0 & 0 & x & x & 0 \\ x & \bullet & x & \bullet & \bullet & x & x & x \\ x & \bullet & 0 & x & \bullet & 0 & x & \bullet \end{pmatrix}$$

onde \bullet representa o elemento não nulo que preencheu uma posição originalmente nula.

Portanto, se a matriz A for esparsa e de grande porte, uma desvantagem dos métodos diretos para a resolução do sistema linear $Ax = b$ é o preenchimento na matriz, exigindo técnicas especiais para escolha do pivô para reduzir este preenchimento. Pode-se conseguir boas implementações para a fatoração LU, empregando-se técnicas de esparsidade, contudo existem situações nas quais pode ser impossível aplicar um método direto, daí a alternativa são os métodos iterativos que têm como principal vantagem não alterar a estrutura da matriz A dos coeficientes.

c) Erros de arredondamento

Vimos que os métodos diretos apresentam sérios problemas com erros de arredondamento. Uma forma de amenizar esses problemas é adotar técnicas de pivoteamento. Os métodos iterativos têm menos erros de arredondamento, visto que a convergência, uma vez

assegurada, independe da aproximação inicial. Desta forma, somente os erros cometidos na última iteração afetam a solução, pois os erros cometidos nas iterações anteriores não levarão à divergência do processo nem à convergência a um outro vetor que não a solução.

3.5 EXEMPLOS FINAIS

Exemplo 17

Retomando o Exemplo 1 da Introdução, com $\alpha = \text{sen}(45^\circ) = \sqrt{2}/2$, resolvemos o sistema linear resultante pelo método da Eliminação de Gauss com pivoteamento parcial. Obtivemos o vetor solução:

$$(-29.247105, 19, 10, -28, 13.853892, 19, 0, -28, 9.235928, 22, 0, -16, -9.235928, 22, 16, -24.629141, 16)^T.$$

Permutando algumas linhas de forma que os elementos da diagonal principal fossem não nulos, conseguimos o esquema iterativo do método de Gauss-Seidel. No entanto, a seqüência gerada divergiu da solução.

Exemplo 18

Seja o sistema linear

$$\begin{bmatrix} 2 & 1 & 7 & 4 & -3 & -1 & 4 & 4 & 7 & 0 \\ 4 & 2 & 2 & 3 & -2 & 0 & 3 & 3 & 4 & 1 \\ 3 & 4 & 4 & 2 & 1 & -2 & 2 & 1 & 9 & -3 \\ 9 & 3 & 5 & 1 & 0 & 5 & 6 & -5 & -3 & 4 \\ 2 & 0 & 7 & 0 & -5 & 7 & 1 & 0 & 1 & 6 \\ 1 & 9 & 8 & 0 & 3 & 9 & 9 & 0 & 0 & 5 \\ 4 & 1 & 9 & 0 & 4 & 3 & 7 & -4 & 1 & 3 \\ 6 & 3 & 1 & 1 & 6 & 8 & 3 & 3 & 0 & 2 \\ 6 & 5 & 0 & -7 & 7 & -7 & 6 & 2 & -6 & 1 \\ 1 & 6 & 3 & 4 & 8 & 3 & -5 & 0 & -6 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} = \begin{bmatrix} 86 \\ 45 \\ 52.5 \\ 108 \\ 66.5 \\ 90.5 \\ 139 \\ 61 \\ -43.5 \\ 31 \end{bmatrix}$$

A solução obtida pelo método da Eliminação de Gauss com pivoteamento parcial foi:

$$\bar{x} = (3, -4.5, 7, 8, 3.5, 2, 4, -3.5, 2, 1.5)^T.$$

Também para este exemplo, trocando apenas a nona equação com a décima, não conseguimos uma seqüência convergente para o método de Gauss-Seidel.

Exemplo 19

Seja o sistema linear

$$\begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 4 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} = \begin{bmatrix} -110 \\ -30 \\ -40 \\ -110 \\ 0 \\ -15 \\ -90 \\ -25 \\ -55 \\ -65 \end{bmatrix}$$

Resolvendo pelo método da Eliminação de Gauss, com estratégia de pivoteamento parcial, obtivemos o seguinte vetor solução:

$$\bar{x} = (-48.646412, -35.4947917, -25.6157408, -49.0908565, -37.7170139, -26.9681713, -39.3142361, -29.5399306, -26.8773148, -22.9693287)^T.$$

Aplicando o método de Gauss-Seidel com o esquema iterativo montado a partir da disposição original das equações, com $x^{(0)} = (20, \dots, 20)^T$ e $\epsilon = 10^{-7}$, obtivemos o mesmo vetor \bar{x} após 28 iterações.

EXERCÍCIOS

1. Escreva um algoritmo para a resolução de um sistema linear triangular inferior.
2. Verifique que o “custo” = número de operações efetuadas para resolver um sistema linear triangular inferior é o mesmo que para multiplicar uma matriz triangular por um vetor.
3. Verifique que o número de operações necessárias no método da Eliminação de Gauss, sem pivoteamento parcial, é $\frac{2n^3}{3} + \frac{n^2}{2} - \frac{7n}{6}$, na fase de triangularização da matriz A , e n^2 , na fase da resolução do sistema triangular superior. Estão sendo contadas as operações de divisão, multiplicação e soma.

(Lembramos que $\sum_{k=1}^{n-1} k^2 = \frac{(n-1)n(2n-1)}{6}$.)

4. Seja $Ax = b$ um sistema $n \times n$ com matriz tridiagonal ($a_{ij} = 0$ se $|i - j| > 1$).
 - a) Escreva um algoritmo para resolver $Ax = b$ através da Eliminação de Gauss com estratégia de pivoteamento parcial de modo que a estrutura especial da matriz A seja explorada.
 - b) Compare o “custo” de resolvê-lo por Eliminação de Gauss via algoritmo tradicional, com o de resolvê-lo pelo algoritmo do item (a).
 - c) Teste seus resultados com o sistema:

$$\begin{cases} 2x_1 - x_2 = 1 \\ -x_{i-1} + 2x_i - x_{i+1} = 0, & 2 \leq i \leq (n-1) \\ -x_{n-1} + 2x_n = 0 \end{cases}$$

para $n = 10$.

5. Resolva o sistema linear abaixo utilizando o método da Eliminação de Gauss:

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

6. Analise os sistemas lineares abaixo com relação ao número de soluções, usando o método da Eliminação de Gauss (trabalhe com três casas decimais):

$$a) \begin{cases} 3x_1 - 2x_2 + 5x_3 + x_4 = 7 \\ -6x_1 + 4x_2 - 8x_3 + x_4 = -9 \\ 9x_1 - 6x_2 + 19x_3 + x_4 = 23 \\ 6x_1 - 4x_2 - 6x_3 + 15x_4 = 11 \end{cases}$$

$$b) \begin{cases} 0.252x_1 + 0.36x_2 + 0.12x_3 = 7 \\ 0.112x_1 + 0.16x_2 + 0.24x_3 = 8 \\ 0.147x_1 + 0.21x_2 + 0.25x_3 = 9 \end{cases}$$

7. O cálculo do determinante de matrizes quadradas pode ser feito usando o método da Eliminação de Gauss.

a) Deduza o método.

b) Aplique-o no cálculo do determinante das matrizes dos sistemas dos Exercícios 5 e 6.

c) Inclua o cálculo do determinante da matriz A do sistema linear $Ax = b$ no algoritmo do método da Eliminação de Gauss.

8. Demonstre que, se no início da etapa k do método da Eliminação de Gauss tivermos $a_{kk}^{(k-1)} = a_{(k+1)k}^{(k-1)} = \dots = a_{nk}^{(k-1)} = 0$, então $\det(A) = 0$ e conseqüentemente A não é inversível. ($a_{ij}^{(k-1)}$ é o elemento da posição ij no início da etapa k .)

9. Podemos encontrar a fatoração LU de A diretamente, usando simplesmente a definição de produto de matrizes. Esquemas deste tipo são conhecidos como esquemas compactos, e o equivalente à fatoração $A = LU$ com L triangular inferior com diagonal unitária e U triangular superior é chamado de *redução de Doolittle*.

Supondo que a fatoração LU de A seja possível, de uma forma única,

a) multiplique a primeira linha de L pela j -ésima coluna de U e iguale a a_{1j} . Verifique que desta forma obtém-se o elemento u_{1j} ;

b) repita o item (a), multiplicando agora a i -ésima linha de L pela primeira coluna de U , e igualando a a_{i1} será possível obter l_{i1} ;

- c) use o mesmo raciocínio de (a) e (b) para deduzir que, se as $(k - 1)$ primeiras linhas de U e colunas de L já foram determinadas, então

$$u_{kj} = a_{kj} - \sum_{m=1}^{k-1} l_{km} u_{mj}, \quad j = k, (k + 1), \dots, n$$

e

$$l_{ik} = \left(a_{ik} - \sum_{m=1}^{k-1} l_{im} u_{mk} \right) / u_{kk}, \quad i = (k + 1), \dots, n;$$

- d) explique por que, se A é não singular, então U também o será, donde $u_{kk} \neq 0$, $k = 1, \dots, n$;
- e) escreva um algoritmo para a fatoração LU de A usando a redução de Doolittle;
- f) teste seu algoritmo, fatorando A e então resolvendo o sistema abaixo, sendo

$$A = \begin{pmatrix} 2 & 3 & 1 & 5 \\ 1 & 3.5 & 1 & 7.5 \\ 1.4 & 2.7 & 5.5 & 12 \\ -2 & 1 & 3 & 28 \end{pmatrix} \text{ e } b = \begin{pmatrix} 11 \\ 13 \\ 21.6 \\ 30 \end{pmatrix}$$

10. Calcule a fatoração LU de A , se possível:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 3 & 2 & 0 \end{pmatrix}$$

11. a) Mostre que resolver $AX = B$, onde A é matriz $n \times n$, X e B são matrizes $n \times m$, é o mesmo que resolver m sistemas do tipo $Ax = b$, onde A é matriz $n \times n$, x e b , vetores $n \times 1$.
- b) Usando o item (a), verifique que A^{-1} pode ser obtida através de resolução de n sistemas lineares.
- c) Entre o método da Eliminação de Gauss e a fatoração LU, qual o mais indicado para o cálculo de A^{-1} ?

d) Aplique o método escolhido no item (c) para obter a inversa da matriz

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

12. Mostre que, se A é matriz não singular e $A = LU$, então $A = LD\bar{U}$, onde D é matriz diagonal e \bar{U} matriz triangular superior com diagonal unitária.
13. Se $A = LDU$, como fica a resolução de $Ax = b$?
14. Escreva um algoritmo para o método da Eliminação de Gauss, usando estratégia de pivoteamento parcial.
15. Seja resolver o sistema linear $Ax = b$ pelo método da Eliminação de Gauss, com estratégia de pivoteamento parcial:

se $M = \max_{i,j} \{ |a_{ij}| \} \ 1 \leq i, j \leq n$, prove que, após o primeiro estágio, $|a_{ij}^{(1)}| \leq 2M$.

16. a) Resolva os itens (b) e (c) do Exercício 11, considerando a estratégia de pivoteamento parcial.
- b) Use os resultados de (a) para encontrar a inversa da matriz

$$A = \begin{pmatrix} 1 & 12 & 3 \\ 2 & 4 & 16 \\ 3 & 15 & 7 \end{pmatrix}$$

17. Trabalhando com arredondamento para dois dígitos significativos em todas as operações, resolva o sistema linear abaixo pelo método da Eliminação de Gauss, sem e com pivoteamento parcial. Discuta seus resultados:

$$\begin{cases} 16x_1 + 5x_2 = 21 \\ 3x_1 + 2.5x_2 = 5.5 \end{cases}$$

Refaça o exercício usando truncamento para dois dígitos significativos.

18. Trabalhando com quatro dígitos significativos, resolva os sistemas lineares a seguir (ou detecte que não há solução). Use pivoteamento parcial. Estabeleça um critério para decidir se números pequenos em lugares importantes são considerados como zero ou não. Confira a solução obtida:

$$a) \begin{cases} 1.12a + 6b = 1.3 \\ 2.21a + 12b = 2.6 \end{cases}$$

$$b) \begin{cases} 1.12a + 6b = 1.3 \\ 2.24a + 12b = 3. \end{cases}$$

19. Justifique se for verdadeira ou dê contra-exemplo se for falsa a afirmação:

“Dada uma matriz A , $n \times n$, sua fatoração LU, obtida com estratégia de pivoteamento parcial, é tal que todos os elementos da matriz L têm módulo menor ou igual a 1”.

20. O vetor p que armazena a informação sobre as permutações realizadas durante a fatoração LU pode ser construído como $p(k) = i$, se na etapa k a linha i da matriz $A^{(k-1)}$ for escolhida como a linha pivotal. Desta forma, o vetor terá dimensão $(n-1) \times 1$. Para o Exemplo 7, teríamos $p = (3, 3)^T$; a dimensão de p é $(n-1) \times 1$, uma vez que são realizados $(n-1)$ etapas. Esta forma para o vetor p é mais eficiente em implementações computacionais porque na fase da resolução dos sistemas triangulares o vetor Pb pode ser armazenado sobre o vetor b original.

Reescreva o algoritmo para a resolução de $Ax = b$ através da fatoração LU com estratégia de pivoteamento parcial, usando o vetor p conforme descrito acima.

21. Prove que se B é matriz $m \times n$, $m \geq n$ com posto completo, então a matriz $C = B^T B$ é simétrica, definida positiva.

22. Em cada caso:

a) verifique se o critério de Sassenfeld é satisfeito;

b) resolva por Gauss-Seidel, se possível:

$$A = \begin{pmatrix} 10 & 1 & 1 \\ 1 & 10 & 1 \\ 1 & 1 & 10 \end{pmatrix}; \quad b = \begin{pmatrix} 12 \\ 12 \\ 12 \end{pmatrix}$$

e

$$A = \begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix}; \quad b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

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

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

- b) Escolha o menor valor inteiro e positivo para k e faça duas iterações do método de Gauss-Seidel para o sistema obtido.
- c) Comente o erro cometido no item (b).

24. a) Considere o sistema linear

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 5 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 1 \end{pmatrix}$$

Verifique, usando eliminação gaussiana, que este sistema não tem solução. Qual será o comportamento do método de Gauss-Seidel?

- b) Através de um sistema 2×2 , dê uma interpretação geométrica do que ocorre com Gauss-Seidel quando o sistema não tem solução e quando existem infinitas soluções.

25. a) Aplique analítica e graficamente os métodos de Gauss-Jacobi e Gauss-Seidel no sistema:

$$\begin{cases} 2x_1 + 5x_2 = -3 \\ 3x_1 + x_2 = 2 \end{cases}$$

- b) Repita o item (a) para o sistema obtido permutando as equações.
- c) Analise seus resultados.

26. Verifique que, se $\lim_{k \rightarrow \infty} x^{(k)} = \alpha$, onde $x^{(j+1)} = Cx^{(j)} + g$, então α é solução de $x = Cx + g$.

27. Prove que, no método de Gauss-Seidel, vale a relação:

$$\begin{cases} e_1^{(k+1)} = \frac{-1}{a_{11}} (a_{12} e_2^{(k)} + a_{13} e_3^{(k)} + \dots + a_{1n} e_n^{(k)}) \\ e_2^{(k+1)} = \frac{-1}{a_{22}} (a_{21} e_1^{(k+1)} + a_{23} e_3^{(k)} + \dots + a_{2n} e_n^{(k)}) \\ \vdots \\ e_n^{(k+1)} = \frac{-1}{a_{nn}} (a_{n1} e_1^{(k+1)} + a_{n2} e_2^{(k+1)} + \dots + a_{n(n-1)} e_{n-1}^{(k+1)}) \end{cases}$$

28. a) Um possível teste de parada para um método iterativo é testar se $Ax^{(k)} - b$ está próximo de zero, quando então $x^{(k)}$ será escolhido como aproximação da solução x^* do sistema. Como realizar computacionalmente este teste?

b) Compare o “custo” computacional de usar o teste acima com o “custo” do teste $(x^{(k+1)} - x^{(k)})$ estar próximo de zero.

29. Considere o sistema linear cuja matriz dos coeficientes é a matriz esparsa

$$A = \begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 16 & 0 \\ 0 & 0 & 4 & 0 & 0 \end{pmatrix} \quad \text{e} \quad b = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 20 \\ 4 \end{pmatrix}.$$

a) Ache a solução por inspeção.

b) Faça mudanças de linhas na matriz original para facilitar a aplicação do método da Eliminação de Gauss. O que você pode concluir, de uma maneira geral?

c) Aplique o método de Gauss-Seidel ao sistema. Comente seu desempenho.

d) Faça uma comparação da utilização de métodos diretos e iterativos na resolução de sistemas lineares esparsos.

30. Ao resolver um sistema linear $Ax = b$ por um método direto, vários problemas podem implicar a obtenção de uma solução, x_0 , apenas aproximada, para \bar{x} . Chamamos $r_0 = Ax_0 - b$ o resíduo associado a x_0 . Note que $r_0 = Ax_0 - b = Ax_0 - A\bar{x} = A(x_0 - \bar{x}) = Az$. Se $Az = r_0$ fosse resolvido sem erro, então $\bar{x} = x_0 - z$ seria a solução do sistema: este não é o caso, mas esta observação pode ser usada como base para um esquema iterativo para “refinamento” de soluções aproximadas.

É razoável supormos que $x_1 = x_0 - \tilde{z}$ (\tilde{z} , solução aproximada de $Az = r_0$) seja uma “solução” de $Ax = b$, melhor do que x_0 . Com x_1 construímos um novo resíduo $r_1 = Ax_1 - b$ e continuamos o processo. Na realidade, podemos continuá-lo tantas vezes quanto quisermos, o que nos fornece o seguinte

Algoritmo: (*Refinamento Iterativo*)

Seja r uma solução (aproximada) para $Ax = b$, obtida por algum método direto, $\varepsilon > 0$ e $itmax$ o número máximo de iterações permitido.

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

$$r = Ax - b$$

z : solução de $Az = r$

$$x = x - z$$

se $\max_{1 \leq i \leq n} |z_i| / \max_{1 \leq i \leq n} |x_i| < \varepsilon$, fim. A solução é x . Se $i > itmax$, envie mensagem de não convergência em $itmax$ iterações.

- a) Justifique por que devemos usar fatoração LU neste caso para resolver os sistemas envolvidos.
- b) Neste caso, não devemos sobrepor L e U em A . Justifique.

31. Para cada um dos sistemas lineares a seguir, analise a existência ou não de solução, bem como unicidade de solução, no caso de haver existência.

a) $3x + 2y = 7$

b)
$$\begin{cases} 4m - 2k = 8 \\ 5m + k = 20 \end{cases}$$

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

$$d) 6x + 4y - 3z + w = 10$$

$$e) \begin{cases} 3x_1 - x_2 + 4x_3 = 6 \\ 6x_1 - 2x_2 + 5x_3 = 8 \end{cases}$$

$$f) \begin{cases} 3x - 2y + z = 8 \\ x - 3y + 4z = 6 \\ 9x + 4y - 5z = 11 \end{cases}$$

$$g) \begin{cases} 2x - y + 3z = 8 \\ x - 5y + z = -1 \\ 4x - 11y + 5z = 6 \end{cases}$$

$$h) \begin{cases} x - 3y + z = 1 \\ 6x - 18y + 4z = 2 \\ 7x - 21y + 5z = 3 \end{cases}$$

$$i) \begin{cases} x - 3y + z = 1 \\ 6x - 18y + 4z = 2 \\ -x + 3y - z = 4 \end{cases}$$

$$j) \begin{cases} 6u - 3v = 6 \\ 3u - 1.5v = 3 \\ 2u - v = 8 \\ 8u - 4v = 1.7 \end{cases}$$

$$k) \begin{cases} 4a + 5b + 7c = 1 \\ -a - b - c = 2 \\ 3a + 4b + 6c = 3 \\ -a + b + 7c = -11 \\ 2a + 5b + 13c = -8 \end{cases}$$

32. Invente um sistema linear com 6 equações e 4 variáveis sem solução, outro com solução única e outro com infinitas soluções. Justifique cada caso.

33. Resolva os sistemas lineares abaixo usando a fatoração de Cholesky:

$$a) \begin{cases} 16x_1 + 4x_2 + 8x_3 + 4x_4 = 32 \\ 4x_1 + 10x_2 + 8x_3 + 4x_4 = 26 \\ 8x_1 + 8x_2 + 12x_3 + 10x_4 = 38 \\ 4x_1 + 4x_2 + 10x_3 + 12x_4 = 30 \end{cases}$$

$$b) \begin{cases} 20x_1 + 7x_2 + 9x_3 = 16 \\ 7x_1 + 30x_2 + 8x_3 = 38 \\ 9x_1 + 8x_2 + 30x_3 = 38 \end{cases}$$

34. Seja $A = \begin{pmatrix} 5 & 7 \\ 7 & 13 \end{pmatrix}$

- a) obtenha o fator de Cholesky de A ;
 b) encontre 3 outras matrizes triangulares inferiores R , tais que $A = RR^T$.

35. Prove que se $A: n \times n$ é simétrica definida positiva então A^{-1} existe e é simétrica definida positiva.
 36. Dizemos que $A: n \times n$ é uma matriz banda com amplitude q se $a_{ij} = 0$ quando $|i - j| > q$.

Escreva um algoritmo para obter a fatoração de Cholesky de uma matriz banda q , simétrica, definida positiva, tirando proveito de sua estrutura.

PROJETO

- a) Compare as soluções dos sistemas lineares

$$\begin{cases} x - y = 1 \\ x - 1.00001 y = 0 \end{cases} \quad \text{e} \quad \begin{cases} x - y = 1 \\ x - 0.99999 y = 0 \end{cases}$$

Fatos como este ocorrem quando a matriz A do sistema está próxima de uma matriz singular e então o sistema é *mal condicionado*.

Dizemos que um sistema linear é *bem condicionado* se pequenas mudanças nos coeficientes e/ou nos termos independentes acarretarem pequenas mudanças na solução do sistema. Caso contrário, o sistema é dito mal condicionado.

Embora saibamos que uma matriz A pertence ao conjunto das matrizes não inversíveis se, e somente se, $\det(A) = 0$, o fato de uma matriz A ter $\det(A) \approx 0$ não implica necessariamente que o sistema linear que tem A por matriz de coeficientes seja mal condicionado.

O *número de condição* de A , $\text{cond}(A) = \|A\| \|A^{-1}\|$, onde $\|\cdot\|$ é uma norma de matrizes [14], é uma medida precisa do bom ou mau condicionamento do sistema que tem A por matriz de coeficientes, pois demonstra-se que

$$\frac{1}{\text{cond}(A)} = \min \left\{ \frac{\|A - B\|}{\|A\|} \text{ tais que } B \text{ é não inversível} \right\}.$$

b) As matrizes de Hilbert, H_n , onde

$h_{ij} = \frac{1}{i+j-1}$, $1 \leq i, j \leq n$ são exemplos clássicos de matrizes mal condicionadas.

(b.1) – Use pacotes computacionais, que estimam ou calculam $\text{cond}(A)$, para verificar que quanto maior for n , mais mal condicionada é H_n .

(b.2) – Resolva os sistemas $H_n x = b_n$, $n = 3, 4, 5, \dots, 10$, onde b_n é o vetor cuja i -ésima componente é

$$\sum_{j=1}^n \frac{1}{i+j-1} \text{ Desta forma a solução exata será: } x^* = (1 \ 1 \ \dots \ 1)^T.$$

(b.3) – Analise seus resultados.