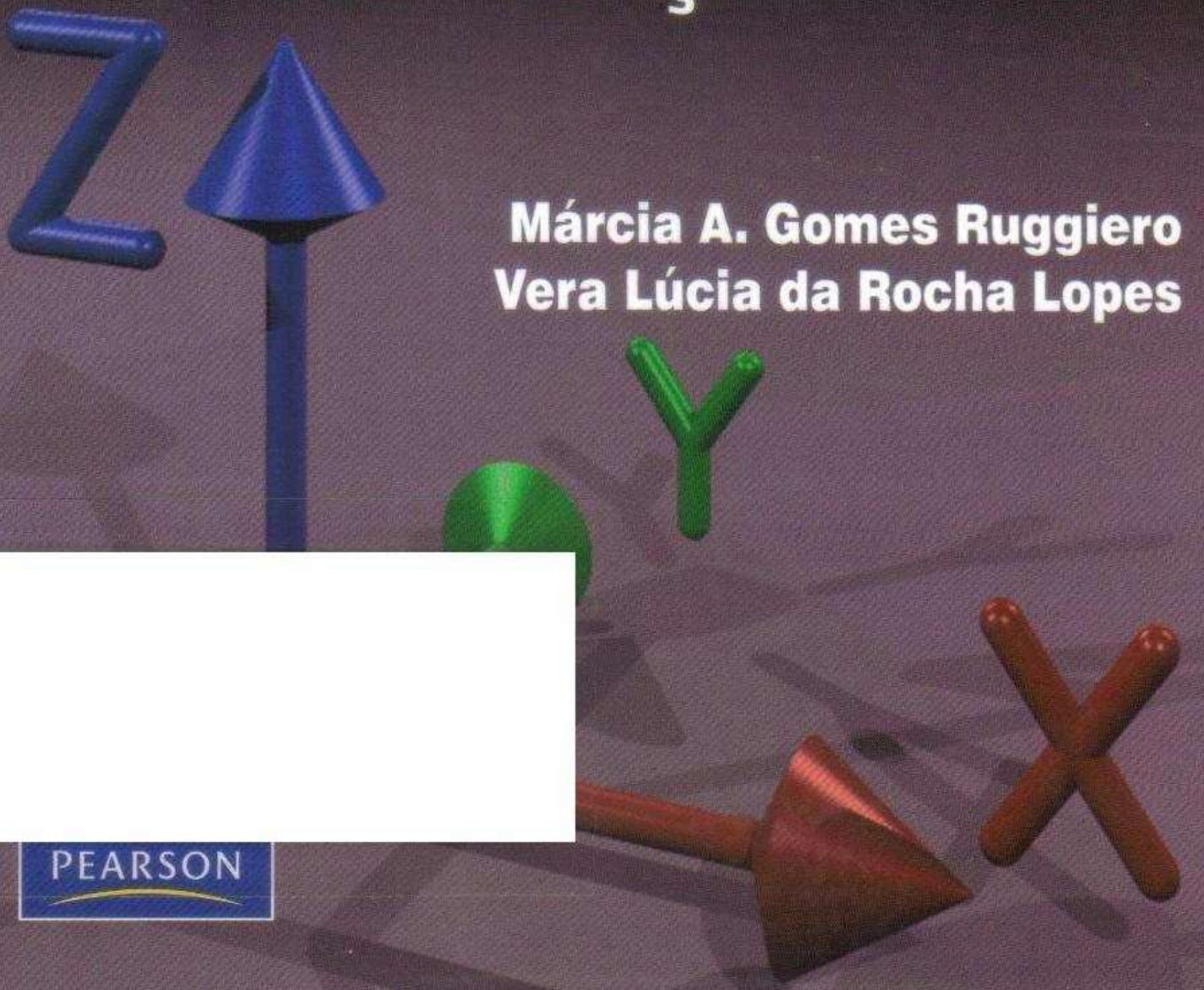
CÁLCULO NUMÉRICO

ASPECTOS TEÓRICOS E COMPUTACIONAIS

2ª Edição



Márcia A. Gomes Ruggiero
Vera Lúcia da Rocha Lopes

PEARSON

CÁLCULO NUMÉRICO

ASPECTOS TEÓRICOS E COMPUTACIONAIS

2ª Edição

OUTROS LIVROS NA ÁREA

Boulos	—	Cálculo Diferencial e Integral (2 vol. + pré-cálculo)
Camargo	—	Geometria Analítica - 3ª edição
Flemming	—	Cálculo A
Gonçalves	—	Cálculo B
Gonçalves	—	Cálculo C
Matos	—	Séries e Equações Diferenciais
Salvetti	—	Algoritmos
Simmons	—	Cálculo com Geometria Analítica (2 vol.)
Sperandio	—	Cálculo Numérico
Thomas	—	Cálculo (2 vol.) - 10ª edição
Zill / Cullen	—	Equações Diferenciais (2 vol.)

Makron Books
é um selo da
PEARSON

www.pearson.com.br

ISBN 978-85-346-0204-4

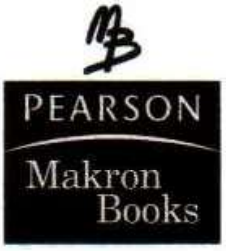
CÁLCULO NUMÉRICO

ASPECTOS TEÓRICOS E COMPUTACIONAIS

2ª edição

Márcia A. Gomes Ruggiero
Vera Lúcia da Rocha Lopes

Departamento de Matemática Aplicada
IMECC – UNICAMP



São Paulo

Brasil Argentina Colômbia Costa Rica Chile Espanha
Guatemala México Peru Porto Rico Venezuela

Dedicamos este livro a nossos pais:

Atayde Gomes (in memorian)
Benedita Germano Gomes

Aarão Soares da Rocha (in memorian)
Hulda Vilhena dos Reis Rocha (in memorian)

MAKRON
Books

PREFÁCIO À SEGUNDA EDIÇÃO

Desde que foi publicada a 1ª edição deste livro, recebemos inúmeras sugestões através de cartas e contatos nos Congressos anuais organizados pela Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC). Atendendo às sugestões de correções, inclusão de tópicos, e sentindo a necessidade de modernização e atualização do texto, optamos por elaborar esta 2ª edição, motivadas pela grande aceitação deste livro como texto básico para cursos de Cálculo Numérico em várias universidades.

Em todos os capítulos já existentes na edição anterior, foram feitas correções e alterações no texto: suprimimos e acrescentamos alguns exercícios e introduzimos uma seção de projetos em quase todos os capítulos; nos projetos o nível de exigência, tanto teórico quanto computacional, é maior que nos exercícios; no Apêndice desta nova edição apresentamos as respostas de todos os exercícios.

Os Capítulos 3, 5 e 8 foram os mais alterados. No Capítulo 3, sobre Resolução de Sistemas Lineares, além de ampliarmos a introdução sobre existência e número de soluções, introduzimos uma seção sobre Fatoração de Cholesky. No Capítulo 5, a seção sobre Splines em Interpolação foi ampliada, acrescentando as Splines Cúbicas em Interpolação. Finalmente, no Capítulo 8, acrescentamos uma seção sobre Problemas de Valor de Contorno em Equações Diferenciais Ordinárias.

O atual Capítulo 4 é inteiramente novo, e trata da Solução de Sistemas Não Lineares.

As listagens de programas computacionais foram eliminadas e sugerimos que alunos e professores utilizem softwares matemáticos que possibilitam que algoritmos de

métodos numéricos sejam facilmente adaptados para que possam ser implementados. Temos usado na UNICAMP o MATLAB[1] como material de apoio computacional no curso de Cálculo Numérico e temos obtido boa aceitação por parte dos alunos. Consideramos que este software é um material de fácil manuseio e eficiente para a resolução dos exercícios e projetos computacionais propostos.

Agradecemos aos professores, alunos e leitores que nos fizeram críticas e sugestões que muito contribuíram na elaboração desta edição. Dedicamos um agradecimento especial ao professor Lúcio Tunes dos Santos que gentilmente escreveu o projeto *Newton e os Fractais* para o Capítulo 4 e aos auxiliares didáticos Ricardo Caetano Azevedo Biloti, Suzana Lima de Campos Castro e Lin Xu, que resolveram a maior parte dos exercícios, e especialmente a Roberto Andreani e a Renato da Rocha Lopes, pela leitura do texto e contribuições para as listas de exercícios e projetos.

1. MATLAB é uma marca registrada do MathWorks, Inc. Uma boa introdução à versão 4.2 do MATLAB é a quarta edição do MATLAB Primer de K. Sigmon, editado em 1994 por CRC Press, Boca Raton, FL.

MAKRON
Books

SUMÁRIO

MAKRON
Books

INTRODUÇÃO

Ao escrever este livro, nosso objetivo principal foi oferecer ao estudante de graduação na área de ciências exatas um texto que lhe apresentasse alguns métodos numéricos com sua fundamentação teórica, suas vantagens e dificuldades computacionais.

Apresentamos oito capítulos e procuramos seguir em todos eles o mesmo desenvolvimento. Em todos os capítulos, incluímos uma lista de exercícios e a maioria deles apresenta propostas de projetos. As respostas de todos os exercícios se encontram no Apêndice.

Supomos que o aluno tenha conhecimentos de Cálculo e de Álgebra Linear e familiaridade com alguma linguagem de programação de computador.

O Capítulo 1 trata de erros em processos numéricos, sem a intensão de fazer um estudo detalhado do assunto; o objetivo é alertar o aluno sobre as dificuldades numéricas que podem ocorrer ao se trabalhar com um computador.

O Capítulo 2 apresenta vários métodos para a resolução de equações não lineares do tipo $f(x) = 0$ e um desenvolvimento específico para o caso de raízes reais de equações polinomiais.

No Capítulo 3, abordamos a resolução de sistemas de equações lineares apresentando métodos diretos e iterativos.

Incluímos nesta edição uma introdução sobre a resolução de sistemas de equações não lineares, que é o tema do Capítulo 4.

O Capítulo 5 apresenta a interpolação polinomial como forma de se obter uma aproximação para uma função f(x). Além disto, introduz as funções spline interpolantes.

O método dos quadrados mínimos, como outra forma de aproximação de funções, é apresentado no Capítulo 6, onde desenvolvemos o método dos quadrados mínimos lineares e damos sugestões de como contornar certos casos não lineares.

O tema do Capítulo 7 é a integração numérica para $\int_a^b f(x)\,dx$, através das fórmulas fechadas de Newton-Cotes. Introduzimos ainda as fórmulas de quadratura Gaussiana.

O Capítulo 8 aborda métodos numéricos para resolução de problemas de valor inicial e de contorno em equações diferenciais ordinárias.

MAKRON
Books

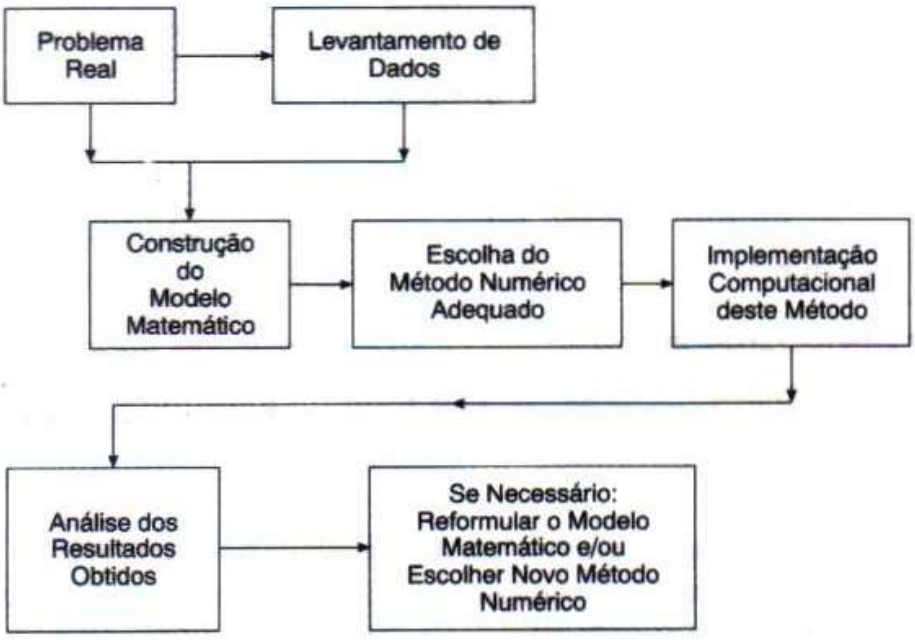
CAPÍTULO 1

NOÇÕES BÁSICAS SOBRE ERROS

1.1 INTRODUÇÃO

Nos capítulos seguintes, estudaremos métodos numéricos para a resolução de problemas que surgem nas mais diversas áreas.

A resolução de tais problemas envolve várias fases que podem ser assim estruturadas:



Não é raro acontecer que os resultados finais estejam distantes do que se esperaria obter, ainda que todas as fases de resolução tenham sido realizadas corretamente.

Os resultados obtidos dependem também:

- da precisão dos dados de entrada;
- da forma como estes dados são representados no computador;
- das operações numéricas efetuadas.

Os dados de entrada contêm uma imprecisão inerente, isto é, não há como evitar que ocorram, uma vez que representam medidas obtidas usando equipamentos específicos, como, por exemplo, no caso de medidas de corrente e tensão num circuito elétrico, ou então podem ser dados resultantes de pesquisas ou levantamentos, como no caso de dados populacionais obtidos num recenseamento.

Neste capítulo, estudaremos os erros que surgem da representação de números num computador e os erros resultantes das operações numéricas efetuadas, [23], [26] e [31].

1.2 REPRESENTAÇÃO DE NÚMEROS

Exemplo 1

Calcular a área de uma circunferência de raio 100 m.

RESULTADOS OBTIDOS

a) A = 31400 m^2

b) A = 31416 m^2

c) A = 31415.92654 m^2

Como justificar as diferenças entre os resultados? É possível obter "exatamente" esta área?

Exemplo 2

Efetuar os somatórios seguintes em uma calculadora e em um computador:

$$S = \sum_{i=1}^{30000} x_i \qquad \text{para } x_i = 0.5 \text{ e para } x_i = 0.11$$

RESULTADOS OBTIDOS

i)	para $x_i = 0.5$:	na calculadora:	S = 15000
		no computador:	S = 15000
ii)	para $x_i = 0.11$:	na calculadora:	S = 3300
		no computador:	S = 3299.99691

Como justificar a diferença entre os resultados obtidos pela calculadora e pelo computador para $x_i = 0.11$?

Os erros ocorridos nos dois problemas dependem da representação dos números na máquina utilizada.

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

O número π, por exemplo, não pode ser representado através de um número finito de dígitos decimais. No Exemplo 1, o número π foi escrito como 3.14, 3.1416 e 3.141592654 respectivamente nos casos (*a*), (*b*) e, (*c*). Em cada um deles foi obtido um resultado diferente, e o erro neste caso depende exclusivamente da aproximação escolhida para π. Qualquer que seja a circunferência, a sua área nunca será obtida exatamente, uma vez que π é um número irracional.

Como neste exemplo, qualquer cálculo que envolva números que não podem ser representados através de um número finito de dígitos não fornecerá como resultado um valor exato. Quanto maior o número de dígitos utilizados, maior será a precisão obtida. Por isso, a melhor aproximação para o valor da área da circunferência é aquela obtida no caso (*c*).

Além disto, um número pode ter representação finita em uma base e não-finita em outras bases. A base decimal é a que mais empregamos atualmente. Na Antiguidade, foram utilizadas outras bases, como a base 12, a base 60. Um computador opera normalmente no sistema binário.

Observe o que acontece na interação entre o usuário e o computador: os dados de entrada são enviados ao computador pelo usuário no sistema decimal; toda esta informação é convertida para o sistema binário, e as operações todas serão efetuadas neste sistema.

Os resultados finais serão convertidos para o sistema decimal e, finalmente, serão transmitidos ao usuário. Todo este processo de conversão é uma fonte de erros que afetam o resultado final dos cálculos.

Na próxima seção, estudaremos os processos para conversão de números do sistema binário para o sistema decimal e vice-versa.

1.2.1 CONVERSÃO DE NÚMEROS NOS SISTEMAS DECIMAL E BINÁRIO

Veremos inicialmente a conversão de números inteiros.

Considere os números $(347)_{10}$ e $(10111)_2$. Estes números podem ser assim escritos:

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

De um modo geral, um número na base β, $(a_j a_{j-1} \ldots a_2 a_1 a_0)_\beta$, $0 \leq a_k \leq (\beta - 1)$, $k = 1, \ldots, j$, pode ser escrito na forma polinomial:

$$a_j\beta^j + a_{j-1}\beta^{j-1} + \ldots + a_2\beta^2 + a_1\beta^1 + a_0\beta^0.$$

Com esta representação, podemos facilmente converter um número representado no sistema binário para o sistema decimal.

Por exemplo:

$$(10111)_2 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

Colocando agora o número 2 em evidência teremos:

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

Deste exemplo, podemos obter um processo para converter um número representado no sistema binário para o sistema decimal:

A representação do número $(a_j a_{j-1} \ldots a_2 a_1 a_0)_2$ na base 10, denotada por b_0, é obtida através do processo:

$$b_j = a_j$$
$$b_{j-1} = a_{j-1} + 2b_j$$

$$b_{j-2} = a_{j-2} + 2b_{j-1}$$
$$\vdots \qquad \vdots$$
$$b_1 = a_1 + 2b_2$$
$$b_0 = a_0 + 2b_1$$

Para $(10111)_2$, a seqüência obtida será:

$$b_4 = a_4 = 1$$
$$b_3 = a_3 + 2b_4 = 0 + 2 \times 1 = 2$$
$$b_2 = a_2 + 2b_3 = 1 + 2 \times 2 = 5$$
$$b_1 = a_1 + 2b_2 = 1 + 2 \times 5 = 11$$
$$b_0 = a_0 + 2b_1 = 1 + 2 \times 11 = 23$$

Veremos agora um processo para converter um número inteiro representado no sistema decimal para o sistema binário. Considere o número $N_0 = (347)_{10}$ e $(a_j a_{j-1} \ldots a_1 a_0)_2$ a sua representação na base 2.

Temos então que:

$$347 = 2 \times (a_j \times 2^{j-1} + a_{j-1} \times 2^{j-2} + \ldots + a_2 \times 2 + a_1) + a_0 = 2 \times 173 + 1$$

e, portanto, o dígito $a_0 = 1$ representa o resto da divisão de 347 por 2. Repetindo agora este processo para o número $N_1 = 173$:

$$173 = a_j \times 2^{j-1} + a_{j-1} \times 2^{j-2} + \ldots + a_2 \times 2 + a_1$$

obteremos o dígito a_1, que será o resto da divisão de N_1 por 2. Seguindo este raciocínio obtemos a seqüência de números N_j e a_j.

$$N_0 = 347 = 2 \times 173 + 1 \Rightarrow a_0 = 1$$
$$N_1 = 173 = 2 \times 86 + 1 \Rightarrow a_1 = 1$$
$$N_2 = 86 = 2 \times 43 + 0 \Rightarrow a_2 = 0$$
$$N_3 = 43 = 2 \times 21 + 1 \Rightarrow a_3 = 1$$
$$N_4 = 21 = 2 \times 10 + 1 \Rightarrow a_4 = 1$$
$$N_5 = 10 = 2 \times 5 + 0 \Rightarrow a_5 = 0$$
$$N_6 = 5 = 2 \times 2 + 1 \Rightarrow a_6 = 1$$

$$N_7 = 2 = 2 \times 1 + 0 \Rightarrow a_7 = 0$$
$$N_8 = 1 = 2 \times 0 + 1 \Rightarrow a_8 = 1$$

Portanto, a representação de $(347)_{10}$ na base 2 será 101011011.

No caso geral, considere um número inteiro N na base 10 e a sua representação binária denotada por: $(a_j a_{j-1} \ldots a_2 a_1 a_0)_2$. O algoritmo a seguir obtém a cada k o dígito binário a_k.

Passo 0: $k = 0$
$N_k = N$

Passo 1: Obtenha q_k e r_k tais que:
$N_k = 2 \times q_k + r_k$
Faça $a_k = r_k$

Passo 2: Se $q_k = 0$, pare.
Caso contrário, faça $N_{k+1} = q_k$.
Faça $k = k + 1$ e volte para o passo 1.

Consideremos agora a conversão de um número fracionário da base 10 para a base 2.

Sejam, por exemplo:

$r = 0.125$, $s = 0.66666...$, $t = 0.414213562...$.

Dizemos que r tem representação finita e que s e t têm representação infinita.

Dado um número entre 0 e 1 no sistema decimal, como obter sua representação binária?

Considerando o número $r = 0.125$, existem dígitos binários: $d_1, d_2, \ldots, d_j \ldots$, tais que $(0.\ d_1 d_2 \ldots d_j \ldots)_2$ será sua representação na base 2.

Assim,

$$(0.125)_{10} = d_1 \times 2^{-1} + d_2 \times 2^{-2} + \ldots + d_j \times 2^{-j} + \ldots$$

Multiplicando cada termo da expressão acima por 2 teremos:

$$2 \times 0.125 = 0.250 = 0 + 0.25 = d_1 + d_2 \times 2^{-1} + d_3 \times 2^{-2} + \ldots + d_j \times 2^{-j+1} + \ldots$$

e, portanto, d_1 representa a parte inteira de 2×0.125 que é igual a zero e $d_2 \times 2^{-1} + d_3 \times 2^{-2} + \ldots + d_j \times 2^{-j+1} + \ldots$ representa a parte fracionária de 2×0.125 que é 0.250.

Aplicando agora o mesmo procedimento para 0.250, teremos:

$$0.250 = d_2 \times 2^{-1} + d_3 \times 2^{-2} + \ldots + d_j \times 2^{-j+1} + \ldots \Rightarrow 2 \times 0.250 = 0.5 =$$
$$= d_2 + d_3 \times 2^{-1} + d_4 \times 2^{-2} + \ldots + d_j \times 2^{-j+2} + \ldots \Rightarrow d_2 = 0$$

e repetindo o processo para 0.5:

$$0.5 = d_3 \times 2^{-1} + d_4 \times 2^{-2} + \ldots + d_j \times 2^{-j+2} + \ldots \Rightarrow$$
$$2 \times 0.5 = 1.0 = d_3 + d_4 \times 2^{-1} + \ldots + d_j \times 2^{-j+3} + \ldots \Rightarrow d_3 = 1$$

Como a parte fracionária de 2×0.5 é zero, o processo de conversão termina, e teremos: $d_1 = 0$, $d_2 = 0$ e $d_3 = 1$ e, portanto, o número $(0.125)_{10}$ tem representação finita na base 2: $(0.001)_2$

Um número real entre 0 e 1 pode ter representação finita no sistema decimal, mas representação *infinita* no sistema binário.

No caso geral, seja *r* um número entre 0 e 1 no sistema decimal e $(0.\, d_1 d_2 \ldots d_j \ldots)_2$ sua representação no sistema binário.

Os dígitos binários $d_1, d_2, \ldots, d_j, \ldots$ são obtidos através do seguinte algoritmo:

Passo 0: $r_1 = r$; $k = 1$

Passo 1: Calcule $2r_k$.
Se $2r_k \geq 1$, faça: $d_k = 1$,
caso contrário, faça: $d_k = 0$

Passo 2: Faça $r_{k+1} = 2r_k - d_k$.
Se $r_{k+1} = 0$, pare.
Caso contrário:

Passo 3: $k = k + 1$.
Volte ao passo 1.

Observar que o algoritmo pode ou não parar após um número finito de passos. Para $r = (0.125)_{10}$ teremos $r_4 = 0$. Já para $r = (0.1)_{10}$, teremos: $r_1 = 0.1$

$$\begin{aligned}
k = 1 \quad 2r_1 = 0.2 \Rightarrow \quad & d_1 = 0 \\
& r_2 = 0.2 \\
k = 2 \quad 2r_2 = 0.4 \Rightarrow \quad & d_2 = 0 \\
& r_3 = 0.4 \\
k = 3 \quad 2r_3 = 0.8 \Rightarrow \quad & d_3 = 0 \\
& r_4 = 0.8 \\
k = 4 \quad 2r_4 = 1.6 \Rightarrow \quad & d_4 = 1 \\
& r_5 = 0.6 \\
k = 5 \quad 2r_5 = 1.2 \Rightarrow \quad & d_5 = 1 \\
& r_6 = 0.2 = r_2
\end{aligned}$$

Como $r_6 = r_2$, teremos que os resultados para k de 2 a 5 se repetirão e então: $r_{10} = r_6 = r_2 = 0.2$ e assim indefinidamente.

Concluímos que:

$$(0.1)_{10} = (0.0001100110011\overline{0011}...)_2$$

e, portanto, o número $(0.1)_{10}$ não tem representação binária finita.

O fato de um número não ter representação finita no sistema binário pode acarretar a ocorrência de erros aparentemente inexplicáveis em cálculos efetuados em sistemas computacionais binários.

Analisando o Exemplo 2 da Seção 1.2 e usando o processo de conversão descrito anteriormente, temos que o número $(0.5)_{10}$ tem representação finita no sistema binário: $(0.1)_2$; já o número $(0.11)_{10}$ terá representação infinita:

$$(0.000111000010100011110\overline{1011100001010001111101}...)_2$$

Um computador que opera no sistema binário irá armazenar uma aproximação para $(0.11)_{10}$, uma vez que possui uma quantidade fixa de posições para guardar os dígitos da mantissa de um número, e esta aproximação será usada para realizar os cálculos. Não se pode, portanto, esperar um resultado exato.

Considere agora um número entre 0 e 1 representado no sistema binário:

$(r)_2 = (0.\ d_1d_2...d_j...)_2$

Como obter sua representação no sistema decimal?

Um processo para conversão é equivalente ao que descrevemos anteriormente. Definindo $r_1 = r$, a cada iteração k, o processo de conversão multiplica o número r_k por $(10)_{10} = (1010)_2$ e obtém o dígito b_k como sendo a parte inteira deste produto convertida para a base decimal. É importante observar que as operações devem ser efetuadas no sistema binário, [26]. O algoritmo a seguir formaliza este processo.

Passo 0: $r_1 = r$; $k = 1$

Passo 1: Calcule $w_k = (1010)_2 \times r_k$.
Seja z_k a parte inteira de w_k
b_k é a conversão de z_k para a base 10.

Passo 2: Faça $r_{k+1} = w_k - z_k$
Se $r_{k+1} = 0$, pare.

Passo 3: $k = k + 1$.
Volte ao passo 1.

Por exemplo, considere o número:

$(r)_2 = (0.000111)_2 = (0.b_1b_2 \ldots b_j)_{10}$

Seguindo o algoritmo, teremos:

$r_1 = (0.000111)_2$;
$w_1 = (1010)_2 \times r_1 = 1.00011 \Rightarrow b_1 = 1$ e $r_2 = 0.00011$;
$w_2 = (1010)_2 \times r_2 = 0.1111 \Rightarrow b_2 = 0$ e $r_3 = 0.1111$;
$w_3 = (1010)_2 \times r_3 = 1001.011 \Rightarrow b_3 = 9$ e $r_4 = 0.011$;
$w_4 = (1010)_2 \times r_4 = 11.11 \Rightarrow b_4 = 3$ e $r_5 = 0.11$;
$w_5 = (1010)_2 \times r_5 = 111.1 \Rightarrow b_5 = 7$ e $r_6 = 0.1$;
$w_6 = (1010)_2 \times r_6 = 101 \Rightarrow b_6 = 5$ e $r_7 = 0$;

Portanto $(0.000111)_2 = (0.109375)_{10}$

Podemos agora entender melhor por que o resultado da operação

$$S = \sum_{i=1}^{30000} 0.11$$

não é obtido exatamente num computador. Já vimos que $(0.11)_{10}$ não tem representação finita no sistema binário. Supondo um computador que trabalhe com apenas 6 dígitos na mantissa, o número $(0.11)_{10}$ seria armazenado como $(0.000111)_2$ e este número representa exatamente $(0.109375)_{10}$. Portanto, todas as operações que envolvem o número $(0.11)_{10}$ seriam realizadas com esta aproximação. Veremos na próxima seção a representação de números em aritmética de ponto flutuante com o objetivo de se entender melhor a causa de resultados imprecisos em operações numéricas.

1.2.2 ARITMÉTICA DE PONTO FLUTUANTE

Um computador ou calculadora representa um número real no sistema denominado *aritmética de ponto flutuante*. Neste sistema, o número r será representado na forma:

$$\pm (. d_1 d_2 ... d_t) \times \beta^e$$

onde:

β é a base em que a máquina opera;
t é o número de dígitos na mantissa; $0 \leq d_j \leq (\beta - 1)$, $j = 1, ..., t$, $d_1 \neq 0$;
e é o expoente no intervalo [l, u].

Em qualquer máquina, apenas um subconjunto dos números reais é representado exatamente, e, portanto, a representação de um número real será realizada através de truncamento ou de arredondamento.

Considere, por exemplo, uma máquina que opera no sistema:

$\beta = 10$; $t = 3$; $e \in [-5, 5]$.

Os números serão representados na seguinte forma nesse sistema:

$0. d_1 d_2 d_3 \times 10^e$, $0 \leq d_j \leq 9$, $d_1 \neq 0$, $e \in [-5, 5]$

O menor número, em valor absoluto, representado nesta máquina é:

$$m = 0.100 \times 10^{-5} = 10^{-6}$$

e o maior número, em valor absoluto, é:

$$M = 0.999 \times 10^{5} = 99900$$

Considere o conjunto dos números reais $\mathbb{R}$ e o seguinte conjunto:

$$G = \{x \in \mathbb{R} \mid m \leqslant |x| \leqslant M\}$$

Dado um número real x, várias situações poderão ocorrer:

Caso 1) $x \in G$:

por exemplo: $x = 235.89 = 0.23589 \times 10^{3}$. Observe que este número possui 5 dígitos na mantissa. Estão representados exatamente nesta máquina os números: 0.235×10^{3} e 0.236×10^{3}. Se for usado o truncamento, x será representado por 0.235×10^{3} e, se for usado o arredondamento, x será representado por 0.236×10^{3} (o truncamento e o arredondamento serão estudados na Seção 1.3.2.);

Caso 2) $|x| < m$:

por exemplo: $x = 0.345 \times 10^{-7}$. Este número não pode ser representado nesta máquina porque o expoente *e* é menor que –5. Esta é uma situação em que a máquina acusa a ocorrência de *underflow*;

Caso 3) $|x| > M$:

por exemplo: $x = 0.875 \times 10^{9}$. Neste caso, o expoente *e* é maior que 5 e a máquina acusa a ocorrência de *overflow*.

Algumas linguagens de programação permitem que as variáveis sejam declaradas em *precisão dupla*. Neste caso, esta variável será representada no sistema de aritmética de ponto flutuante da máquina, mas com aproximadamente o dobro de dígitos disponíveis na mantissa. É importante observar que, neste caso, o tempo de execução e requerimentos de memória aumentam de forma significativa.

O *zero* em ponto flutuante é, em geral, representado com o menor expoente possível na máquina. Isto porque a representação do zero por uma mantissa nula e um expoente qualquer para a base β pode acarretar perda de dígitos significativos no resultado da adição deste *zero* a um outro número. Por exemplo, em uma máquina que opera na base 10 com 4 dígitos na mantissa, para $x = 0.0000 \times 10^4$ e $y = 0.3134 \times 10^{-2}$ o resultado de $x + y$ seria 0.3100×10^{-2}, isto é, são perdidos dois dígitos do valor exato y. Este resultado se deve à forma como é efetuada a adição em ponto flutuante, que estudaremos na Seção 1.3.3.

Exemplo 3

Dar a representação dos números a seguir num sistema de aritmética de ponto flutuante de três dígitos para $\beta = 10$, $m = -4$ e $M = 4$.

x	Representação obtida por arredondamento	Representação obtida por truncamento
1.25	0.125×10	0.125×10
10.053	0.101×10^2	0.100×10^2
–238.15	-0.238×10^3	-0.238×10^3
2.71828...	0.272×10	0.271×10
0.000007	(expoente menor que –4)	=
718235.82	(expoente maior que 4)	=

1.3 ERROS

1.3.1 ERROS ABSOLUTOS E RELATIVOS

No Exemplo 1 da Seção 1.2, diferentes resultados para a área da circunferência foram obtidos, dependendo da aproximação adotada para o valor de π.

Definimos como *erro absoluto* a diferença entre o valor exato de um número x e de seu valor aproximado $\bar{x}$: $EA_x = x - \bar{x}$.

Em geral, apenas o valor $\bar{x}$ é conhecido, e, neste caso, é impossível obter o valor exato do erro absoluto. O que se faz é obter um limitante superior ou uma estimativa para o módulo do erro absoluto.

Por exemplo, sabendo-se que $\pi \in (3.14, 3.15)$ tomaremos para π um valor dentro deste intervalo e teremos, então, $|EA_{\pi}| = |\pi - \bar{\pi}| < 0.01$.

Seja agora o número x representado por $\bar{x} = 2112.9$ de tal forma que $|EA_x| < 0.1$, ou seja, $x \in (2112.8, 2113)$ e seja o número y representado por $\bar{y} = 5.3$ de tal forma que $|EA_y| < 0.1$, ou seja, $y \in (5.2, 5.4)$. Os limitantes superiores para os erros absolutos são os mesmos. Podemos dizer que ambos os números estão representados com a mesma precisão?

É preciso comparar a ordem de grandeza de x e y. Feito isto, é fácil concluir que o primeiro resultado é mais preciso que o segundo, pois a ordem de grandeza de x é maior que a ordem de grandeza de y. Então, dependendo da ordem de grandeza dos números envolvidos, o erro absoluto não é suficiente para descrever a precisão de um cálculo. Por esta razão, o erro relativo é amplamente empregado.

O *erro relativo* é definido como o erro absoluto dividido pelo valor aproximado:

$$ER_x = \frac{EA_x}{\bar{x}} = \frac{x - \bar{x}}{\bar{x}}$$

No exemplo anterior, temos

$$|ER_x| = \frac{|EA_x|}{|\bar{x}|} < \frac{0.1}{2112.9} \approx 4.7 \times 10^{-5}$$

e

$$|ER_y| = \frac{|EA_y|}{|\bar{y}|} < \frac{0.1}{5.3} \approx 0.02,$$

confirmando, portanto, que o número x é representado com maior precisão que o número y.

1.3.2 ERROS DE ARREDONDAMENTO E TRUNCAMENTO EM UM SISTEMA DE ARITMÉTICA DE PONTO FLUTUANTE

Vimos na Seção 1.2 que a representação de um número depende fundamentalmente da máquina utilizada, pois seu sistema estabelecerá a base numérica adotada, o total de dígitos na mantissa etc.

Seja um sistema que opera em aritmética de ponto flutuante de t dígitos na base 10, e seja x, escrito na forma

$$x = f_x \times 10^e + g_x \times 10^{e-t} \text{ onde } 0.1 \leqslant f_x < 1 \text{ e } 0 \leqslant g_x < 1.$$

Por exemplo, se t = 4 e x = 234.57, então

$$x = 0.2345 \times 10^3 + 0.7 \times 10^{-1}, \text{ donde } f_x = 0.2345 \text{ e } g_x = 0.7.$$

É claro que na representação de x neste sistema $g_x \times 10^{e-t}$ não pode ser incorporado totalmente à mantissa. Então, surge a questão de como considerar esta parcela na mantissa e definir o erro absoluto (ou relativo) máximo cometido.

Vimos na Seção 1.2.2 que podem ser adotados dois critérios: o do arredondamento e o do truncamento (ou cancelamento).

No *truncamento*, $g_x \times 10^{e-t}$ é desprezado e $\bar{x} = f_x \times 10^e$. Neste caso, temos

$$|EA_x| = |x - \bar{x}| = |g_x| \times 10^{e-t} < 10^{e-t}, \text{ visto que } |g_x| < 1$$

$$|ER_x| = \frac{|EA_x|}{|\bar{x}|} = \frac{|g_x| \times 10^{e-t}}{|f_x| \times 10^e} < \frac{10^{e-t}}{0.1 \times 10^e} = 10^{-t+1}$$

visto que 0.1 é o menor valor possível para f_x.

No *arredondamento*, f_x é modificado para levar em consideração g_x. A forma de arredondamento mais utilizada é o arredondamento simétrico:

$$\bar{x} = \begin{cases} f_x \times 10^e & \text{se } |g_x| < \frac{1}{2} \\ f_x \times 10^e + 10^{e-t} & \text{se } |g_x| \geqslant \frac{1}{2} \end{cases}$$

Portanto se $|g_x| < \frac{1}{2}$, g_x é desprezado, caso contrário, somamos o número 1 ao último dígito de f_x.

Então, se $|g_x| < \frac{1}{2}$ teremos

$$|EA_x| = |x - \bar{x}| = |g_x| \times 10^{e-t} < \frac{1}{2} \times 10^{e-t}$$

e

$$|ER_x| = \frac{|EA_x|}{|\bar{x}|} = \frac{|g_x| \times 10^{e-t}}{|f_x| \times 10^e} < \frac{0.5 \times 10^{e-t}}{0.1 \times 10^e} = \frac{1}{2} \times 10^{-t+1}$$

E se $|g_x| \geqslant 1/2$, teremos

$$\begin{aligned} |EA_x| &= |x - \bar{x}| = |(f_x \times 10^e + g_x \times 10^{e-t}) - (f_x \times 10^e + 10^{e-t})| \\ &= |g_x \times 10^{e-t} - 10^{e-t}| = |(g_x - 1)| \times 10^{e-t} \leqslant \frac{1}{2} \times 10^{e-t} \end{aligned}$$

e

$$|ER_x| = \frac{|EA_x|}{|\bar{x}|} \leqslant \frac{1/2 \times 10^{e-t}}{|f_x \times 10^e + 10^{e-t}|} < \frac{1/2 \times 10^{e-t}}{|f_x| \times 10^e} <$$

$$< \frac{1/2 \times 10^{e-t}}{0.1 \times 10^e} = \frac{1}{2} \times 10^{-t+1}$$

Portanto, em qualquer caso teremos

$$|EA_x| \leqslant \frac{1}{2} \times 10^{e-t} \quad \text{e} \quad |ER_x| < \frac{1}{2} \times 10^{-t+1}$$

Apesar de incorrer em erros menores, o uso do arredondamento acarreta um tempo maior de execução e por esta razão o truncamento é mais utilizado.

1.3.3 ANÁLISE DE ERROS NAS OPERAÇÕES ARITMÉTICAS DE PONTO FLUTUANTE

Dada uma seqüência de operações, como, por exemplo, $u = [(x + y) - z - t] + w$, é importante a noção de como o erro em uma operação propaga-se ao longo das operações subseqüentes.

O erro total em uma operação é composto pelo erro das parcelas ou fatores e pelo erro no resultado da operação.

Nos exemplos a seguir, vamos supor que as operações são efetuadas num sistema de aritmética de ponto flutuante de quatro dígitos, na base 10, e com acumulador de precisão dupla.

Exemplo 4

Dados $x = 0.937 \times 10^4$ e $y = 0.1272 \times 10^2$, obter $x + y$.

A adição em aritmética de ponto flutuante requer o alinhamento dos pontos decimais dos dois números. Para isto, a mantissa do número de menor expoente deve ser deslocada para a direita. Este deslocamento deve ser de um número de casas decimais igual à diferença entre os dois expoentes.

Alinhando os pontos decimais dos valores acima, temos

$$x = 0.937 \times 10^4 \text{ e } y = 0.001272 \times 10^4.$$

Então,

$$x + y = (0.937 + 0.001272) \times 10^4 = 0.938272 \times 10^4.$$

Este é o resultado exato desta operação. Dado que em nosso sistema t é igual a 4, este resultado dever ser arredondado ou truncado.

Então, $\overline{x + y} = 0.9383 \times 10^4$ no arredondamento e $\overline{x + y} = 0.9382 \times 10^4$ no truncamento.

Exemplo 5

Sejam x e y do Exemplo 4. Obter xy:

$$\begin{aligned} xy &= (0.937 \times 10^4) \times (0.1272 \times 10^2) \\ &= (0.937 \times 0.1272) \times 10^6 = \\ &= 0.1191864 \times 10^6. \end{aligned}$$

Então, $\overline{xy} = 0.1192 \times 10^6$, se for efetuado o arredondamento, e $\overline{xy} = 0.1191 \times 10^6$, se for efetuado o truncamento.

Os Exemplos 4 e 5 mostram que ainda que as parcelas ou fatores de uma operação estejam representados exatamente no sistema, não se pode esperar que o resultado armazenado seja exato.

Na maioria dos sistemas, o resultado exato da operação que denotaremos por OP é normalizado e em seguida arredondado ou truncado para t dígitos, obtendo assim o resultado aproximado $\overline{OP}$ que é armazenado na memória da máquina.

Então, conforme vimos na Seção 1.3.2, o erro relativo no resultado de uma operação (supondo que as parcelas ou fatores estão representados exatamente) será:

$|ER_{OP}| < 10^{-t+1}$ no truncamento

$|ER_{OP}| < \frac{1}{2} \times 10^{-t+1}$ no arredondamento.

Veremos a seguir as fórmulas para os erros absoluto e relativo nas operações aritméticas com erros nas parcelas ou fatores.

Vamos supor que o erro final é arredondado.

Sejam x e y, tais que $x = \bar{x} + EA_x$ e $y = \bar{y} + EA_y$.

Adição: $x + y$

$$x + y = (\bar{x} + EA_x) + (\bar{y} + EA_y) = (\bar{x} + \bar{y}) + (EA_x + EA_y).$$

Então, o erro absoluto na soma, denotado por EA_{x+y} é a soma dos erros absolutos das parcelas:

$$EA_{x+y} = EA_x + EA_y.$$

O erro relativo será:

$$ER_{x+y} = \frac{EA_{x+y}}{\bar{x} + \bar{y}} = \frac{EA_x}{\bar{x}}\left(\frac{\bar{x}}{\bar{x} + \bar{y}}\right) + \frac{EA_y}{\bar{y}}\left(\frac{\bar{y}}{\bar{x} + \bar{y}}\right) =$$

$$= ER_x\left(\frac{\bar{x}}{\bar{x} + \bar{y}}\right) + ER_y\left(\frac{\bar{y}}{\bar{x} + \bar{y}}\right).$$

Subtração: $x - y$

Analogamente temos:

$$EA_{x-y} = EA_x - EA_y \text{ e}$$

$$ER_{x-y} = \frac{EA_x - EA_y}{\bar{x} - \bar{y}} = ER_x\left(\frac{\bar{x}}{\bar{x} - \bar{y}}\right) - ER_y\left(\frac{\bar{y}}{\bar{x} - \bar{y}}\right).$$

Multiplicação: xy

$$xy = (\bar{x} + EA_x)(\bar{y} + EA_y) =$$
$$= \bar{x}\,\bar{y} + \bar{x}\,EA_y + \bar{y}\,EA_x + (EA_x)(EA_y)$$

Considerando que $(EA_x)(EA_y)$ é um número pequeno, podemos desprezar este termo na expressão acima.

Teremos então $EA_{xy} \approx \bar{x}\,EA_y + \bar{y}\,EA_x$

$$ER_{xy} \approx \frac{\bar{x}\,EA_y + \bar{y}\,EA_x}{\bar{x}\,\bar{y}} = \frac{EA_x}{\bar{x}} + \frac{EA_y}{\bar{y}} = ER_x + ER_y.$$

Divisão: x/y

$$\frac{x}{y} = \frac{\bar{x} + EA_x}{\bar{y} + EA_y} = \frac{\bar{x} + EA_x}{\bar{y}}\left(\frac{1}{1 + \dfrac{EA_y}{\bar{y}}}\right).$$

Representando o fator $\dfrac{1}{1 + \dfrac{EA_y}{\bar{y}}}$ sob a forma de uma série infinita, teremos

$$\frac{1}{1 + \dfrac{EA_y}{\bar{y}}} = 1 - \frac{EA_y}{\bar{y}} + \left(\frac{EA_y}{\bar{y}}\right)^2 - \left(\frac{EA_y}{\bar{y}}\right)^3 + \ldots$$

e desprezando os termos com potências maiores que 1, teremos

$$\frac{x}{y} \approx \frac{\bar{x} + EA_x}{\bar{y}} \left(1 - \frac{EA_y}{\bar{y}}\right) = \frac{\bar{x}}{\bar{y}} + \frac{EA_x}{\bar{y}} - \frac{\bar{x}\,EA_y}{\bar{y}^2} - \frac{EA_x\,EA_y}{\bar{y}^2}.$$

Então

$$\frac{x}{y} \approx \frac{\bar{x}}{\bar{y}} + \frac{EA_x}{\bar{y}} - \frac{\bar{x}\,EA_y}{\bar{y}^2}.$$

$$\text{Assim, } EA_{x/y} \approx \frac{EA_x}{\bar{y}} - \frac{\bar{x}\,EA_y}{\bar{y}^2} = \frac{\bar{y}\,EA_x - \bar{x}\,EA_y}{\bar{y}^2}$$

e

$$ER_{x/y} \approx \left(\frac{\bar{y}\,EA_x - \bar{x}\,EA_y}{\bar{y}^2}\right)\frac{\bar{y}}{\bar{x}} = \frac{EA_x}{\bar{x}} - \frac{EA_y}{\bar{y}} = ER_x - ER_y$$

Escrevemos todas essas fórmulas sem considerar o erro de arredondamento ou truncamento no resultado final.

A análise completa da propagação de erros se faz considerando os erros nas parcelas ou fatores e no resultado de cada operação efetuada.

Exemplo 6

Supondo que x, y, z e t estejam representados exatamente, qual o erro total no cálculo de u = (x + y)z – t? (Calcularemos o erro relativo e denotaremos por RA o erro relativo de arredondamento no resultado da operação.)

Seja s = x + y. O erro relativo nesta operação será:

$$ER_s = ER_x \left(\frac{\bar{x}}{\bar{x} + \bar{y}}\right) + ER_y \left(\frac{\bar{y}}{\bar{x} + \bar{y}}\right) + RA = 0 + RA .$$

Assim, $|ER_s| = |RA| < \frac{1}{2} \times 10^{-t+1}$.

Calculando agora s × z, teremos m = s × z, e o erro relativo desta operação será:

$$ER_m = ER_s + ER_z + RA = ER_s + \frac{EA_z}{\bar{z}} + RA = RA_s + 0 + RA.$$

Então,

$$|ER_m| \leq |RA_s| + |RA| < \frac{1}{2} \times 10^{-t+1} + \frac{1}{2} \times 10^{-t+1} = 10^{-t+1}.$$

Calcularemos u = m – t e o erro relativo desta operação será:

$$ER_u = \frac{EA_m - EA_t}{\bar{m} - \bar{t}} + RA = \frac{EA_m}{\bar{m} - \bar{t}} + RA = \frac{EA_m}{\bar{m}} \left(\frac{\bar{m}}{\bar{m} - \bar{t}}\right) + RA$$

$$= ER_m \left(\frac{\bar{m}}{\bar{m} - \bar{t}}\right) + RA.$$

Então,

$$|ER_u| \leq |ER_m| \left|\frac{\bar{m}}{\bar{m} - \bar{t}}\right| + RA < 10^{-t+1} \left|\frac{\bar{m}}{\bar{m} - \bar{t}}\right| + \frac{1}{2} \times 10^{-t+1}$$

Finalmente,

$$|ER_u| < \left(\frac{|\bar{m}|}{|\bar{m} - \bar{t}|} + \frac{1}{2}\right) \times 10^{-t+1} .$$

Exemplo 7

Vimos que dados x, y e z = x – y, então:

$$ER_z = \frac{EA_x - EA_y}{\bar{x} - \bar{y}} = \frac{EA_x}{\bar{x}}\left(\frac{\bar{x}}{\bar{x} - \bar{y}}\right) - \frac{EA_y}{\bar{y}}\left(\frac{\bar{y}}{\bar{x} - \bar{y}}\right).$$

Se x e y são números positivos arredondados, então:

$$\left|\frac{EA_x}{\bar{x}}\right| < \frac{1}{2} \times 10^{-t+1} \quad \text{e} \quad \left|\frac{EA_y}{\bar{y}}\right| < \frac{1}{2} \times 10^{-t+1}.$$

É claro que se $x \approx y$, então o erro relativo em z pode ser grande, como por exemplo, se $\bar{x} = 0.2357 \times 10^3$ e $\bar{y} = 0.2353 \times 10^3$, então $\bar{z} = \bar{x} - \bar{y} = 0.0004 \times 10^3$ e isto é denominado *cancelamento subtrativo.*

O erro relativo em z é limitado superiormente por:

$$|ER_z| < \left(\frac{0.2357 \times 10^3 + 0.2353 \times 10^3}{0.0004 \times 10^3}\right) \times \frac{1}{2} \times 10^{-3} \ (t = 4) \Rightarrow$$

$$\Rightarrow |ER_z| < 0.5888 \approx 59\%.$$

Este erro relativo é propagado nas próximas operações (pois cada expressão para o erro relativo depende do erro relativo em cada parcela ou fator).

Por exemplo, para w = zt, se $\bar{t} = 0.4537 \times 10^3$ teremos $\bar{w} = 0.1815 \times 10^3$ e

$$|ER_w| \leq |ER_z| + |ER_t| < 0.59 + \frac{1}{2} \times 10^{-3} = 0.5905 \approx 59\%.$$

Desta forma, o cancelamento subtrativo pode dar origem a grandes erros nas operações seguintes.

EXERCÍCIOS

1. Converta os seguintes números decimais para sua forma binária:

   $x = 37$ $y = 2345$ $z = 0.1217$

2. Converta os seguintes números binários para sua forma decimal:

   $x = (101101)_2$ $y = (110101011)_2$

   $z = (0.1101)_2$ $w = (0.111111101)_2$

3. Seja um sistema de aritmética de ponto flutuante de quatro dígitos, base decimal e com acumulador de precisão dupla. Dados os números:

   $x = 0.7237 \times 10^4$ $y = 0.2145 \times 10^{-3}$ e $z = 0.2585 \times 10^1$

   efetue as seguintes operações e obtenha o erro relativo no resultado, supondo que x, y e z estão exatamente representados:

   a) $x + y + z$ *d*) $(xy)/z$

   b) $x - y - z$ *e*) $x(y/z)$

   c) x/y

4. Supondo que x é representado num computador por $\overline{x}$, onde $\overline{x}$ é obtido por arredondamento, obtenha os limites superiores para os erros relativos de $u = 2\overline{x}$ e $w = \overline{x} + \overline{x}$.

5. Idem para $u = 3\overline{x}$ e $w = \overline{x} + \overline{x} + \overline{x}$.

6. Sejam $\overline{x}$ e $\overline{y}$ as representações de x e y obtidas por arredondamento em um computador. Deduza expressões de limitante de erro para mostrar que o limitante do erro relativo de $u = 3\,\overline{x}\,\overline{y}$ é menor que o de $v = (\overline{x} + \overline{x} + \overline{x})\,\overline{y}$.

7. Verifique de alguma forma que, se r_F tem representação finita na base 2 com k dígitos, ou seja, $r_F = (0.\ d_1d_2...d_k)_2$, então sua representação na base 10 é também finita com k dígitos.

8. É interessante observar que a adição em aritmética de ponto flutuante nem sempre é associativa. Prova-se que dados n números $x_1, x_2, ..., x_n$ representados exatamente, o limite do erro total devido ao arredondamento na soma $S = x_1 + x_2 + ... x_n$ é:

$$|EA_s| \leq [(n-1)x_1 + (n-1)x_2 + (n-2)x_3 + ... 2x_{n-1} + x_n] \times \frac{1}{2} \times 10^{-t+1}.$$

a) Faça n = 5 e prove que esta relação é verdadeira.

b) Analise cuidadosamente a expressão acima e verifique que, dados n números, o erro total será menor se forem ordenados em ordem crescente antes de serem somados.

9. Considere uma máquina cujo sistema de representação de números é definido por: $\beta = 10$, $t = 4$, $l = -5$ e $u = 5$. Pede-se:

a) qual o menor e o maior número em módulo representados nesta máquina?

b) como será representado o número 73.758 nesta máquina, se for usado o arredondamento? E se for usado o truncamento?

c) se a = 42450 e b = 3 qual o resultado de a + b?

d) qual o resultado da soma

$$S = 42450 + \sum_{k=1}^{10} 3$$

nesta máquina?

e) idem para a soma:

$$S = \sum_{k=1}^{10} 3 + 42450.$$

(Obviamente o resultado deveria ser o mesmo. Contudo, as operações devem ser realizadas na ordem em que aparecem as parcelas, o que conduzirá a resultados distintos).

f) o resultado da operação: wz/t pode ser obtido de várias maneiras, bastando modificar a ordem em que os cálculos são efetuados. Para determinados valores de w, z e t, uma seqüência de cálculos pode ser melhor que outra. Faça uma análise para o caso em que w = 100, z = 3500 e t = 7.

10. Escreva um programa em alguma linguagem para obter o resultado da seguinte operação:

$$S = 10000 - \sum_{k=1}^{n} x$$

para: *a)* n = 100000 e x = 0.1; *b)* n = 80000 e x = 0.125.

PROJETOS

1. PRECISÃO DA MÁQUINA

A precisão da máquina é definida como sendo o menor número positivo em aritmética de ponto flutuante ε, tal que $(1 + \varepsilon) > 1$. Este número depende totalmente do sistema de representação da máquina: base numérica, total de dígitos na mantissa, da forma como são realizadas as operações e do compilador utilizado. É importante conhecermos a *precisão da máquina* porque em vários algoritmos precisamos fornecer como dado de entrada um valor positivo, próximo de zero, para ser usado em testes de comparação com zero.

O algoritmo a seguir estima a precisão da máquina:

Passo 1: A = 1;
s = 2;

Passo 2: Enquanto (s) > 1, faça:
A = A/2
s = 1 + A;

Passo 3: Faça Prec = 2A e imprima Prec.

a) Teste este algoritmo escrevendo um programa usando uma linguagem conhecida. Declare as variáveis do programa em precisão simples e execute o programa; em seguida, declare as variáveis em precisão dupla e execute novamente o programa.

b) Interprete o passo 3 do algoritmo, isto é, por que a aproximação para Prec é escolhida como sendo o dobro do último valor de A obtido no passo 2?

c) Na definição de precisão da máquina, usamos como referência o número 1. No algoritmo a seguir, a variável VAL é um dado de entrada, escolhido pelo usuário:

Passo 1: A = 1;
s = VAL + A;

Passo 2: Enquanto (s) > VAL, faça:
A = A/2
s = VAL + A

Passo 3: Faça Prec = 2A e imprima Prec.

c.1) Teste seu programa atribuindo para VAL os números: 10, 17, 100, 184, 1000, 1575, 10000, 17893.

c.2) Para cada valor diferente para VAL, imprima o valor correspondente obtido para Prec. Justifique por que Prec se altera quando VAL é modificado.

2. CÁLCULO DE e^x

Escreva um programa em uma linguagem conhecida, para calcular e^x pela série de Taylor com n termos. O valor de x e o número de termos da série, n, são dados de entrada deste programa. Para valores negativos de x, o programa deve calcular e^x de duas formas: em uma delas o valor de x é usado diretamente na série de Taylor e, na outra forma, o valor usado na série é $y = -x$, e, em seguida, calcula-se o valor de e^x através de $1/e^x$.

a) Teste seu programa com vários valores de x (x próximo de zero e x distante de zero) e, para cada valor de x, teste o cálculo da série com vários valores de n. Analise os resultados obtidos.

b) Dificuldades com o cálculo do fatorial:

O cálculo de k! necessário na série de Taylor pode ser feito de modo a evitar a ocorrência de *overflow*. Os cálculos podem ser organizados de modo a se evitar o "estouro" no cálculo de k!; para isto é preciso analisar cuidadosamente o k-ésimo termo $x^k/k!$, tentar misturar o cálculo do numerador e do denominador e realizar divisões intermediárias. Estude uma maneira de realizar esta operação de modo a evitar que k! estoure.

c) Com a modificação do item (*b*), a série de Taylor pode ser calculada com quantos termos se queira. Qual seria um critério de parada para se interromper o cálculo da série?

MAKRON
Books

CAPÍTULO 2

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:

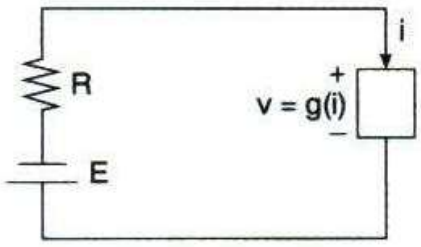


Figura 2.1

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 abcissas dos pontos onde uma curva intercepta o eixo $\overrightarrow{ox}$.

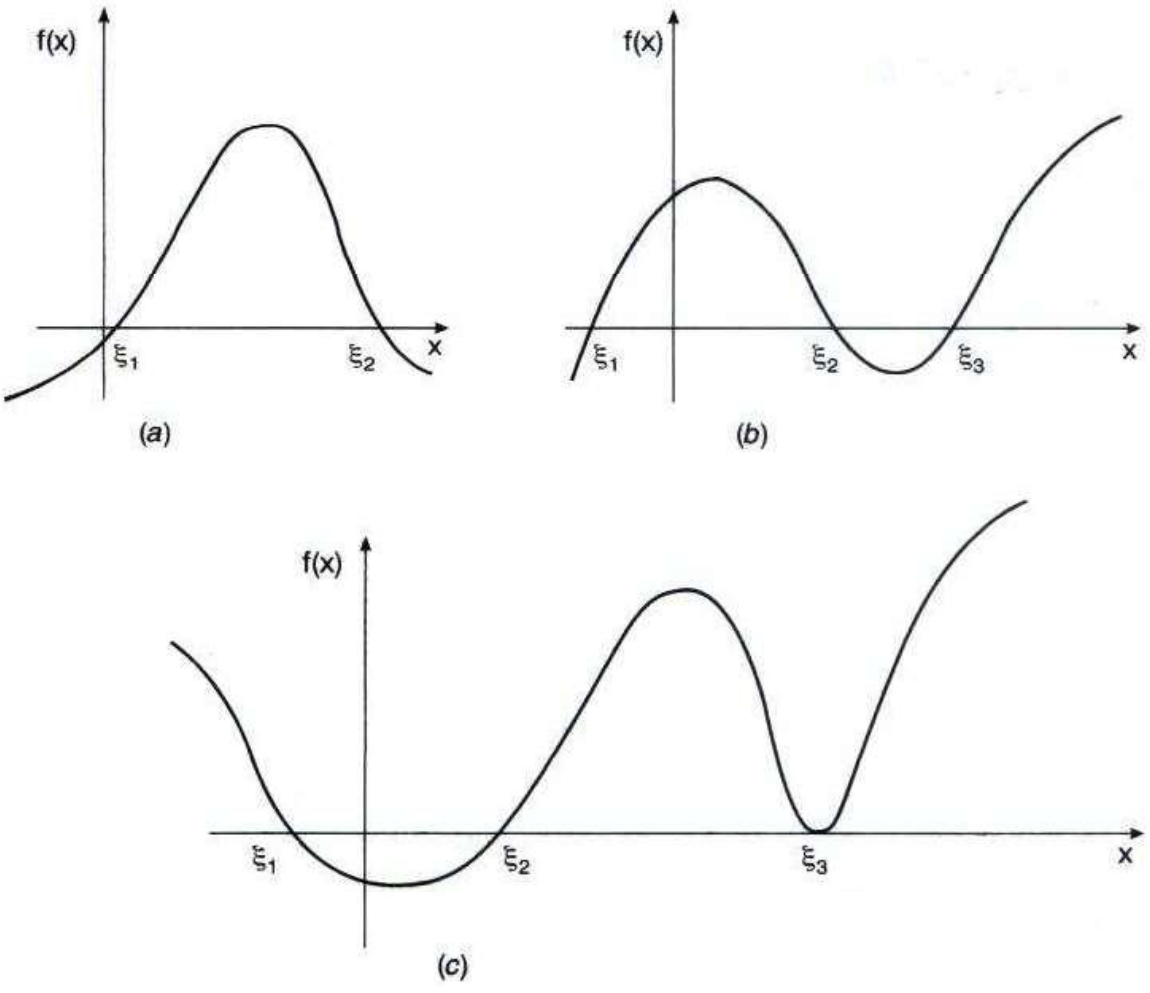


Figura 2.2

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

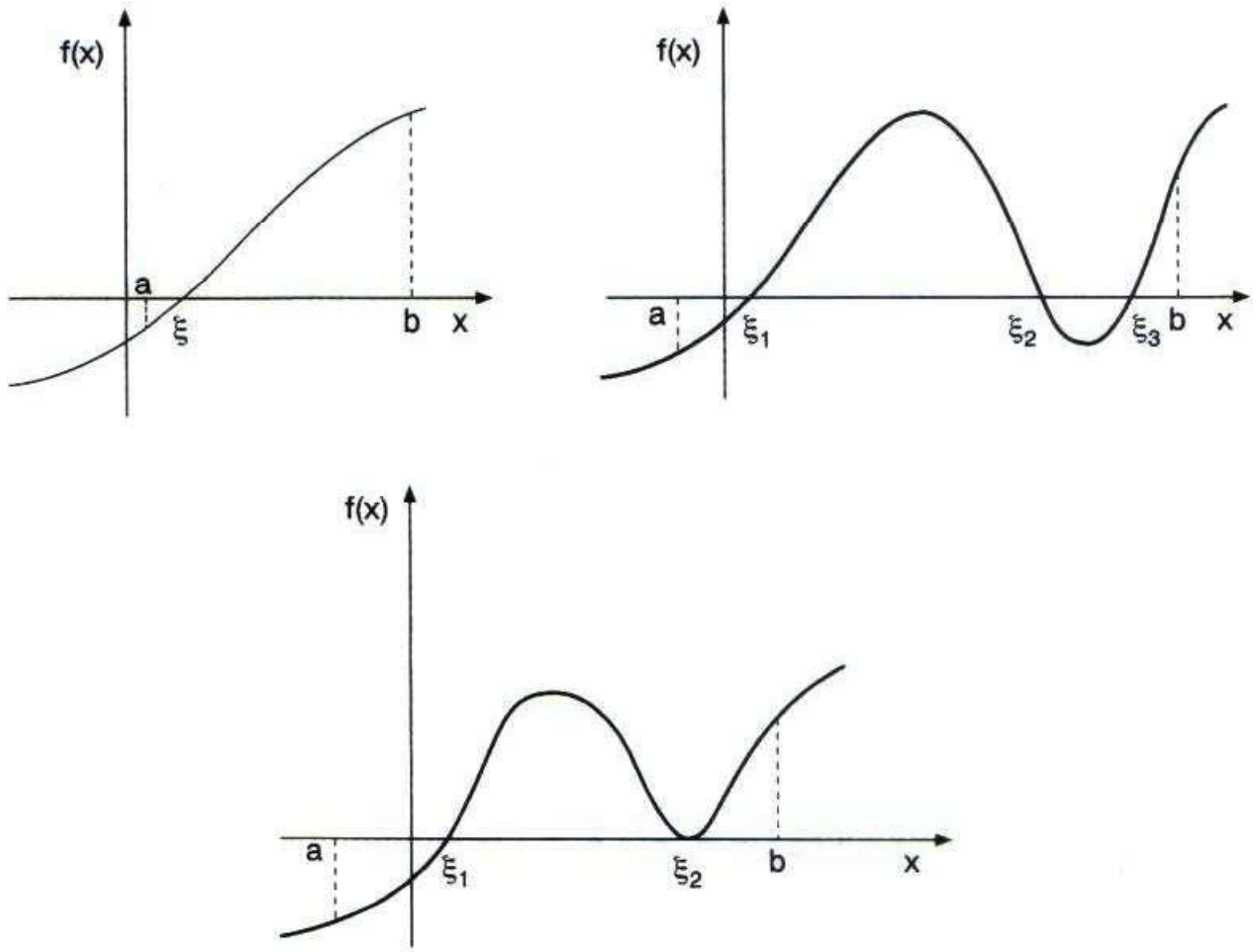


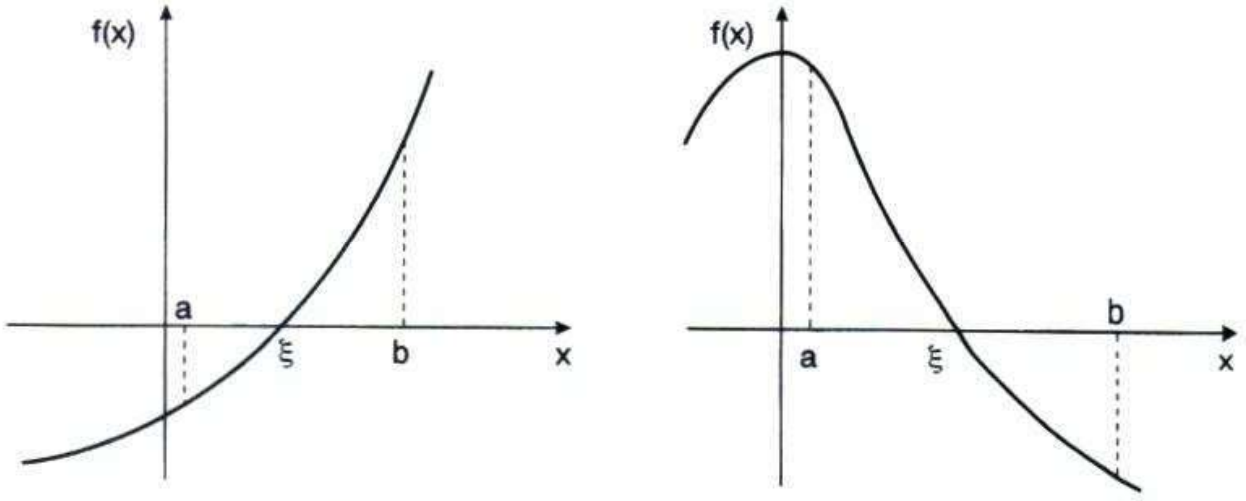
Figura 2.3

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



$f'(x) > 0, \forall x \in [a, b]$ $f'(x) < 0, \forall x \in [a, b]$

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) ≡ 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):

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

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:

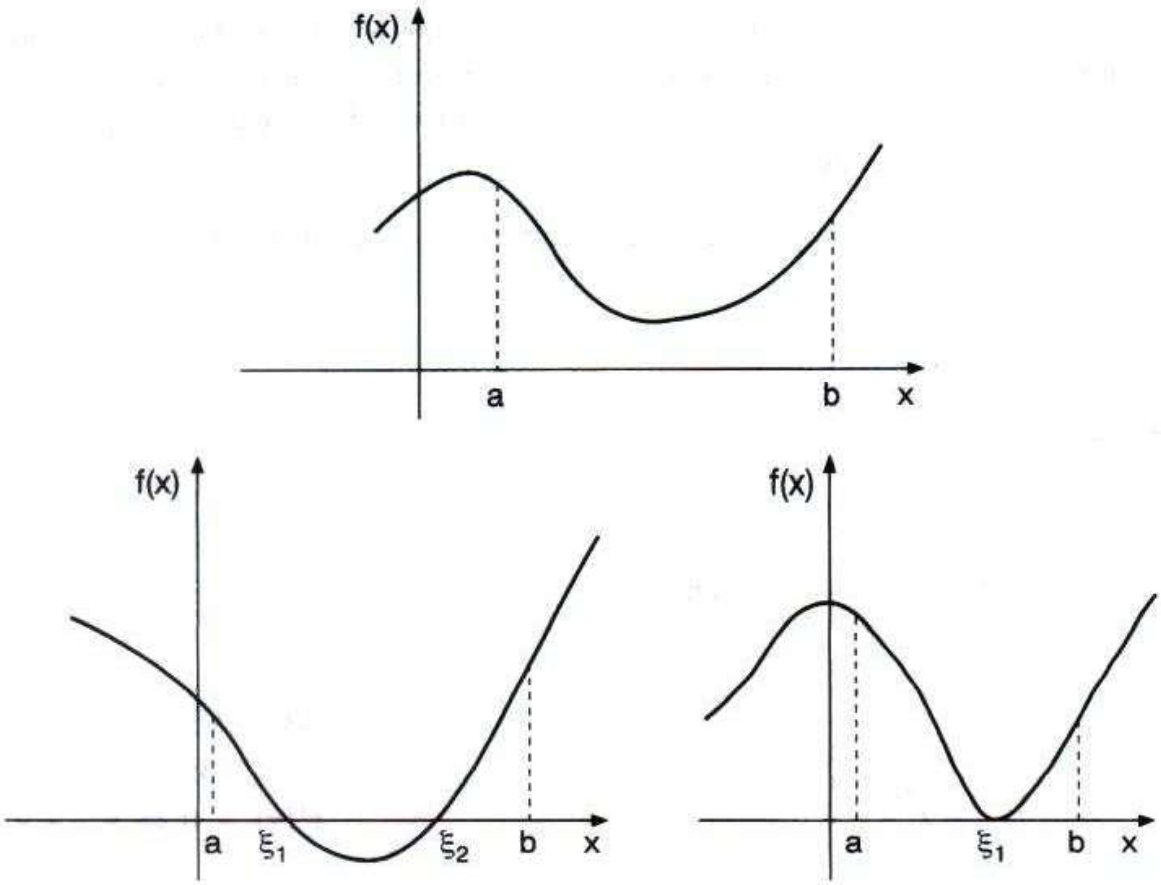


Figura 2.5

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 $\overrightarrow{ox}$;

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.

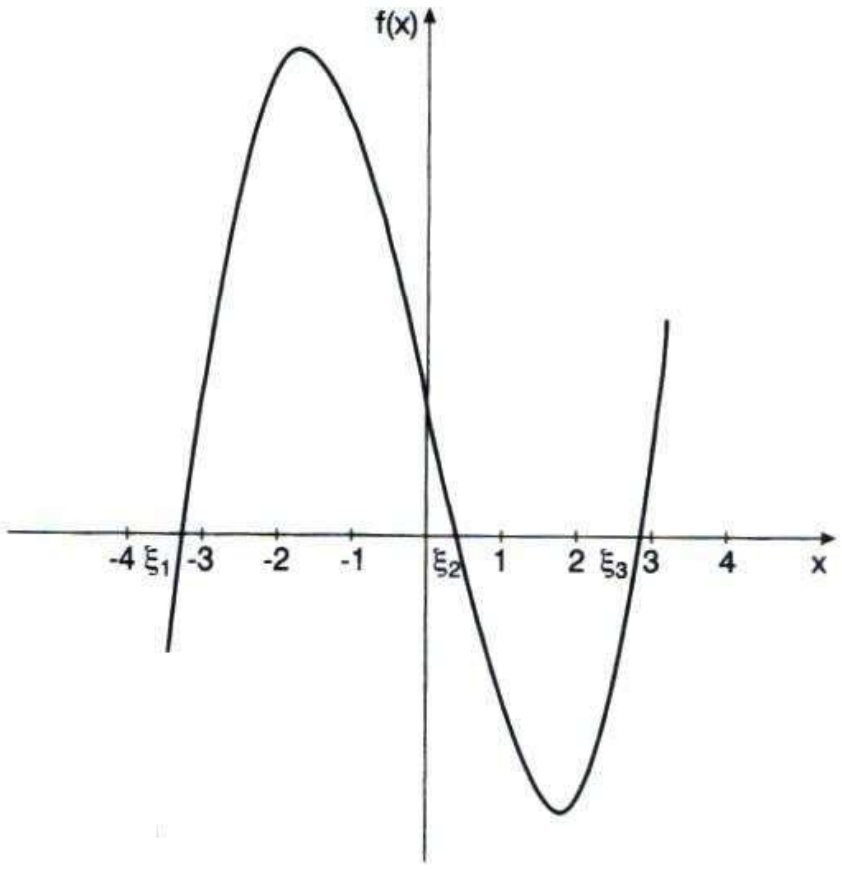
O esboço do gráfico de uma função requer um estudo detalhado do comportamento desta função, que envolve basicamente os itens: domínio da função; pontos de descontinuidade; intervalos de crescimento e decrescimento; pontos de máximo e mínimo; concavidade; pontos de inflexão e assíntotas da função.

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:



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

$f'(x) = 3x^2 - 9$

$f'(x) = 0 \Leftrightarrow x = \pm\sqrt{3}$

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,

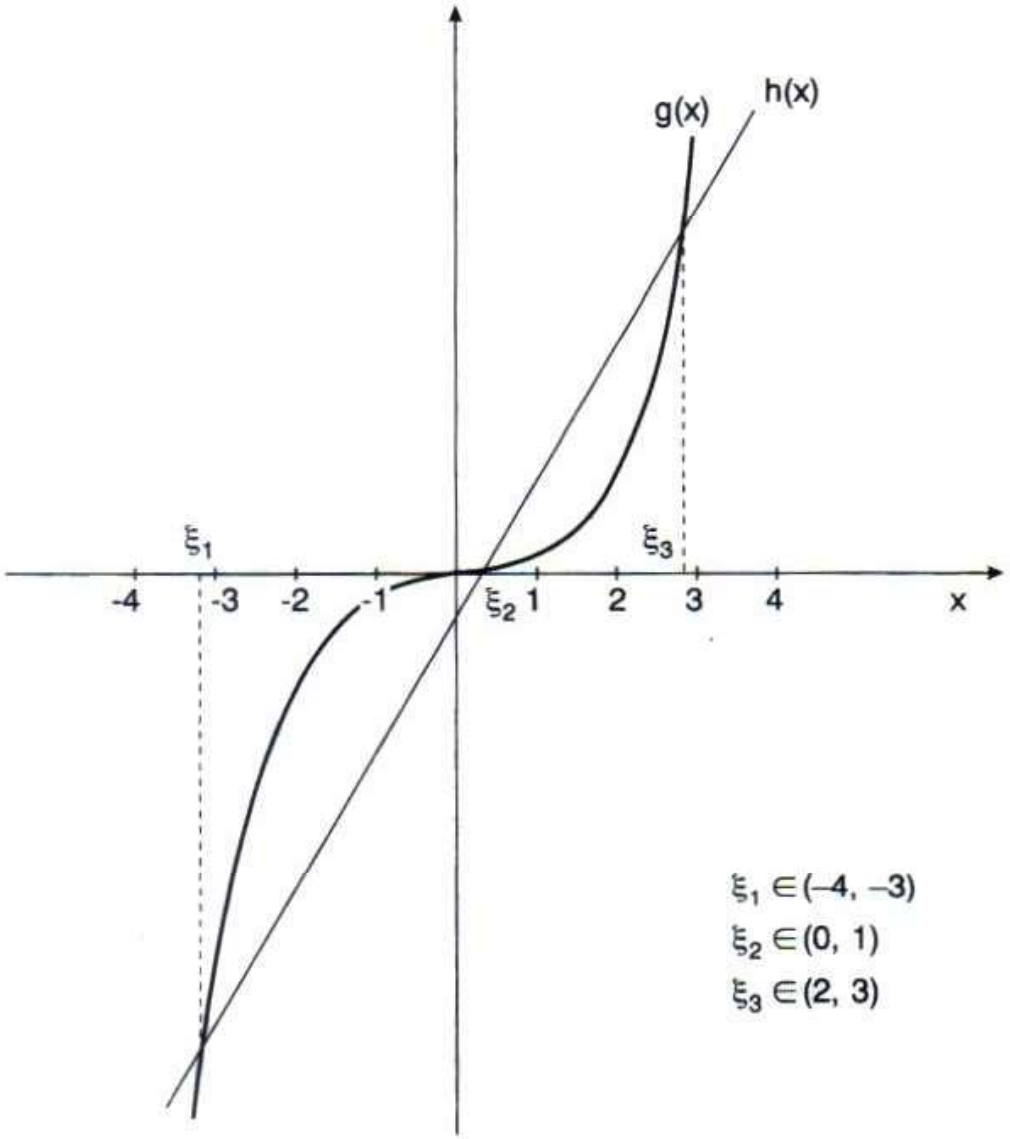


Figura 2.7

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

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

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

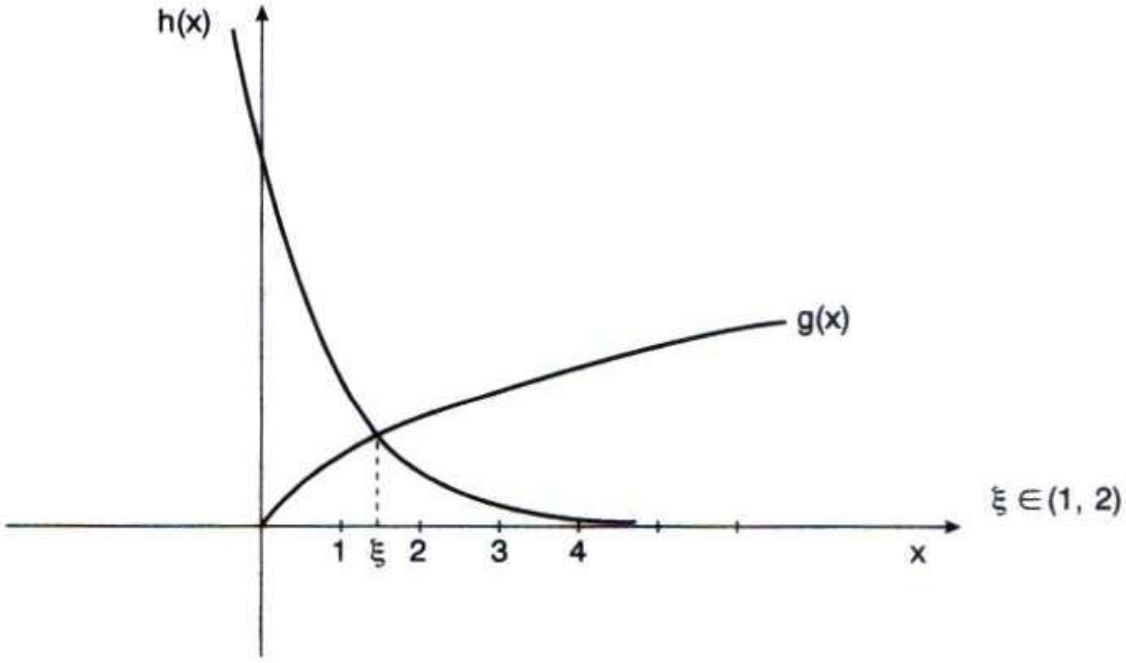


Figura 2.8

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

Usando novamente o processo (*ii*) temos que

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

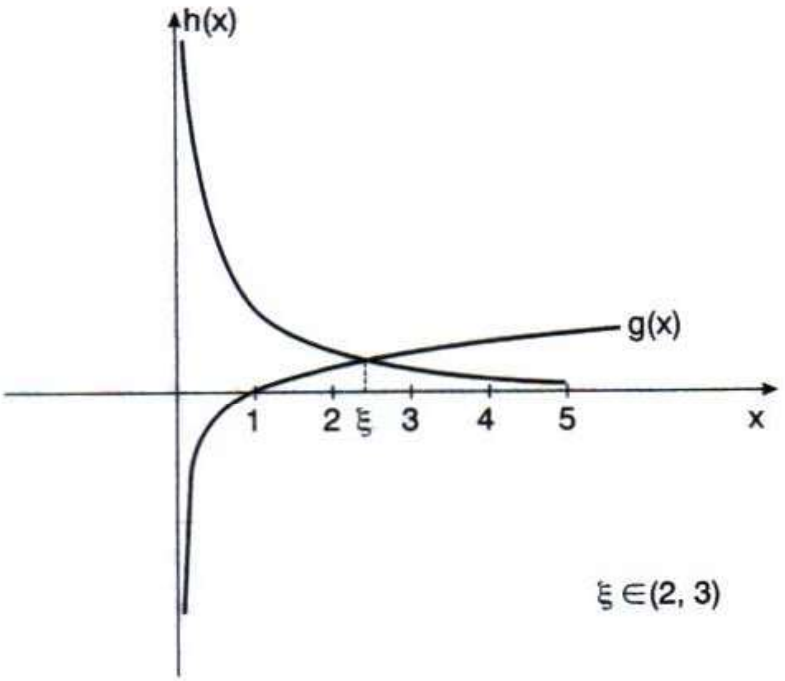


Figura 2.9

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:

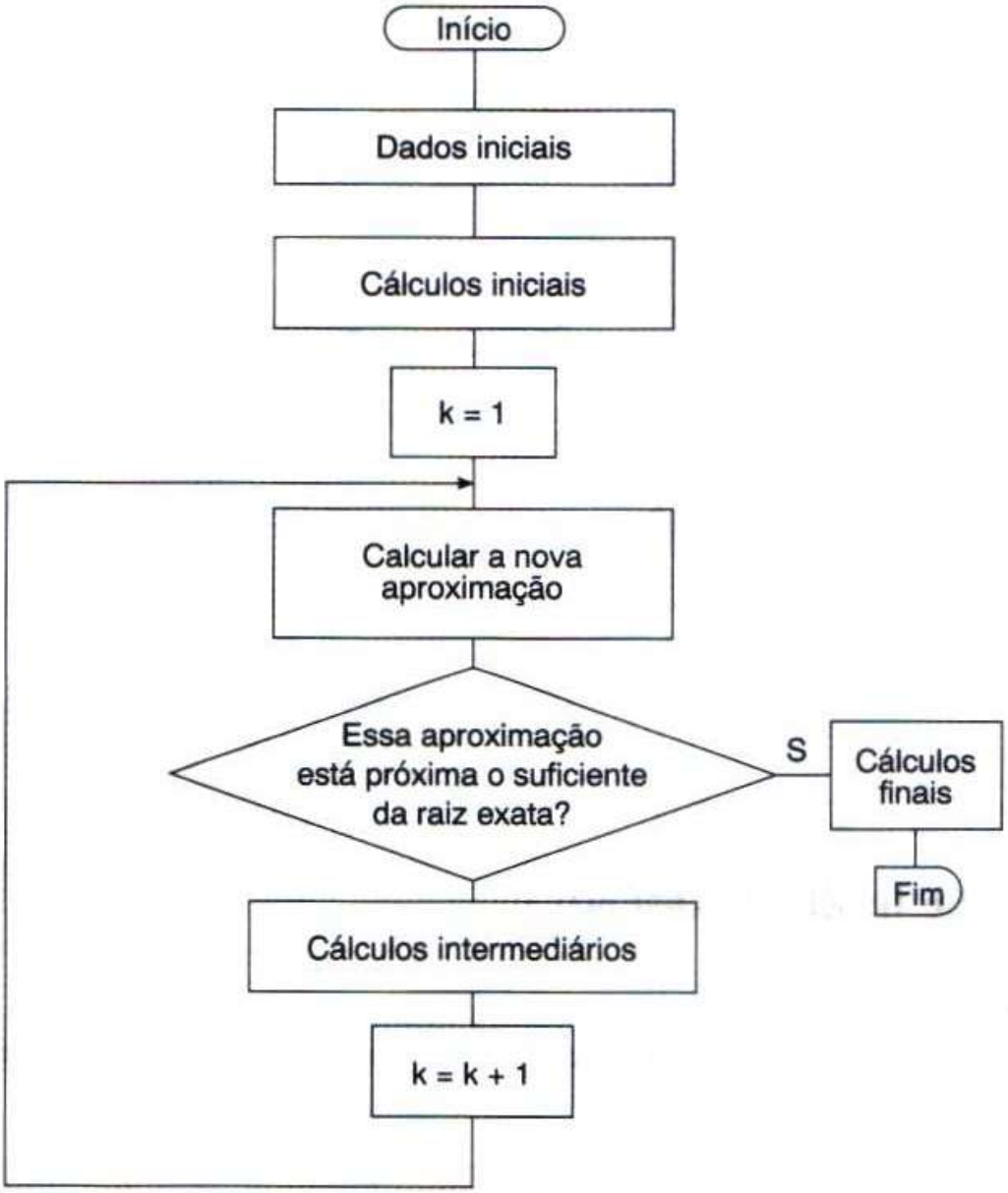


Figura 2.10

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

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

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:

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

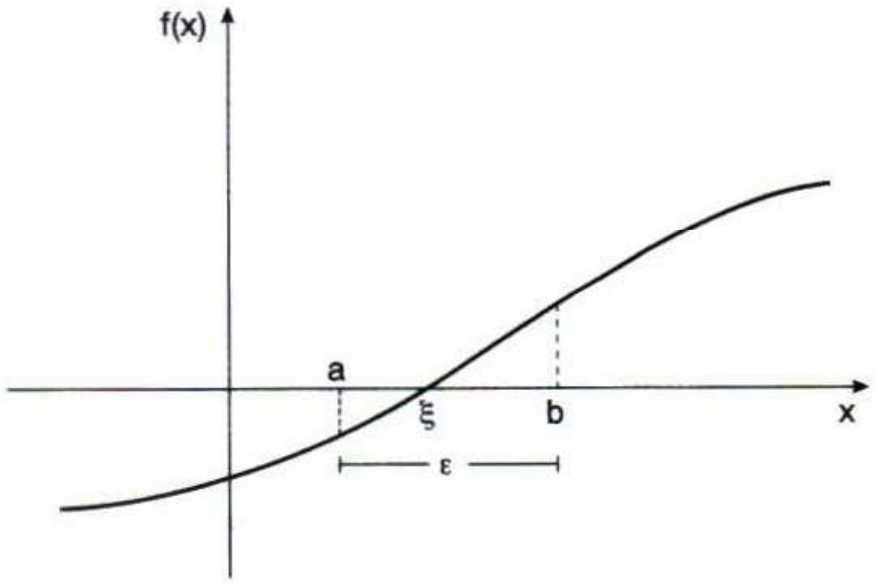


Figura 2.11

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

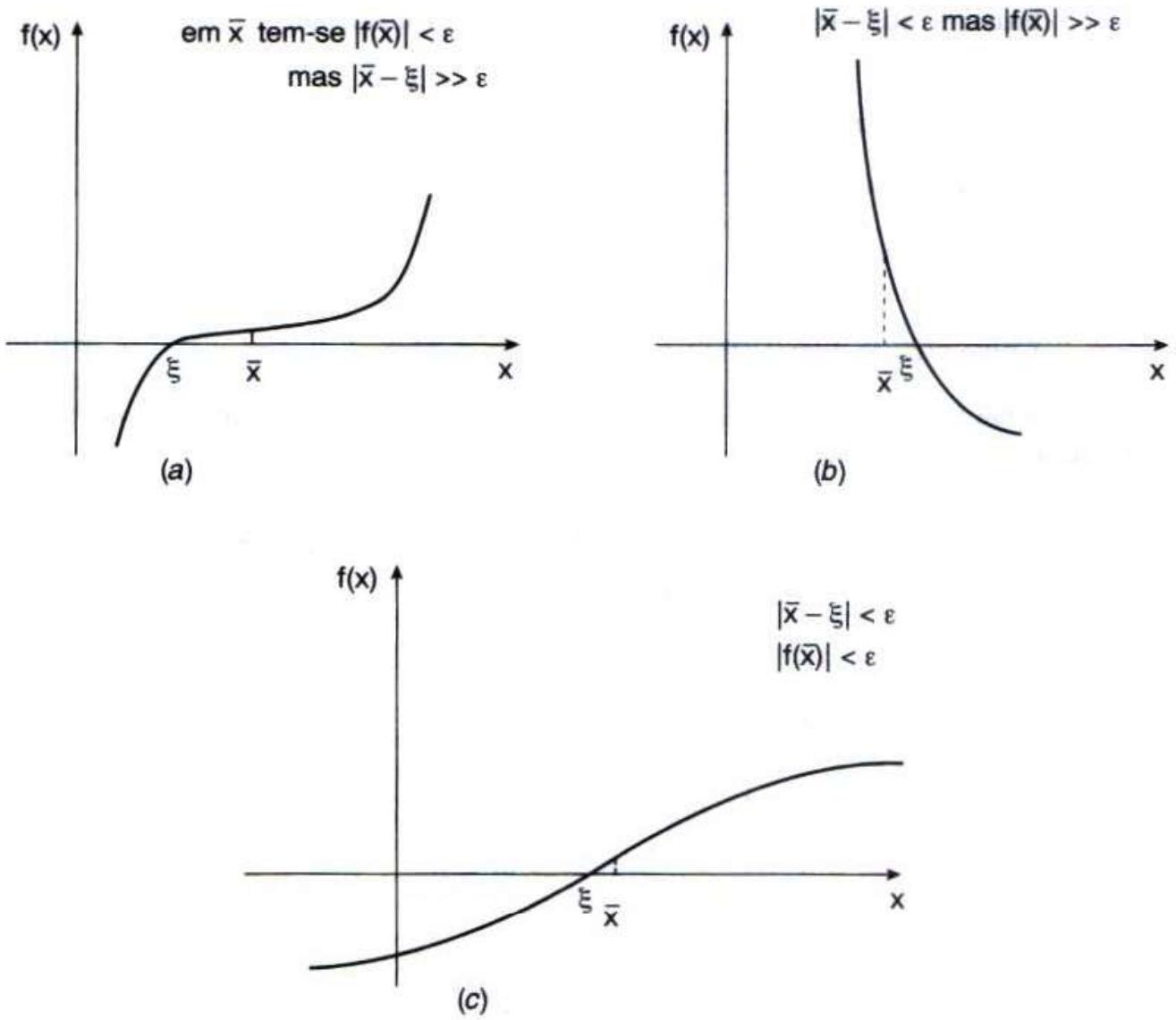


Figura 2.12

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 $\frac{|f(\bar{x})|}{L} < \varepsilon$ 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: $(b - a) < \varepsilon$, usando para isto a sucessiva divisão de [a, b] ao meio.

GRAFICAMENTE

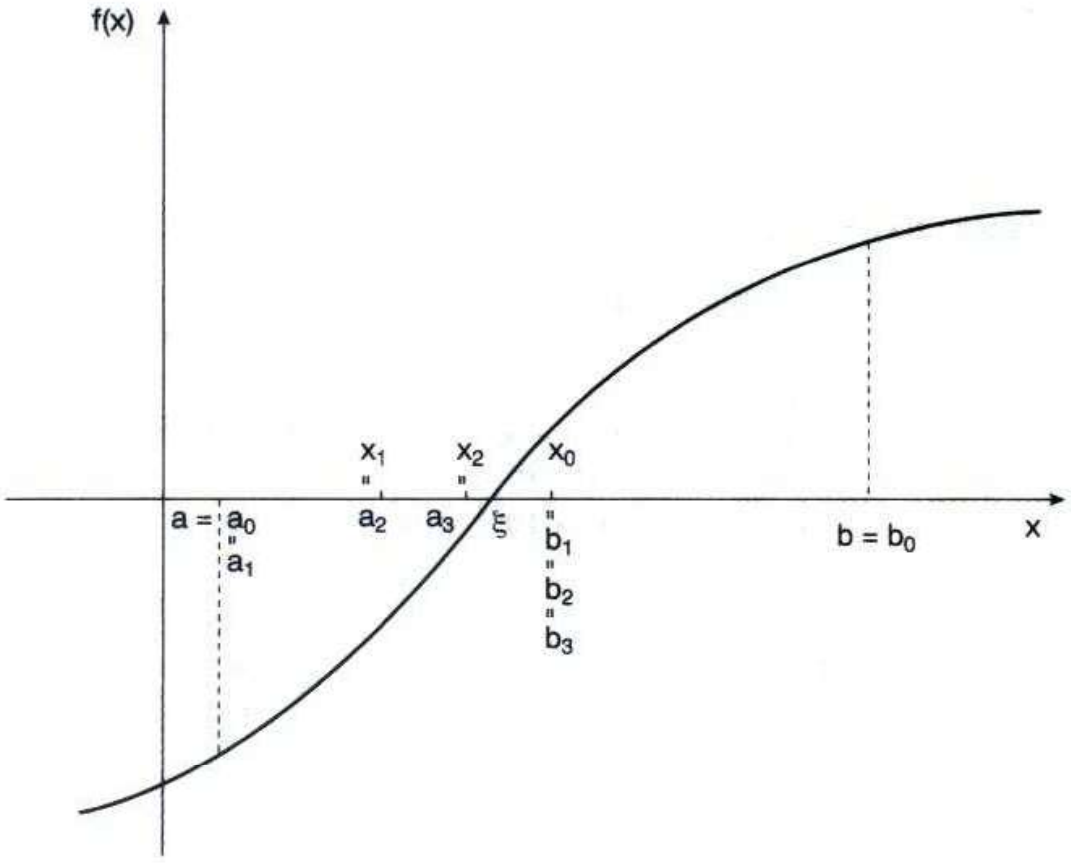


Figura 2.13

As iterações são realizadas da seguinte forma:

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

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

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

. .
. .
. .

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:

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

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

.
.
.

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) $x = \dfrac{a + b}{2}$

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

$f(x) = x^3 - 9x + 3$ $\quad I = [0, 1]$ $\quad \varepsilon = 10^{-3}$

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 $\bar{x} = \dfrac{a + b}{2}$.

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

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

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

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

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

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

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

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

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

$\lim_{k \to \infty} b_k - \lim_{k \to \infty} a_k = 0 \Rightarrow \lim_{k \to \infty} b_k = \lim_{k \to \infty} a_k$. 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

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

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

$$0 \geqslant \lim_{k \to \infty} f(a_k) f(b_k) = \lim_{k \to \infty} f(a_k) \lim_{k \to \infty} f(b_k) = f(\lim_{k \to \infty} a_k)\, f(\lim_{k \to \infty} b_k) =$$

$$= f(r)\, f(s) = f(\ell)\, f(\ell) = [f(\ell)]^2$$

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

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

Concluímos, 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 $b - a < \varepsilon$, 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 $b_k - a_k < \varepsilon$, ou seja,

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

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

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

Por exemplo, se desejarmos encontrar ξ, o zero da função $f(x) = x \log(x) - 1$ que está no intervalo [2, 3] com precisão $\varepsilon = 10^{-2}$, 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 $b_0 - a_0 >> \varepsilon$ e se ε for muito pequeno, o número de iterações tende a ser muito grande, como por exemplo:

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

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 $\overrightarrow{ox}$ e a reta r(x) que passa por (a, f(a)) e (b, f(b)):

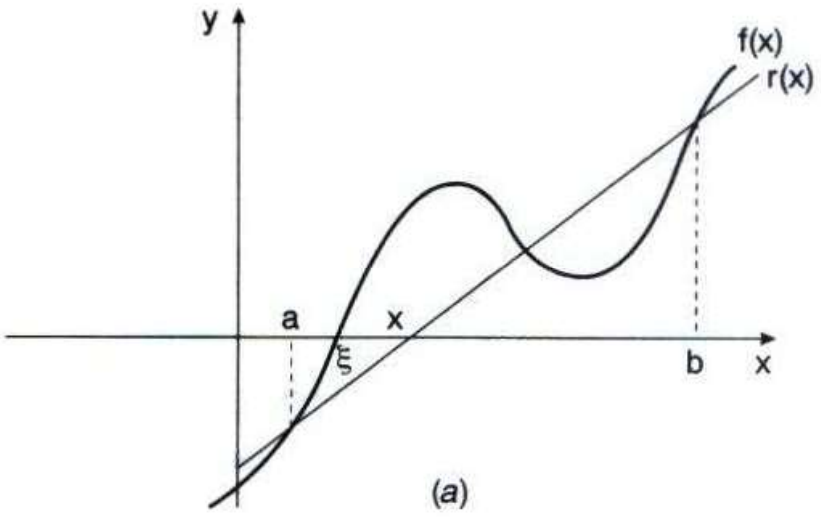


Figura 2.14

E as iterações são feitas assim:

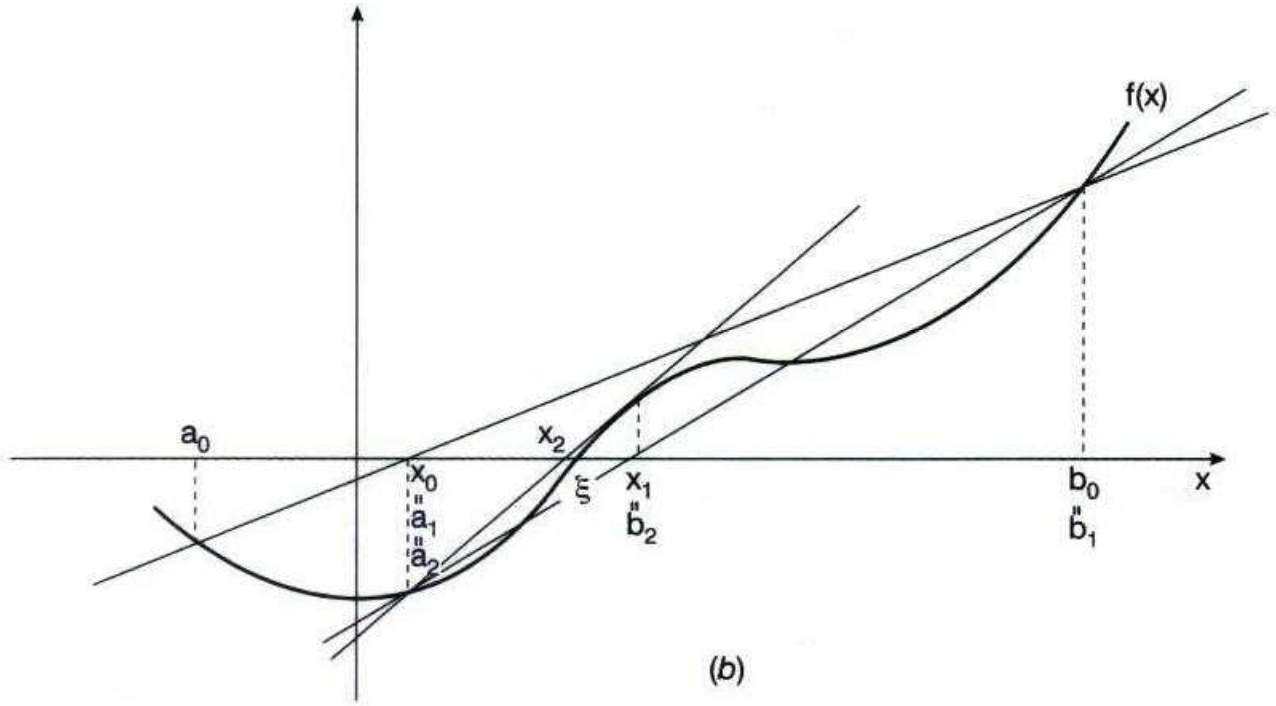


Figura 2.14

Exemplo 5

O método da posição falsa aplicado a xlog(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,

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

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

$f(x_1) = -0.0011$. 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.

   $$\left.\begin{array}{l} \text{se } |f(a)| < \varepsilon_2 \\ \text{ou se } |f(b)| < \varepsilon_2 \end{array}\right] \text{escolha a ou b como } \bar{x}. \text{ FIM.}$$

3) $k = 1$

4) $M = f(a)$

5) $x = \dfrac{af(b) - bf(a)}{f(b) - f(a)}$

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

$f(x) = x^3 - 9x + 3 \qquad I = [0, 1] \qquad \varepsilon_1 = \varepsilon_2 = 5 \times 10^{-4}$

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:

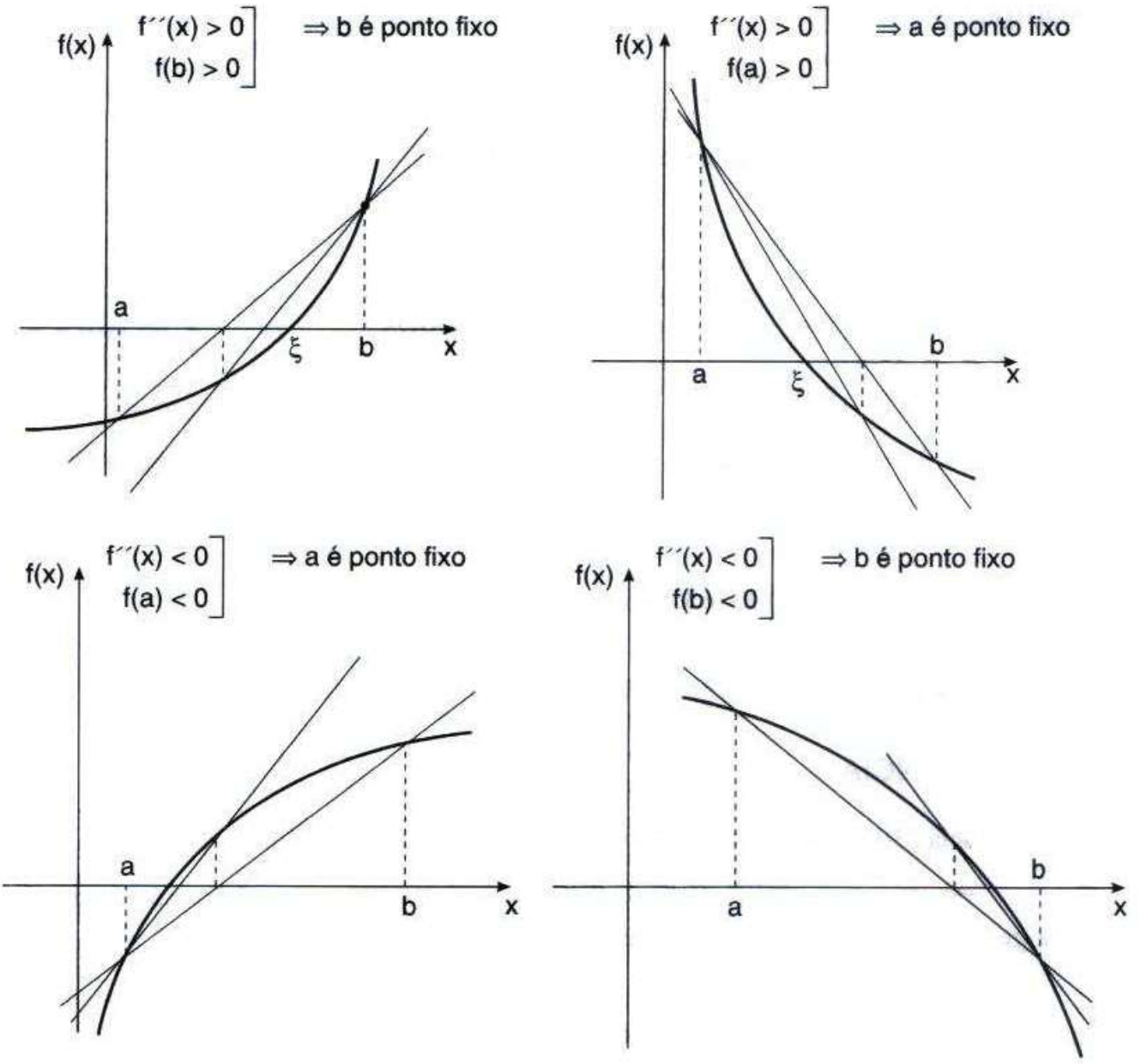


Figura 2.15

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 $\lim_{k \to \infty} x_k = \xi$.

Analisando ainda estes gráficos, podemos concluir que em geral o método da posição falsa obtém como raiz aproximada um ponto $\overline{x}$, no qual $| f(\overline{x}) | < \varepsilon$, 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) $\varphi_3(x) = \dfrac{6}{x} - 1$;

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

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

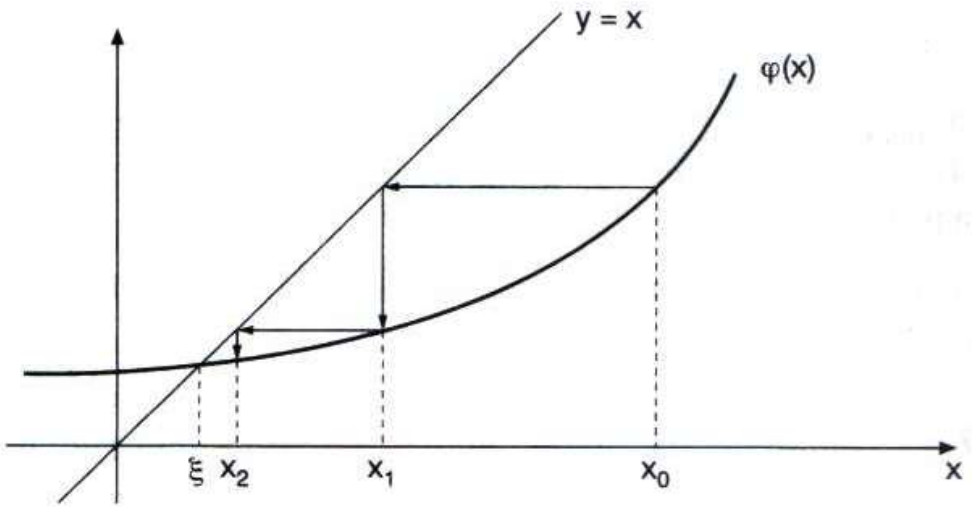
($\Rightarrow$) seja ξ tal que $f(\xi) = 0$.

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

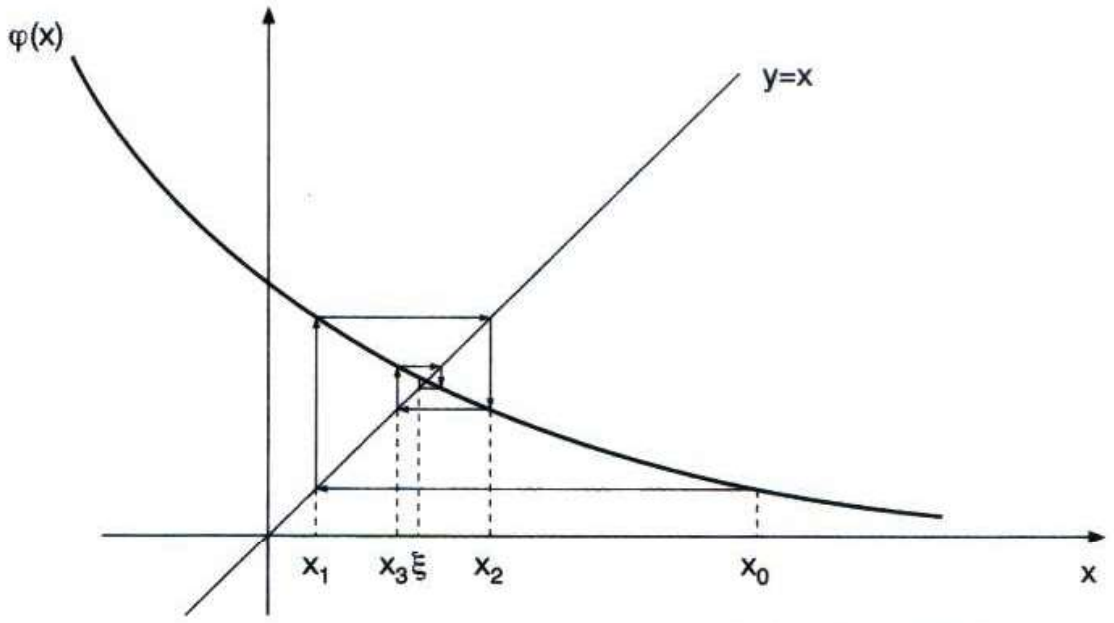
Com isto vemos que, dada uma equação f(x) = 0, existem infinitas funções de iteração φ(x) para a equação f(x) = 0.

Graficamente, uma raiz da equação x = φ(x) é a abcissa do ponto de intersecção da reta y = x e da curva y = φ(x):



$\{x_k\} \to \xi$ quando $k \to \infty$

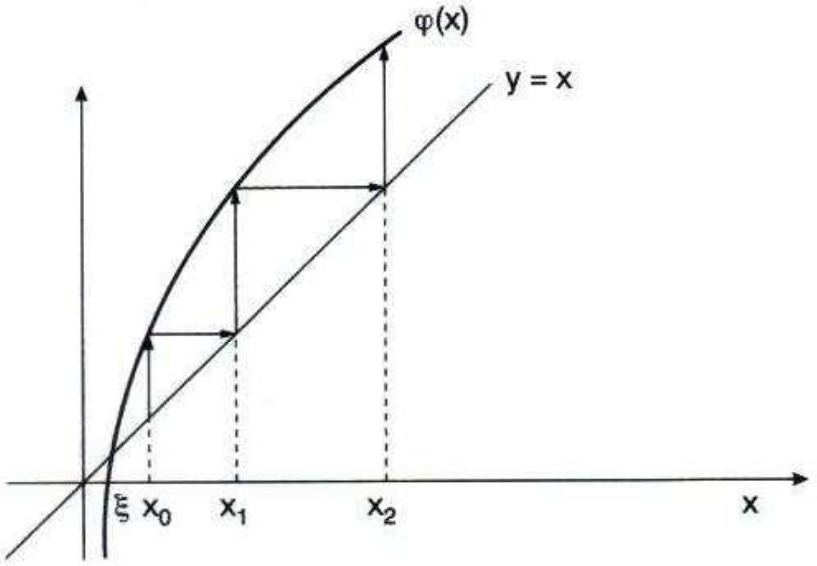
(*a*)



$\{x_k\} \to \xi$ quando $k \to \infty$

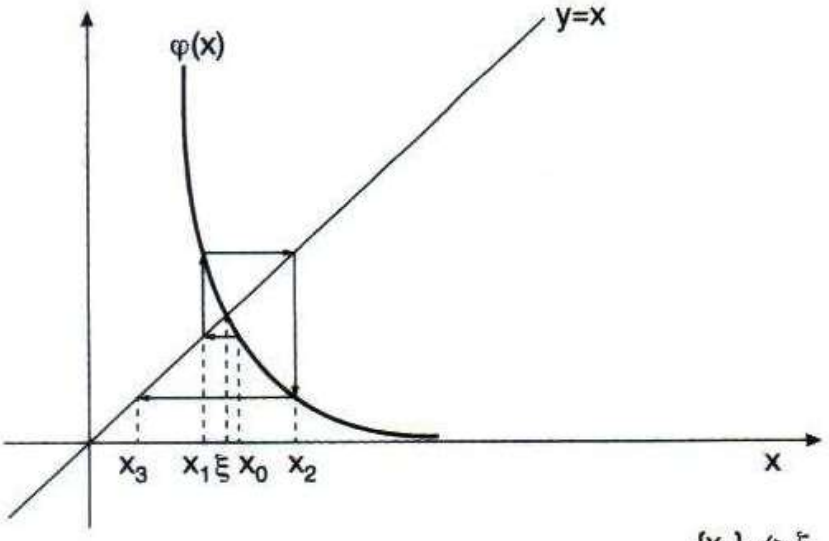
(*b*)

Figura 2.16



$\{x_k\} \nrightarrow \xi$

(c)



$\{x_k\} \nrightarrow \xi$

(d)

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

$$x_1 = \varphi(x_0) = 6 - 1.5^2 = 3.75$$
$$x_2 = \varphi(x_1) = 6 - (3.75)^2 = -8.0625$$
$$x_3 = \varphi(x_2) = 6 - (-8.0625)^2 = -59.003906$$
$$x_4 = \varphi(x_3) = -(-59.003906)^2 + 6 = -3475.4609$$
$$\vdots$$

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

GRAFICAMENTE

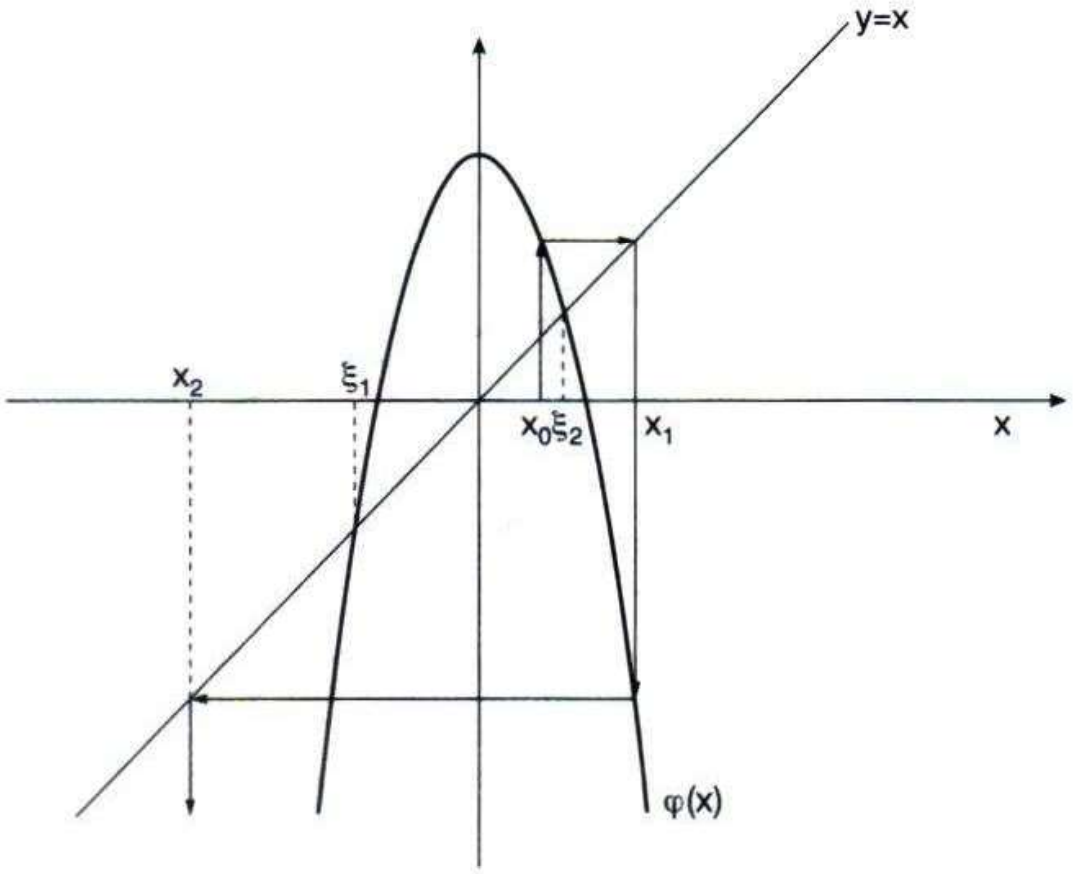


Figura 2.17

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

$$x_1 = \varphi(x_0) = \sqrt{6 - 1.5} = 2.12132$$
$$x_2 = \varphi(x_1) = 1.96944$$
$$x_3 = \varphi(x_2) = 2.00763$$
$$x_4 = \varphi(x_3) = 1.99809$$
$$x_5 = \varphi(x_4) = 2.00048$$
$$\vdots$$

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

GRAFICAMENTE

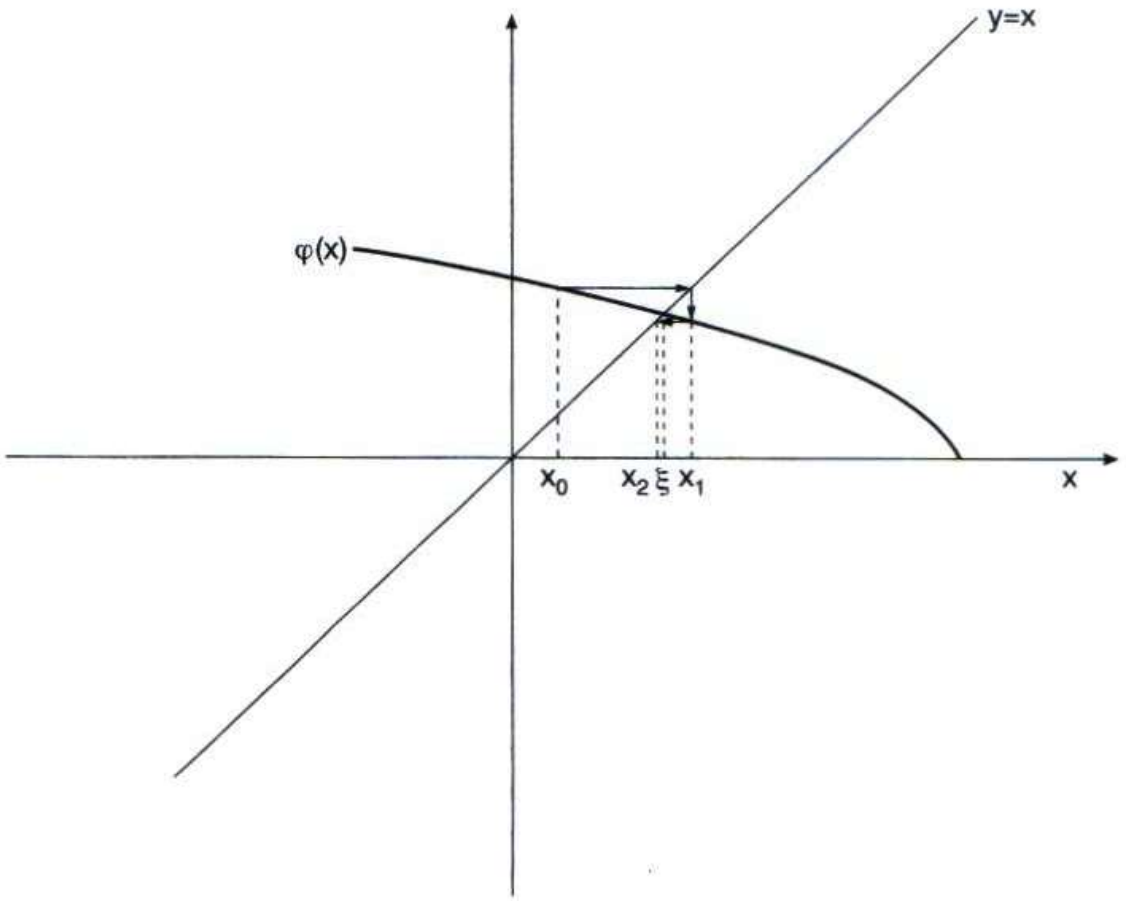


Figura 2.18

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 $\lim_{k \to \infty} x_k = \xi$.

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

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

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

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

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

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

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

De (1), segue que:

$$|x_1 - \xi| = |\varphi(x_0) - \varphi(\xi)| = \underbrace{|\varphi'(c_0)|}_{\leq M} |x_0 - \xi| \leq M |x_0 - \xi|$$

(c_0 está entre x_0 e ξ)

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

(c_1 está entre x_1 e ξ)

.
.
.

$$|x_k - \xi| = |\varphi(x_{k-1}) - \varphi(\xi)| = \underbrace{|\varphi'(c_{k-1})|}_{\leq M} |x_{k-1} - \xi| \leq M |x_{k-1} - \xi| \leq \ldots \leq M^k |x_0 - \xi|$$

(c_k está entre x_k e ξ)

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

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

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 $\mathbb{R}$.

$|\varphi_1'(x)| < 1 \Leftrightarrow |2x| < 1 \Leftrightarrow -\frac{1}{2} < x < \frac{1}{2}$. Então, não existe um intervalo I 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 $\varphi_2'(x) = \frac{-1}{2\sqrt{6 - x}}$

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

$\varphi_2'(x)$ é contínua em $S' = \{x \in \mathbb{R} \mid x < 6\}$ (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 $x_0 = -2.5$:

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

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

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

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

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

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

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:

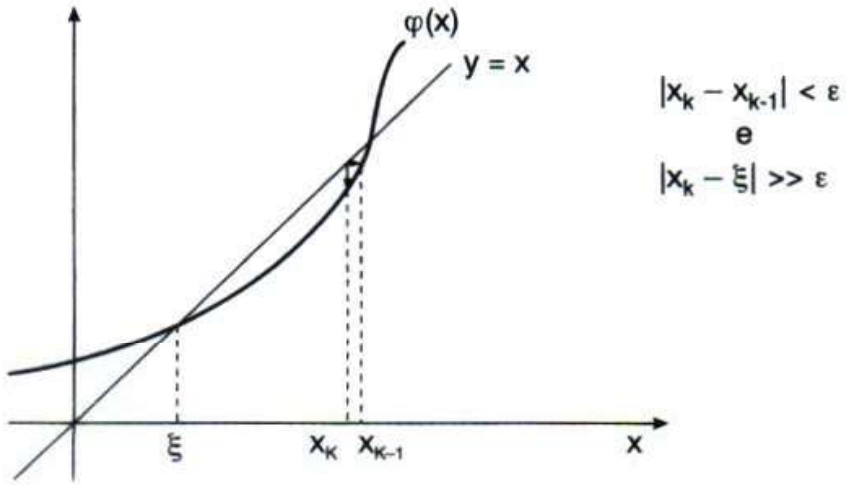


Figura 2.19

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 $|x_k - \xi| < |x_k - x_{k-1}|$.

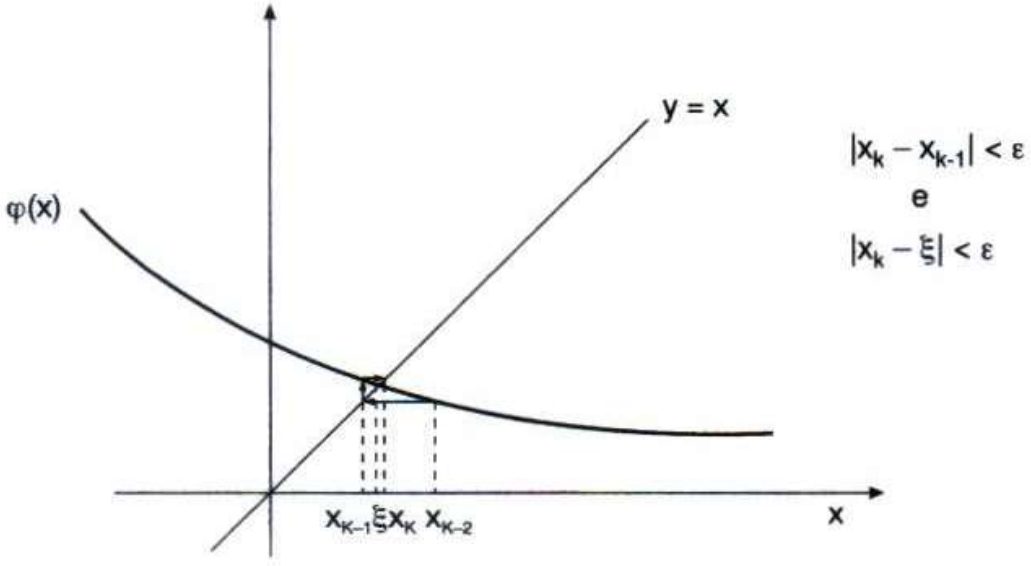


Figura 2.20

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

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

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

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

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

Se $\lim_{k \to \infty} \frac{e_{k+1}}{e_k} = C, \quad 0 \leq |C| < 1$, 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:

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

Considerando que a seqüência $\{x_k\}$ é convergente, temos que $e_k \to 0$ quando $k \to \infty$, 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 $k \to \infty$

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

Portanto, $\lim\limits_{k \to \infty} \dfrac{e_{k+1}}{e_k} = \varphi'(\xi) = C$ 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 $|\varphi'(x)| \leqslant M < 1$, $\forall\, x \in I$, 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, $\varphi'(\xi) = 0 \Leftrightarrow 1 + A(\xi)f'(\xi) = 0 \Rightarrow A(\xi) = \dfrac{-1}{f'(\xi)}$, donde tomamos $A(x) = \dfrac{-1}{f'(x)}$.

Então, dada f(x), a função de iteração $\varphi(x) = x - \dfrac{f(x)}{f'(x)}$ 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 $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$, $k = 0, 1, 2, \ldots$.

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

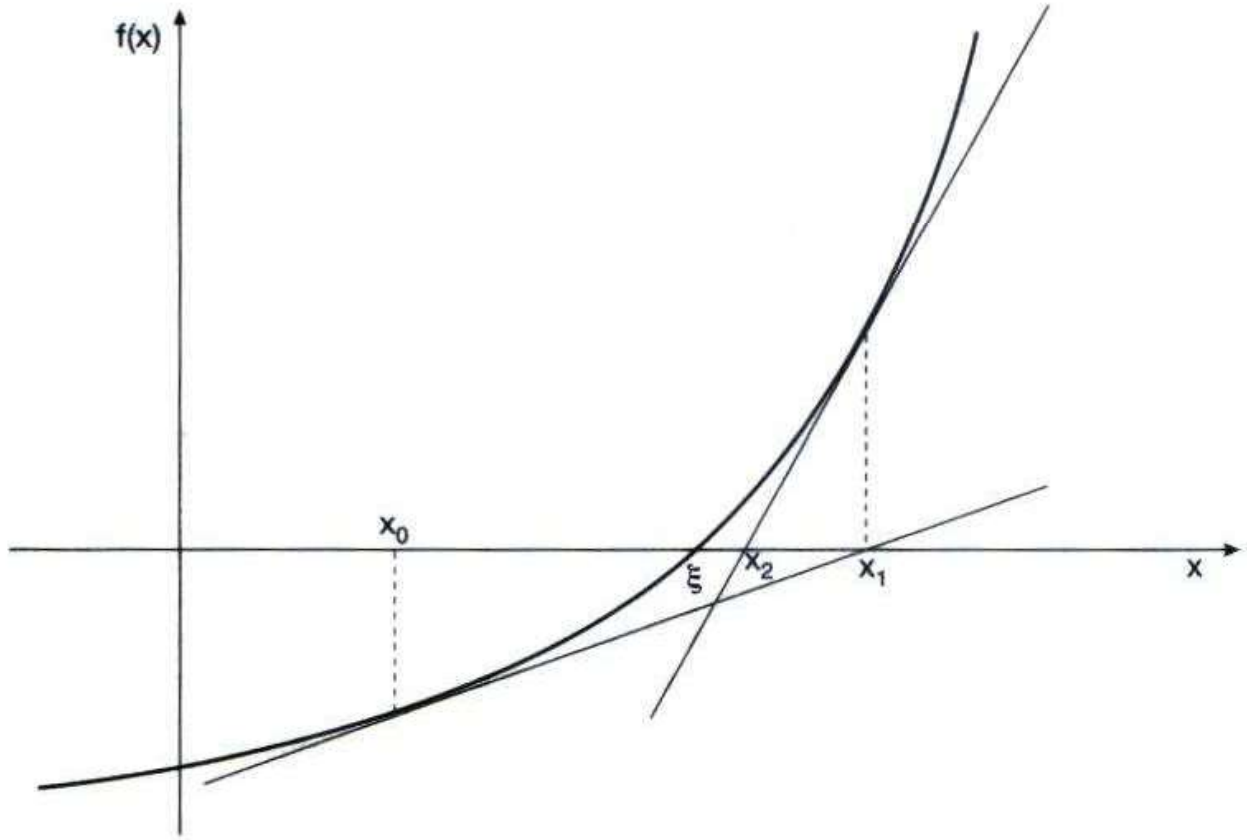


Figura 2.21

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 $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$ convergirá para a raiz.

DEMONSTRAÇÃO

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

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, $\varphi'(x) = \dfrac{f(x)\,f''(x)}{[f'(x)]^2}$. 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 $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$ 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) $x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$

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

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

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.

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

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

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

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

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

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

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

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

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

$$= \frac{1}{2}\,\frac{f''[\lim\limits_{k \to \infty} (c_k)]}{f'[\lim\limits_{k \to \infty} (x_k)]} = \frac{1}{2}\,\frac{f''(\xi)}{f'(\xi)} = \frac{1}{2}\,\varphi''(\xi) = C$$

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

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

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

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

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 abcissa do ponto de intersecção do eixo $\overrightarrow{ox}$ e da reta secante que passa por $(x_{k-1}, f(x_{k-1}))$ e $(x_k, f(x_k))$:

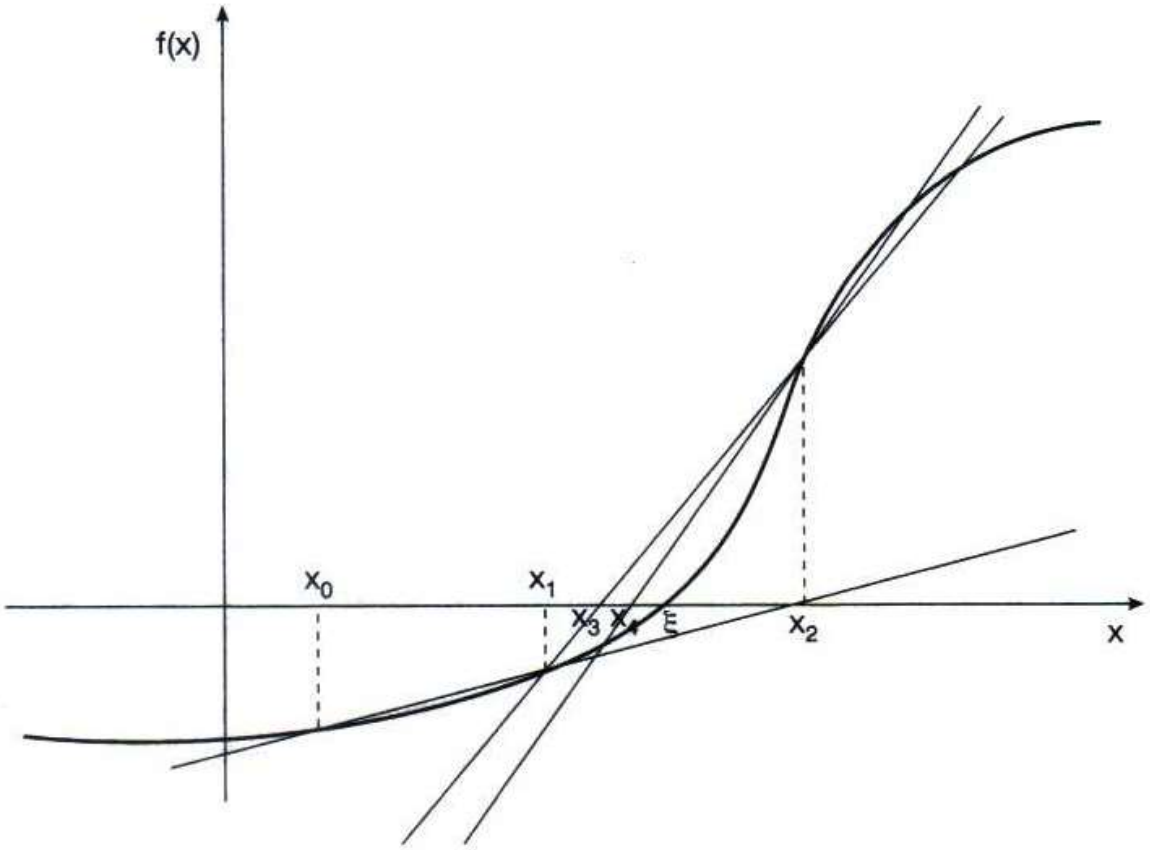


Figura 2.22

Exemplo 16

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

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

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

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

$$= 1.99999$$

. . .
. .
. .

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 $|x_1 - x_0| < \varepsilon_2$ 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; acrescente-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

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

	Bissecção	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

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

	Bissecção	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

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

	Bissecção	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

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

	Bissecção	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".

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 $x_2 = 4/3 = 1.33333...$ Observe que nos testes 2 e 3 tomamos propositadamente 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:

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

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 $\xi_1 = 1$.

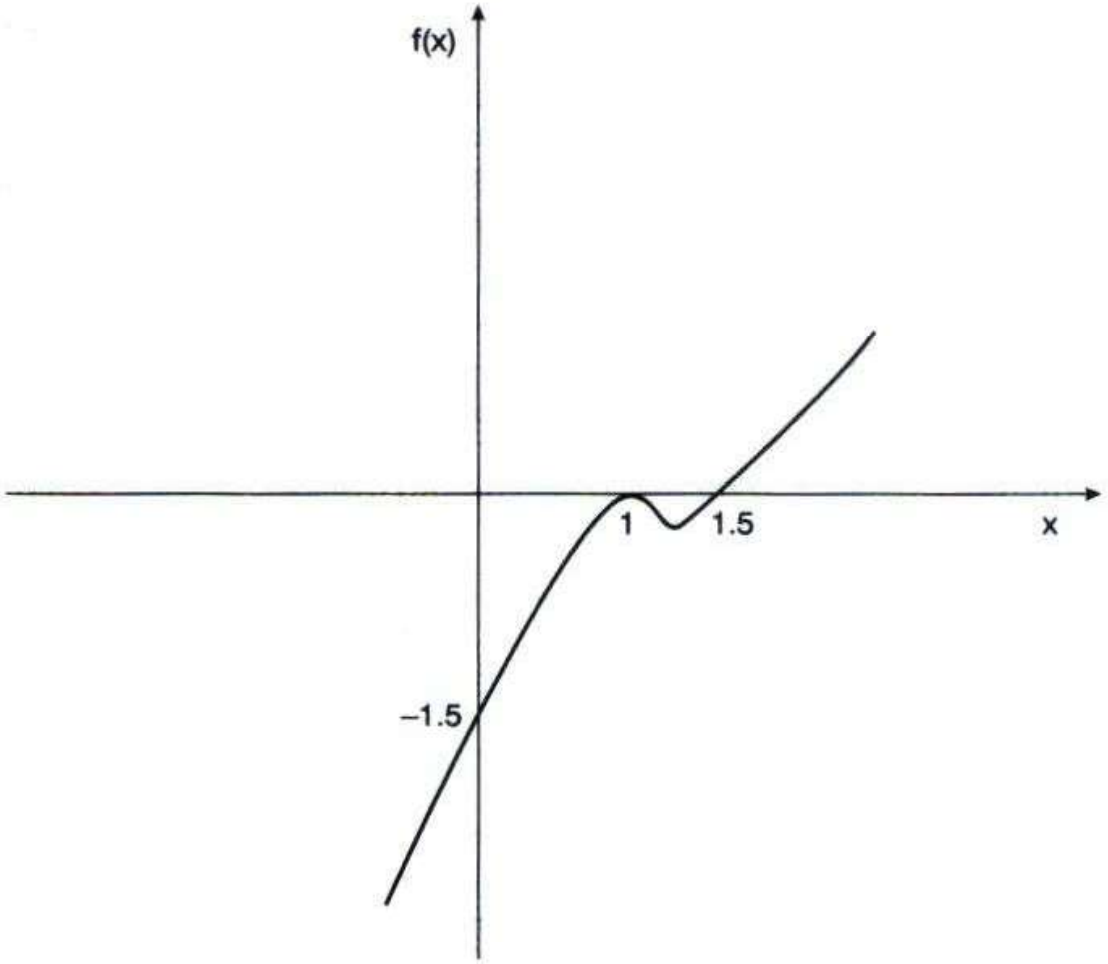


Figura 2.23

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 freqüência em aplicações faz com que lhes dediquemos especial atenção.

Normalmente, um polinômio de grau n é escrito na forma

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

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 $n \geqslant 5$, 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)

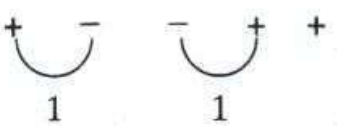
"Se $p_n(x)$ é um polinômio de grau $n \geqslant 1$, ou seja, $p_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$, $a_0, a_1, \ldots, a_n$ reais ou complexos, com $a_n \neq 0$, então $p_n(x)$ tem pelo menos um zero, ou seja, existe um número complexo ξ tal que $p_n(\xi) = 0$."

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".

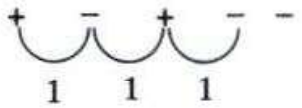
Exemplo 23

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

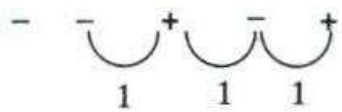
$\Rightarrow v = 0$ e p: $(v - p \geqslant 0) \Rightarrow p = 0$.

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

Exemplo 24

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

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



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

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

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

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

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

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

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 $g_0(\alpha)$, $g_1(\alpha)$, ... $g_n(\alpha)$, 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

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

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

$g_2(x) = ?$

$$\begin{array}{l|l} x^3 + x^2 - x + 1 & 3x^2 + 2x - 1 \\ \hline -x^3 - \frac{2}{3}x^2 + \frac{1}{3}x & \frac{1}{3}x + \frac{1}{9} \\ \hline \frac{1}{3}x^2 - \frac{2}{3}x + 1 & \\ -\frac{1}{3}x^2 - \frac{2}{9}x + \frac{1}{9} & \\ \hline -\frac{8}{9}x + \frac{10}{9} & \Rightarrow g_2(x) = \frac{8}{9}x - \frac{10}{9}. \end{array}$$

$g_3(x) = ?$

$3x^2 + 2x - 1$	$\frac{8}{9}x - \frac{10}{9}$
.	$\frac{27}{8}x + \frac{207}{32}$
.	
.	
$\frac{99}{16}$	$\Rightarrow g_3(x) = -\frac{99}{16}$.

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 $\alpha \leqslant x \leqslant \beta$ é exatamente $\tilde{v}(\alpha) - \tilde{v}(\beta)$.

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

$g_0(\beta) = 34 > 0$

$g_1(\beta) = 32 > 0$

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

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

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,..., 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

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

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,

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

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

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

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

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

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

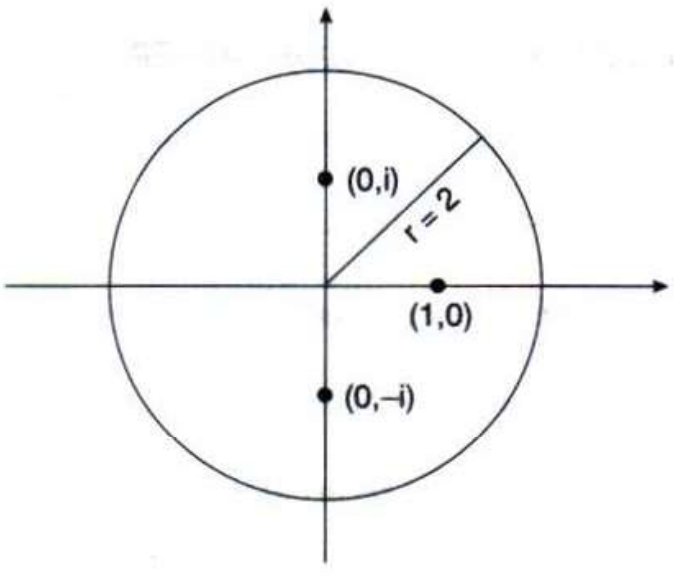


Figura 2.24

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 freqüentemente 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,..., n, ou seja:

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

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:

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

Este polinômio pode ser escrito na forma:

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

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 $n + (n-1) + \ldots + 2 + 1 = \frac{(1+n)n}{2}$ desde que x^j seja calculado por $x \cdot x \cdot x \ldots \cdot x$, j vezes, pois a potenciação calculada desta forma introduz erros menores de arredondamento.

Agora,

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

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

Temos então, no caso de n = 4, que

$$p_4(x) = (((\underbrace{a_4 x + a_3}_{b_4})x + a_2)x + a_1)x + a_0$$

$$\underbrace{\qquad\qquad}_{b_3}$$

$$\underbrace{\qquad\qquad\qquad}_{b_2}$$

$$\vdots$$

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

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

$\Rightarrow p(c) = b_0$.

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

$$b_n = a_n$$

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

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

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

$$\Rightarrow \quad p_4'(x) = 4a_4x^3 + 3a_3x^2 + 2a_2x + a_1.$$

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:

$$\begin{aligned}
p_4'(c) &= 4a_4c^3 + 3a_3c^2 + 2ac + a_1 \\
&= 4b_4c^3 + 3(b_3 - b_4c)c^2 + 2(b_2 - b_3c)c + (b_1 - b_2c) \\
&= 4b_4c^3 - 3b_4c^3 + 3b_3c^2 - 2b_3c^2 + 2b_2c + b_1 - b_2c.
\end{aligned}$$

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, $j = n, n-1, \ldots, 1$ da seguinte forma:

$c_n = b_n$
$c_j = b_j + c_{j+1}\, c \qquad j = n-1, \ldots, 1.$

Teremos então $p'(c) = c_1$.

MÉTODO DE NEWTON PARA ZEROS DE POLINÔMIOS

Seja $p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$ 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 $a_0, a_1, \ldots, a_n$, 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) deltax = x

2) Para k = 1, . . . , itmax, faça:
   b = a_n
   c = b
   Para i = (n − 1), . . . , 1, faça:
      b = a_i + bx
      c = b + cx
   b = a_0 + bx
   Se |b| ≤ ε_1 vá para o passo 4
   deltax = b / c
   x = x − deltax
   Se |deltax| ≤ ε_2 vá para o passo 4

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

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

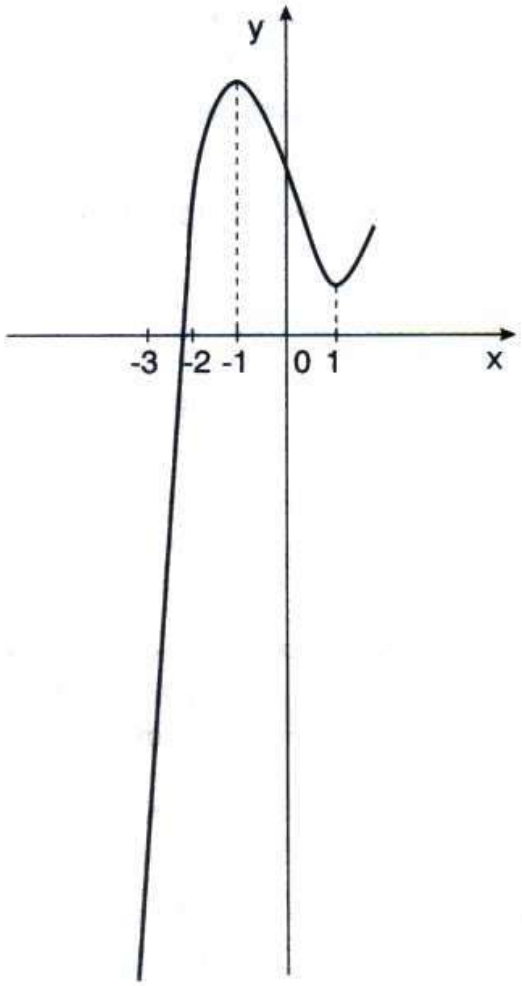


Figura 2.25

EXERCÍCIOS

1. Localize graficamente as raízes das equações a seguir:

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

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

   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:

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

   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 $|\xi - x_k| \leq \dfrac{M}{1 - M} |x_k - x_{k-1}|$;

   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 $\varphi(x) = \dfrac{1}{x} + \dfrac{1}{x^2}$ 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 $\varphi(x) = x - \dfrac{f(x)}{f'(x_0)}$ onde x_0 é a aproximação inicial e é tal que $f'(x_0) \neq 0$.

a) Com o auxílio de um gráfico, escreva a interpretação geométrica deste método.

b) 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:

a) o MPF com $\varphi(x) = \dfrac{1}{2} e^{x/2}$;

b) o método de Newton.
Compare a rapidez de convergência.

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

a) $\text{sen}(x) = 0$

b) $\cos(x) + 1 = 0$

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

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

a) Encontre os valores positivos de k para que f tenha apenas uma raiz estritamente positiva.

b) Encontre os valores positivos de k para que f tenha três raízes estritamente positivas.

18. Seja $f(x) = \dfrac{x^2}{2} + x(\ln(x) - 1)$. 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 $\varepsilon = 10^{-5}$

    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 $\varepsilon = 10^{-5}$.

    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 $\varepsilon = 5 \times 10^{-6}$.

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

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 $|f'(x_0)| \geq \lambda$, a seqüência x_k é gerada através de:

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

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 $\varepsilon = 0.05$.

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 $f(\xi) = f'(\xi) = \ldots = f^{(p-1)}(\xi) = 0$ 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:

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

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,

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

e então

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

O MÉTODO

Se ξ é uma raiz de $f(x) = 0$ com multiplicidade p, dados x_0 uma aproximação inicial para ξ e ε_1, ε_2 precisões desejadas, para k = 1, 2,..., 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

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

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

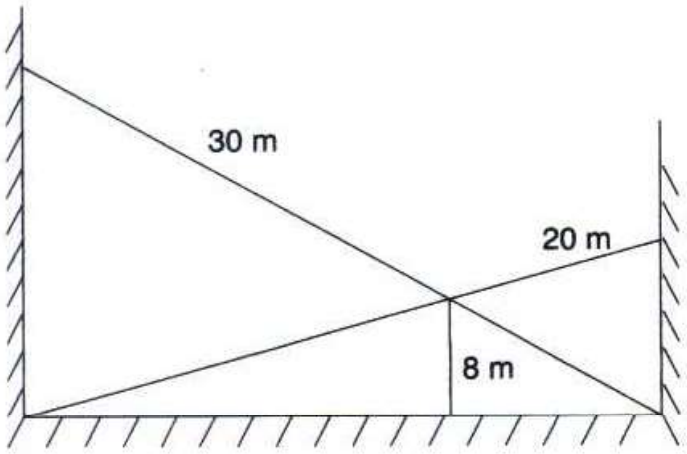
SOL = valor da "solução" encontrada

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)

$x_0 > 2.236067977$

$x_0 < -2.236067977$.