Matemática Computacional

Paulo Rebelo

Índice

Lista de Figuras

Capítulo 1

Introdução

A análise numérica é a disciplina da matemática que se ocupa da elaboração e estudo de métodos que permitem obter, de forma efectiva, soluções numéricas para problemas matemáticos, quando por uma qualquer razão não podemos ou não desejamos usar métodos analíticos.

Um método numérico apresenta uma sucessão que converge para o valor exacto. Cada termo dessa sucessão deve ser visto como uma aproximação - que é possível calcular com um número finito de operações elementares. O objectivo da análise numérica é encontrar sucessões que aproximem os valores exatos com um número mínimo de operações elementares.

Um dos escritos matemáticos mais antigos é a tábua Babilónica YBC 7289, que fornece uma aproximação sexagesimal de $\sqrt{2}$, o comprimento da diagonal de um quadrado unitário.

Figura 1.1: Tábua babilónica YBC 7289

Ser capaz de calcular as faces de um triângulo (e assim, sendo capaz de calcular raízes quadradas) é extremamente importante, por exemplo, em carpintaria e construção. Numa parede quadrada que tem dois metros por dois metros, uma diagonal deve medir $\sqrt{8} \approx 2.83$ metros.

Para perceber melhor o que se pretende dizer por de forma efectiva, consideremos o problema do cálculo do determinante. Como é sabido, o determinante de uma matriz quadrada $A = [a_{ij}]$ para $i, j = 1, \ldots, n$ é dado pela expressão

$$\det(A) = \sum a_{1i_1} a_{2i_2} \cdots a_{ni_n},$$

onde a soma é efectuada sobre todas as $n!$ permutações $(i_1; \ldots; i_n)$ dos números $S_n = \{1, 2, \ldots, n\}$. Esta fórmula teórica só permite o cálculo efectivo do determinante se a dimensão da matriz for muito pequena. Por exemplo, se $n = 25$ o número de permutações possíveis é superior a 15 quatriliões (o valor exacto é 15511210043330985984000000)! Se possuirmos um computador que calcule cada termo da expressão anterior num bilionésimo de segundo, para calcular todas as parcelas necessitamos de 15 biliões de segundos, ou seja 400.000 anos! Os problemas que a análise numérica pretende dar solução são geralmente originários das ciências naturais e sociais, da engenharia, e, como foi dito, não podem, geralmente, ser resolvidos por processos analíticos e que são resolvidos recorrendo a algoritmos.

Um algoritmo é uma sequência finita e não ambígua de instruções para solucionar um problema. Mais especificamente, em matemática, constitui o conjunto de processos (e símbolos que os representam) para efectuar um cálculo. Algoritmos podem ser implementados por programas de computadores.

O conceito de algoritmo é frequentemente ilustrado pelo exemplo de uma receita, embora muitos algoritmos sejam mais complexos. Eles podem repetir passos (fazer iterações) ou necessitar de decisões (tais como comparações ou lógica) até que a tarefa seja completada. Um algoritmo corretamente executado não irá resolver um problema se o algoritmo estiver incorreto ou não for apropriado ao problema.

Um algoritmo não representa, necessariamente, um programa de computador, e sim os passos necessários para realizar uma tarefa. Sua implementação pode ser feita por um computador, por outro tipo de autômato ou mesmo por um ser humano.

Diferentes algoritmos podem realizar a mesma tarefa usando um conjunto diferenciado de instruções em mais ou menos tempo, espaço ou esforço do que outros.

Os algoritmos numéricos são quase tão antigos quanto a civilização humana. Os babilónios, vinte séculos antes de Cristo, já possuiam tabelas de quadrados de todos os inteiros entre 1 e 60. Os egípcios, que já usavam fracções, inventaram o chamado "método da falsa posição"para aproximar as raízes de uma equação. Esse método encontra-se descrito no papiro de Rhind (cerca de 1650 anos antes da era cristã). Na Grécia antiga muitos, foram os matemáticos que deram contributos para o impulso desta disciplina. Por exemplo, Arquimedes de Siracusa ($278 - 212$, a.C.) mostrou que

$$3\frac{10}{71} < \pi < 3\frac{1}{7}$$

e apresentou o chamado método da exaustão para calcular comprimentos, áreas e volumes de figuras geométricas. Este método, quando usado como método para calcular aproximações, está muito próximo do que hoje se faz em análise numérica; por outro lado, foi

também um importante precursor do desenvolvimento do cálculo integral por Isaac Newton (1643-1727) e Gottfried Wilhelm Leibniz (1646-1716). Heron, o velho, no século I a.C., deduziu um procedimento para determinar $\sqrt{a}$ da forma

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right). \tag{1.1}$$

No ano 250 da nossa era, Diofanto obteve um processo para a determinação das soluções de uma equação quadrática. Durante a Idade Média, os grandes contributos para o desenvolvimento da matemática algorítmica vieram, sobretudo, do médio oriente, Índia e China. O contributo maior foi, sem dúvida, a simplificação introduzida com a chamada numeração hindu-árabe. O aparecimento do cálculo e a criação dos logaritmos, no século XVII, vieram dar um grande impulso ao desenvolvimento de procedimentos numéricos. Os novos modelos matemáticos propostos não podiam ser resolvidos de forma explícita e assim tornava-se imperioso o desenvolvimento de métodos numéricos para obter soluções aproximadas. O próprio Newton criou vários métodos numéricos para a resolução de muitos problemas, métodos esses que possuem, hoje, o seu nome. Tal como Newton, muitos vultos da matemática dos séculos XVIII e XIX trabalharam na construção de métodos numéricos. De entre eles podemos destacar Leonhard Euler (1707-1783), Joseph-Louis Lagrange (1736-1813) e Carl Friedrich Gauss (1777-1875). Foi, no entanto, o aparecimento, na década de 40 do século XX, dos primeiros computadores que contribuiu decisivamente para o forte desenvolvimento da disciplina. Apesar de tanto Pascal como Leibniz terem construído, já no séc. XVII, as primeiras máquinas de calcular e de Charles Babage, milionário inglês, ter construído o que é considerado o primeiro computador (nunca funcionou!), foi apenas com o aparecimento do ENIAC, nos anos 40, que a ciência usufruiu, de facto, desses dispositivos de cálculo.

Os métodos numéricos conduzem a soluções aproximadas de um modelo ou sistema exacto. Porquê usar métodos numéricos?

a) Existem situações em que é preferível um método numérico ao método analítico ainda que este exista, por exemplo se a solução para um problema envolve muitos cálculos.

b) A maior parte dos problemas concretos são, em geral, complexos e envolvem fenómenos não lineares pelo que é comum encontrarmo-nos numa situação em que os nossos conhecimentos de matemática não são suficientes para a descoberta de uma solução para um problema real.

c) Quando os dados do problema são os de uma tabela de valores, qualquer tratamento (a sua diferenciação ou integração por exemplo) terá de ser feito através de um método numérico.

Assim, quando estudamos determinado fenómeno é necessário fazer

i) Formulação de um modelo matemático que descreve uma situação real. Tal formulação pode ser feita recorrendo a (sistemas de) equações algébricas, transcendentes, integrais, equações diferenciais, etc... É necessário ter muito cuidado nesta fase uma vez que a grande complexidade dos problemas físicos pode-nos obrigar a fazer simplificações no modelo, simplificações essas que não devem alterar grandemente o comportamento da solução.

ii) Obtenção de um método numérico que permite construir uma solução aproximada para o problema. Um método numérico que possa ser usado para resolver o problema é traduzido por algoritmo que não é mais do que um completo e não ambíguo conjunto de passos que conduzem á solução do problema. Esta fase constitui o cerne da análise numérica. Assim, dado um determinado método numérico, temos necessidade de saber em que condições as soluções por ele obtidas convergem para a solução exacta; em que medida pequenos erros de arredondadmento (e outros) poderão afectar a solução final; qual o grau de precisão da solução aproximada obtida, etc...

iii) Programação automática do algoritmo. Nesta fase teremos necessidade de recorrer a uma linguagem de programação como o Fortran, o Pascal, o C++, entre outras. Mais recentemente é usual o recurso a programas como o Mathematica ou o Matlab.

Um dos problemas mais simples é a avaliação de uma função num determinado ponto. Mas mesmo a avaliação de um polinómio não é sempre trivial: o método de Horner é muitas vezes mais eficiente do que o método óbvio. De forma geral, é importante estimar e controlar o erro de arredondamento que resulta do uso do sistema de ponto flutuante na aritmética. Alguns dos temas abordados pela Análise Numérica são o cálculo de valores de funções, resolução de equações e sistemas de equações, resolução de problemas de valores próprios, cálculo de integrais e solução de equações diferenciais.

Estas notas encontram-se divididas da seguinte forma: no capítulo 2 são apresentados alguns resultados importantes relacionados com a Análise Real; no capítulo 3 são apresentados métodos que nos permitem determinar soluções aproximadas de equações não lineares; no capítulo 4 são apresentados métodos que nos permitem determinar uma solução aproximada para sistemas de equações lineares (métodos directos e indirectos) e para sistemas de equações não lineares (métodos indirectos); no capítulo 5 é abordado o problema da Interpolação polinomial; no capítulo 7 são apresentados métodos que nos permitem determinar o valor aproximado de um integral e no último capítulo é dedicado à solução numérica de equações diferenciais ordinárias.

Capítulo 2

Alguns conceitos e resultados importantes

2.1 Representação de números

Um número real x é representado na forma decimal (isto é, na base 10) por:

- o seu sinal ($+$ ou $-$);
- por uma sequência (finita ou não) de algarismos do conjunto $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ posicionada relativamente ao ponto (ou vírgula) decimal (.), ou seja:

$$x = \pm d_n d_{n-1} \dots d_1 d_0 . d_{-1} d_{-2} d_{-3} \dots$$

A necessidade de representar números de diferentes grandezas de uma forma compacat levou à introdução da designada **notação científica**, que não é mais do que a introdução na representação de um factor multiplicativo correspondente a uma potência inteira da base de representação, ou seja, de base 10, isto é,

$$x = \pm d_n d_{n-1} \dots d_1 d_0 . d_{-1} d_{-2} d_{-3} \dots \times 10^n$$

A parte da representação $d_n d_{n-1} \dots d_1 d_0 . d_{-1} d_{-2} d_{-3}$ é designada por **mantissa** e o número inteiro n é designada por **expoente**. A localização do ponto decimal pode ser alterada bastando para isso modificar o valor do expoente.

Por exemplo, o número 10.23 pode ser representado por

$$1.023 \times 10^1, \;\; 0.1023 \times 10^2, 102.3 \times 10^{-1}, \dots$$

Na prática, podemos utilizar unicamente representações finitas e por vezes, não queremos ou não podemos utilizar mais do que um dado número de algarismos da mantissa, coloca-se o problema de como representar um valor exacto que à partida não será representável. Isto é, suponhamos que temos um determinado número

$$d_1 d_2 \dots d_n d_{n+1} d_{n+2},$$

e que apenas podemos utilizar os n primeiros algarismos. Este problema pode ser resolvido por dois processos distintos: **truncatura** e **arredondamento**.

No caso da truncatura, ignoram-se os algarismos da mantissa a partir do índice $n+1$, tendo em conta que os que correspondam a algarismo inteiros devem ser substituídos por zeros e posteriormente eliminados por alteração do expoente. A representação assim obtido difere do valor original menos do que uma unidade da última casa decimal não eliminada.

2.2 Sistemas de vírgula flutuante

A representação mais comum de números reais em sistemas computacionais é realizada em *vírgula flutuante*. Um sistema de vírgula flutuante é habitualmente caracterizado por 3 parâmetros:

1. a base de representação (β),
2. o número de dígitos da mantissa (n),
3. número máximo de dígitos no expoente (ou então, os valores máximos e mínimos do expoente m e M, respectivamenteSe bem que neste caso é caracterizado por 4 parâmetros.).

Para representar um sistema de numeração vamos utilizar a seguinte notação:

$$FP\left(\beta, t, \varepsilon\right), \tag{2.1}$$

para designar um número expresso na base β, cuja mantissa tem no máximo t dígitos e cujo expoente pode ter no máximo ε dígitos.

Diz-se ainda que um sistema de vírgula flutuante se encontra *normalizado* se apenas permitir representações de números cujo primeiro algarismo da mantissa seja diferente de zero, isto é, $d \neq 0$, isto para além de permitir a representação do número zero. Independentemente de se tratar de um sistema normalizado ou não, qualquer sistema de vírgula flutuante terá a si associado o número diferente de zero com menor valor absoluto representável bem como o número com o maior valor absoluto representável.

Dizemos que um número está representado no sistema de ponto flutuante $FP\left(t, b, \varepsilon\right)$ se estiver na forma

$$x = \pm \left(m_1 m_2 \cdots m_t\right)_b \times b^{\pm (e)_b}, \tag{2.2}$$

ou de forma equivalente,

$$x = \pm \left(0.m_1 m_2 \cdots m_t\right)_b \times b^{\pm (c_{\varepsilon-1} c_{\varepsilon-2} c_{\varepsilon-3} \cdots c_1 c_0)}. \tag{2.3}$$

Em (2.2), b é a base, $(m_1 m_2 \cdots m_t)$ é a mantissa e $(e)_b$ é o expoente. Quando em (2.3), $m_1 \neq 0$, dizemos que o número está representado na sua forma **normalizada**.

Em resumo, um número de $n-$dígitos numa base b tem a representação

$$x = \pm \left(.m_1 m_2 m_3 \ldots m_n\right)_b b^{\pm e}, \tag{2.4}$$

onde $(.m_1 m_2 m_3 \ldots m_n)_b$é uma fração na base b, designada por mantissa, e e é designado por expoente. O número encontra-se na forma normalizada se $m_1 \neq 0$ ou então $m_1 =$

$m_2 = m_3 = \cdots = m_n = 0$. Habitualmente, os computadores utilizam $b = 2$, em alguns $b = 16$ e na maioria das máquinas de calcular e máquinas de escritório, $b = 10$.

Consideremos o sistema $FP(10, 6, 2)$, isto é, utilizamos a base 10, a mantissa tem 6 dígitos e o expente tem no máximo 2 dígitos.

Obviamente, este sistema de numeração não permite a representação de todos os números reais. Por exemplo

- o número 0.1×10^{500} não se pode representar no formato $FP(10, 6, 2)$ pois o expoente "ocupa"3 dígitos e o sistema só permite a utilização de 2 - esta situação é designada por *overflow*;
- o número 0.555×10^{-100} não se pode representar no formato $FP(10, 6, 2)$ pois o expoente "ocupa"3 dígitos e o sistema só permite a utilização de 2 - esta situação é designada por *underflow*;
- o número $\sqrt{2} \approx 1.4142135623730950488$ também não pode ser representado neste formato pois é um número irracional (a mantissa tem um número infinito de dígitos).

Se o número x tiver representação exacta em $FP(b, t, \varepsilon)$, então escrevemos $Fl(x) = x$. Portanto, coloca-se o problema de como representar os números que não se podem representar neste formato. Existem basicamente duas técnicas para resolver este problema:

- **Truncatura**: desprezam-se simplesmente os dígitos do número real x que não cabem na mantissa, isto é, os dígitos da mantissa além dos t primeiros são desprezados.
- **Arredondamento**: o número real x é representado pelo número do sistema $FP(\beta, t, \varepsilon)$ que lhe está mais próximo.

Se o número x tiver representação exacta em $FP(b, t, \varepsilon)$, então escrevemos $Fl(x) = x$. Se tal representação não for possível, isto é, se a representação não for exacta, ex Para indicar o tipo de técnica que é escolhida, empregamos as notações $FP(\beta, t, \varepsilon, T)$ e $FP(\beta, t, \varepsilon, A)$ para **T**runcatura e **A**rredondamento, respectivamente

Exemplo 2.2.1 *Represente, na forma normalizada, o número $x = 372.526$ nos sistemas de vírgula flutuante $FP(10, 5, 2, A)$ e $FP(10, 5, 2, T)$.*

Resolução 2.2.1 *Movendo o ponto decimal para o início do número fica,*

$$x = 0.372526 \times 10^3.$$

Uma vez que $t = 5$, isto é, só podemos ter 5 dígitos na mantissa, obtemos:

- *em $FP(10, 5, 2, A)$, 0.37253×10^3;*
- *em $FP(10, 5, 2, T)$, 0.37252×10^3;*

■

Exemplo 2.2.2 *Calcule, em $FP(10, 4, 2, T)$, $x + y$, para $x = 1.256879$ e $y = 0.985441$.*

Resolução 2.2.2 *Temos* $\hat{x} = 1.256$ *e* $\hat{y} = 0.9854$, *logo* $z = \hat{x}+\hat{y} = 0.1256+0.9854 = 1.111$. *Então* $\hat{z} = 0.1111$. ■

Nota 2.2.1 *Uma outra notação para representar um sistema de numeração é a seguinte:*

$$FP(\beta, n, m, M).$$

Portanto, dizer que $x \in FP(\beta, n, m, M)$ *é equivalente a escrever*

$$x = \pm (0.d_1 d_2 \cdots d_n) \times \beta^r, \tag{2.5}$$

onde r *é um inteiro tal que* $m \leq r \leq M$, *e* $0 \leq d_i \leq \beta - 1$, *para* $i = 0, 1, \ldots, n-1$ *são dígitos na base* β.

Habitualmente tem-se que $m < 0 < M$, *de forma a tornar possível representar números com valores absolutos menores e maiores do que a unidade. Habitualmente, os sistemas computacionais utilizam sistemas de vírgula flutuante de base* 2, *de forma a que apenas seja necessário utilizar os dígitos "*0*"e "*1*". Obviamente que um sistema de vírgula flutuante apenas permite representar um subconjunto finito de números reais. Nestes sistemas, o conjunto de expoentes permitidos limita a gama de valores representáveis e o número de dígitos da mantissa caracteriza a precisão com que se podem aproximar números que não tenham representação exacta.*

No caso em que há limites superior e inferior para os expoentes, facilmente se verifica que uma vez fixado sistema de vírgula flutuante $FP(\beta, n, m, M)$ *e dois elementos* $\alpha,\ \beta \in FP(\beta, n, m, M)$ *o resultado das operações aritméticas usuais,* $+, -, \times, \div$ *pode não pertencer a* $FP(\beta, n, m, M)$.

Para a soma e para a subtracção, obtemos a seguinte regra

$$\alpha \pm \beta = m_1 \times \beta^{n_1} \pm m_2 \times \beta^{n_2} = \begin{cases} \left(m_1 \pm m_2 \times \beta^{-(n_1-n_2)}\right) \times \beta^{n_1} & se \quad n_1 > n_2 \\ \left(m_1 \times \times \beta^{-(n_2-n_1)} \pm m_2\right) \times \beta^{n_2} & se \quad n_1 \leq n_2 \end{cases} \tag{2.6}$$

enquanto que para a multiplicação e para a divisão, temos:

$$\alpha \times \beta = (m_1 \times m_2) \times \beta^{n_1+n_2} \ e \ \frac{\alpha}{\beta} = \left(\frac{m_1}{m_2}\right) \times \beta^{n_1-n_2}. \tag{2.7}$$

2.3 Mudança de base

Um dos primeiros aspectos numéricos que apreendemos é o sistema de numeração árabe. Há nesse sistema um grande avanço face aos que o precederam historicamente (ex: gregos, romanos), quer na facilidade de execução de operações, quer no optimizar da extensão da notação. O sistema de numeração árabe é um avanço numérico que passou despercebido à matemática grega, mais preocupada com propriedades conceptuais relacionadas com a geometria. O sistema árabe permitiu que o cálculo numérico fosse facilmente mecanizado no que diz respeito às operações elementares, libertando-se de qualquer interpretação aplicada à geometria ou a qualquer outro modelo intuitivo. O exemplo

máximo dessa mecanização teve como pioneiros B. Pascal ou C. Babbage que construiram as primeiras máquinas de calcular mecânicas e que culminaram, passados quase três séculos, no aparecimento de computadores electrónicos. Por uma questão de eficácia, a nível interno os computadores trabalham com um sistema de numeração binária ao invés do habitual sistema decimal que utilizamos correntemente. Os sistemas de numeração tem por objectivo prover símbolos e convenções para representar quantidades, de forma a registrar a informação quantitativa e poder processá-la.

Nos sistemas de *numeração posicional*, o valor representado pelo algarismo no número depende da posição em que ele aparece na representação.

O método ao qual estamos acostumados usa um sistema de numeração posicional. Isso significa que a posição ocupada por cada algarismo em um número altera seu valor de uma potência de 10 (na base 10) para cada casa à esquerda.

Por exemplo, no sistema decimal (base 10), no número 125 o algarismo 1 representa 100 (uma centena ou 10^2), o 2 representa 20 (duas dezenas ou 1×10^1) e o 5 representa 5 mesmo (5 unidades ou 5×10^0). Assim, em nossa notação,

$$125 = 1 \times 10^2 + 2 \times 10^1 + 5 \times 10^0 \tag{2.8}$$

Definição 2.3.1 *A base de um sistema é a quantidade de algarismos disponível na representação.*

Como há dígitos comuns a várias bases, é necessário indicar qual a base que estamos a utilizar. Por exemplo, o número

$$x = 11.01,$$

pode estar representado em várias bases. Desde a base 2 até ... Porquê?

Assim, para evitar confusões, indicamos da seguinte forma a base em que o número está representado:

$$x = (11.01)_2$$

Os computadores utilizam a base 2 (sistema binário) e os programadores, por facilidade, usam em geral uma base que seja uma potência de 2, tal como 24 (base 16 ou sistema hexadecimal) ou eventualmente ainda 23 (base 8 ou sistema octal).

Na base 10, dispomos de 10 algarismos para a representação do número: 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9.

Na base 2, seriam apenas 2 algarismos: 0 e 1.

Na base 16, seriam 16: os 10 algarismos aos quais estamos acostumados, mais os símbolos A, B, C, D, E e F, representando respectivamente 10, 11, 12, 13, 14 e 15 unidades.

Generalizando, temos que uma base b qualquer disporá de b algarismos, variando entre 0 e $(b-1)$.

A representação $125,38_{10}$ na base 10 significa

$$125,38_{10} = 1 \times 10^2 + 2 \times 10^1 + 5 \times 10^0 + 3 \times 10^{-1} + 8 \times 10^{-2}$$

Generalizando, representamos uma quantidade N qualquer, numa dada base b, com um número tal como segue:

$$N_b = a_n b^n + ... + a_2.b^2 + a_1.b^1 + a_0.b_0 + a_{-1}.b^{-1} + a_{-2}.b^{-2} + ... + a_{-n}.b^{-n} \tag{2.9}$$

sendo que

$$a_n b^n + ... + a_2.b^2 + a_1.b^1 + a_0.b_0 \tag{2.10}$$

é a parte inteira e

$$a_{-1}b^{-1} + a_{-2}b^{-2} + ... + a_{-n}b^{-n} \tag{2.11}$$

é a parte fraccionaria.

Portanto, uma vez que vamos efectuar divisões inteiras, convém rever o algoritmo da divisão inteira:

$$N = Q \times D + R, \tag{2.12}$$

onde n é o dividendo, Q quociente, D divisor e R é o resto, que satisfaz a condição $R < D$.

Intuitivamente, sabemos que o maior número que podemos representar, com n algarismos, na base b, será o número composto n vezes pelo maior algarismo disponível naquela base (ou seja, $b-1$). Por exemplo, o maior número que pode ser representado na base 10 usando 3 algarismos será 999 (ou seja, $10^3 - 1 = 999$).

Generalizando, podemos ver que o maior número inteiro N que pode ser representado, em uma dada base b, com n algarismos (n "casas"), será $N = b^n - 1$. Assim, o maior número de 2 algarismos na base 16 será $FF16$ que, na base 10, equivale a 255_{10}.

2.3.1 Números em decimal e sua representação em binário, octal e hexadecimal

De seguida apresentamos uma tabela para alguns números nas bases 2, 8, 10 e 16:

Base 10	Base 2	Base 8	Base 16
0	0	0	0
1	1	1	1
2	10	2	2
3	11	3	3
4	100	4	4
5	101	5	5
6	110	6	6
7	111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

Nota 2.3.1 *A base 16 ou sistema hexadecimal pode ser indicada também por um "H"ou "h"após o número; por exemplo: FFH significa que o número FF (ou 255 em decimal) está em hexadecimal. Não confundir o "H"ou "h"com mais um dígito, mesmo porque em hexadecimal só temos algarismos até "F"e portanto não existe um algarismo "H".*

Exercício 2.3.1 *Qual é a representação do número 16_{10} em binário, octal e hexadecimal?*

2.3.2 Conversão de Números em uma base b qualquer para a base 10

Vamos lembrar a expressão geral já apresentada:

$$N_b = a_n b^n + ... + a_2.b^2 + a_1.b^1 + a_0.b_0 + a_{-1}.b^{-1} + a_{-2}.b^{-2} + ... + a_{-n}.b^{-n} \qquad (2.13)$$

A melhor forma de fazer a conversão é usando essa expressão.

Tomando como exemplo o número 101101_2, vamos calcular seu valor representado na base dez. Usando a expressão acima, fazemos:

$101101_2 = 1\times 2^5 + 0\times 2^4 + 1\times 2^3 + 1\times 2^2 + 0\times 2^1 + 1\times 2^0 = 32+0+8+4+0+1 = 45_{10}$

Podemos fazer a conversão de números em qualquer base para a base 10 usando o algoritmo acima.

Exemplo 2.3.1 *1. Converter $4F5_H$ para a base 10. Lembramos que o H significa que a representação é hexadecimal (base 16). Sabemos ainda que $F_{16} = 15_{10}$. Então:*

$$\begin{aligned} 4F5_H &= 4\times 16^2 + 15\times 16^1 + 5\times 16^0 \\ &= 4\times 256 + 15\times 16 + 5 \\ &= 1024 + 240 + 5 \\ &= 1269_{10} \end{aligned}$$

2. *Converter* 3485_9 *para a base 10.*

$$\begin{aligned} 3485_9 &= 3 \times 9^3 3 + 4 \times 9^2 + 8 \times 9^1 + 5 \times 9^0 \\ &= 3 \times 729 + 4 \times 81 + 8 \times 9 + 5 \\ &= 2187 + 324 + 72 + 5 \\ &= 2588_{10}. \end{aligned}$$

3. *Converter* $7G16$ *para a base 10.*

   Uma base b dispõe dos algarismos entre 0 *e* $(b-1)$*. Assim, a base* 16 *dispõe dos algarismos* 0 *a* F *e portanto o símbolo* G *não pertence à representação hexadecimal.*

4. *Converter* $1001,01_2$ *para a base 10.*

$$\begin{aligned} 1001,01_2 &= 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} \\ &= 8 + 0 + 0 + 1 + 0 + 0,25 \\ &= 9,25_{10} \end{aligned}$$

5. *Converter* $34,3_5$ *para a base 10.*

$$3 \times 5^1 + 4 \times 5^0 + 3 \times 5^{-1} = 15 + 4 + 0,6 = 19,6_{10}$$

6. *Converter* $38,3_8$ *para a base 10.*

   Uma base b dispõe dos algarismos entre 0 e $(b-1)$*. Assim, a base 8 dispõe dos algarismos* 0 *a* 7 *e portanto o algarismo 8 não existe nessa base. A representação* 38.3 *não existe na base 8.*

A conversão de números da base dez para uma base qualquer emprega algoritmos que serão o inverso dos acima apresentados. Os algoritmos serão melhor entendidos pelo exemplo que por uma descrição formal. Vamos a seguir apresentar os algoritmos para a parte inteira e para a parte fraccionaria:

- **Parte Inteira**: O número decimal será dividido sucessivas vezes pela base; o resto de cada divisão ocupará sucessivamente as posições de ordem 0, 1, 2 e assim por diante até que o resto da última divisão (que resulta em quociente zero) ocupe a posição de mais alta ordem. Veja o exemplo da conversão do número 19 (10) para a base 2:

19	2				
$\boxed{a_0 = 1}$	9	2			
	$\boxed{a_1 = 1}$	4	2		
		$\boxed{a_2 = 0}$	2	2	
			$\boxed{a_3 = 0}$	1	2
				$\boxed{a_4 = 1}$	0

Logo

$$19_{10} = 10011_2 \tag{2.14}$$

Experimente fazer a conversão contrária (retornar para a base 10) e ver se o resultado está correcto.

- **Parte fraccionária**: Se o número for fracionário, a conversão se fará em duas etapas distintas: primeiro a parte inteira e depois a parte fracionária. Os algoritmos de conversão são diferentes. O algoritmo para a parte fraccionaria consiste de uma série de multiplicações sucessivas do número fracionário a ser convertido pela base; a parte inteira do resultado da primeira multiplicação será o valor da primeira casa fracionária e a parte fracionária será de novo multiplicada pela base; e assim por diante, até o resultado dar zero ou até encontrarmos o número de casas decimais desejado.

  Isto é, seja x um número real dado através da sua representação decimal e suponhamos que pretendemos obter a sua representação na base b. Como a conversão da parte inteira segue as regras deduzidas anteriormente, vamos admitir, para simplificar que x é um número fraccionário puro, isto é, sem parte inteira. Nestas condições temos

$$x = (.d_{-1}d_{-2}\ldots d_{-k})_b = d_{-1} \times b^{-1} + d_{-2} \times b^{-2} + \cdots + d_{-k} \times b^{-k}. \tag{2.15}$$

  Se multiplicarmos x por b, verificamos que d_{-1} é a parte inteira do resultado e $d_{-2} \times b^{-1} + \cdots + d_{-k} \times b^{-k+1}$ a parte fraccionária. Multiplicando esta novamente por b e tomando novamente a parte inteira do resultado obtemos d_{-2}, e assim sucessivamente.

Exemplo 2.3.2 *Escreva na base* 2 *o número* .625. *De acordo com o exposto temos sucessivamente:*

$$\begin{aligned} .625 \times 2 &= 1.250 \\ .250 \times 2 &= 0.500 \\ 0.500 \times 2 &= 1.000 \end{aligned}$$

Logo,

$$.0625_{10} = (.101)_2 .$$

Exemplo 2.3.3 *Escreva na base* 2 *o número* .1.
De igual modo temos

$$\begin{aligned} .1 \times 2 &= 0.2 \\ 0.2 \times 2 &= 0.4 \\ 0.4 \times 2 &= 0.8 \\ 0.8 \times 2 &= 1.6 \\ 0.6 \times 2 &= 1.2 \\ 0.2 \times 2 &= 0.4 \\ &\vdots \end{aligned}$$

Logo,

$$.1 = (.000110011\ldots)$$

Verificamos que, enquanto o número .625 tem uma representação binária finita, já o mesmo não acontece com o número .1 cuja representação binária é infinita com o grupo 0011 repetindo-se periodicamente. Este número, não é pois representável numa máquina com capacidade finita.

Consideremos o seguinte

Exemplo 2.3.4 *Escreva o número* 0.562510 *na base* 2*. A parte inteira é* $0_{10} = 0_2$ *e a parte fraccionaria é igual a* 0.562510*. Multiplica-se a parte fraccionaria por* 2 *sucessivamente, até que ela seja igual a zero ou cheguemos na precisão desejada.*

$$\begin{aligned} \textit{fracção} \times 2 &= \textit{vai-um} + \textit{fracção seguinte} \\ 0.5625 \times 2 &= 1 + 0.1250 \\ 0.1250 \times 2 &= 0 + 0.2500 \\ 0.2500 \times 2 &= 0 + 0.5000 \\ 0.5000 \times 2 &= 1 + 0.0000 \end{aligned}$$

Nesta linha a fracção ficou igual a zero e finalizamos a conversão.
Anotando a sequência de vai-um na ***ordem de cima para baixo****, temos:* 1001.
Portanto,

$$0.5625_{10} = 0.1001_2$$

No entanto, é mais comum nunca pararmos o processo na multiplicação seguinte. Neste caso, devemos parar as multiplicações quando atingirmos uma certa precisão desejada.

Vamos agora considerar o caso em que pretende-mos passar de uma base b para a base 10. Seja $x = (.d_{-1}d_{-2}\ldots d_{-k})$. A sua representação decimal obtêm-se aplicando directamente a expressão

$$\begin{aligned} x &= .d_{-1}d_{-2}\ldots d_{-k} && (2.16) \\ &= \sum_{d=1}^{k} d_{-j} \times b^{-j} && (2.17) \end{aligned}$$

Deste modo,

$$\begin{aligned}(.561)_8 &= 5\times 8^{-1}+6\times 8^{-2}+1\times 8^{-3}\\ &= .720703125\end{aligned}$$

Outro método para conversão de decimal para binário

Considere alguns resultados da potência 2^n e exponha-os em tabela por ordem decrescente:

2^{11}	2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
2048	1024	512	256	128	64	32	16	8	4	2	1

Deste modo é possível converter grandes quantidades de números decimais para binários:

Número decimal	2048	1024	512	256	128	64	32	16	8	4	2	1	Resultado Binário
354	0	0	0	1	0	1	1	0	0	0	1	0	101100010
1634	0	1	1	0	0	1	1	0	0	0	1	0	11001100010
104	0	0	0	0	0	1	1	0	1	0	0	0	1101000
2	0	0	0	0	0	0	0	0	0	0	1	0	10
38	0	0	0	0	0	0	1	0	0	1	1	0	100110
57	0	0	0	0	0	0	1	1	1	0	0	1	111001

O procedimento é igual a qualquer caso. Vamos acompanhar de perto o caso do 1634 por exemplo: O procedimento se inicia do extremo esquerdo, e consiste na verificação de uma possível subtracção não-negativa.

- $2048 > 1634$. Logo fica "0"(por exemplo 1634-2048 resultava num número negativo. Logo atribuí-se o "0")
- $1024 < 1634$. Logo fica "1"
- $1634 - 1024 = 610$
- $512 < 610$. Logo fica "1"
- $610 - 512 = 98$
- $256 > 98$. Logo fica "0"
- $128 > 98$. Logo fica "0"
- $64 < 98$. Logo fica "1"
- $98 - 64 = 34$
- $32 < 34$. Logo fica "1"
- $34 - 32 = 2$
- $16 > 2$. Logo fica "0"
- $8 > 2$. Logo fica "0"

- $4 > 2$. Logo fica "0"
- $2 = 2$. Logo fica "1"
- $2 - 2 = 0$
- $1 > 0$. Logo fica "0"

Onde, "1"significa verdadeiro e "0"os que não interessam ou que dão números negativos. Este método, quando executado mentalmente, permite uma velocidade incrível na conversão. Bem como na facilidade para converter muitos números devido à sua estructura em tabela.

Repara que existem "0"à esquerda do primeiro "1"que são ignorados? Eles são ignorados porque qualquer "0"à esquerda do primeiro "1"da sequência, não vale nada, portanto omite-se.

Observação 2.3.1 *Em ambos os casos, a conversão foi interrompida quando encontramos o número de algarismos fracionarios solicitadas no enunciado. No entanto, como não encontramos resultado 0 em nenhuma das multiplicações, poderíamos continuar efectuando multiplicações indefinidamente até encontrar (se encontrarmos) resultado zero. No caso de interrupção por chegarmos ao número de dígitos especificado sem encontramos resultado zero, o resultado encontrado é aproximado e essa aproximação será função do número de algarismos que calcularmos.*

Para converter números de uma base b para uma outra base b' quaisquer (isso é, que não sejam os casos particulares anteriormente estudados), o processo prático utilizado é converter da base b dada para a base 10 e depois da base 10 para a base b' pedida.

Exemplo: Converter 43_5 para $()_9$. Temos então

$$43_5 = (4 \times 5 + 3)_{10} = 23_{10} \Rightarrow \frac{23}{9} = 2(\text{ resto } 5) \text{ logo } 43_5 = 23_{10} = 25_9.$$

2.3.3 Bases que são potências entre si

As conversões mais simples são as que envolvem bases que são potências entre si, como por exemplo, conversão entre as bases 2 e $8 = 2^3$. Vamos apresentar como exemplo, a conversão entre a base 2 e a base 8. Como $2^3 = 8$, separando os bits de um número binário em grupos de três bits (começando sempre da direita para a esquerda!) e convertendo cada grupo de três bits para seu equivalente em octal, teremos a representação do número em octal. Por exemplo:

$$10101001_2 = 10.101.001_2$$

(separando em grupos de 3, sempre começando da direita para a esquerda).

Sabemos que

$$010_2 = 28;\ 101_2 = 58\ ; 001_2 = 1_8,$$

portanto

$$10101001_2 = 251_8$$

Vamos agora exemplificar com uma conversão entre as bases 2 e 16. Como $2^4 = 16$, basta separarmos em grupos de 4 bits (começando sempre da direita para a esquerda!) e converter. Por exemplo:

$$11010101101_2 = 110.1010.1101_2$$

(separando em grupos de 4 bits, sempre começando da direita para a esquerda) Sabemos que $1102 = 6_{16}$; $10102 = A_{16}$; $11012 = D_{16}$; portanto $11010101101_2 = 6AD_{16}$

Vamos agora exercitar a conversão inversa. Quanto seria $3F5_H$ (lembrar que o H está designando "hexadecimal") em octal? O método mais prático seria converter para binário e em seguida para octal.

$$3F5H = 11.1111.01012$$

(convertendo cada dígito hexadecimal em 4 dígitos binários) = = 1.111.110.1012 (agrupando de três em três bits) = = 17658 (convertendo cada grupo de três bits para seu valor equivalente em octal).

2.4 Erros

Os dados de um determinado problema, podem estar à partida afectados de imprecisões resultantes de medições incorrectas. Note-se que a escala de um instrumento de medição nos dá uma possibilidade de saber um limite superior para o erro com que esses valores vêm afectados. Por exemplo, com uma régua usual, a medição de uma distância de 2mm pode vir afectada com um erro de 0.5mm o que dá um *erro relativo* de $2,5\%$. Outra causa de erro resulta das simplificações impostas ao modelo matemático usado para descrever um determinado fenómeno físico. Por exemplo, é usual considerar que, para um dado problema, não há perdas de calor, o atrito é nulo, etc... Este tipo de erros fogem ao controlo do analista numérico e são muito difíceis de quantificar.

Outra causa de erros resulta da forma como representamos os números reais. De facto, quando usamos um computador, a mantissa de um número tem que ser limitada. Assim, existem números que não possuem representação na máquina que estamos a trabalhar. Por exemplo, o número $x = 123.9346$ não tem representação numa máquina de base decimal cuja mantissa só permita armazenar 6 dígitos. Temos assim necessidade de o aproximar por um outro que possa ser representado na referida máquina. Essa aproximação vai ser efectuada por um processo conhecido por *arredondamento*.

Quase todos os cálculos envolvem erros. Em cálculo numérico lidamos quase exclusivamente com valores aproximados daí que não podemos usar métodos numéricos e ignorar a existência de erros. É por essa razão que é importante indicar quais são as *regras de arredondamento*:

1. Se o primeiro dígito a desprezar for maior que 5, ou for 5 seguido não só de zeros, soma-se uma unidade ao último dígito a reter. Caso contrário, o último dígito a reter não será alterado. Se o primeiro e único dígito a desprezar é 5, ou cinco seguido de zeros, o último dígito a reter deverá ser aumentado de uma unidade apenas se esse último dígito for ímpar.

2. Na adição e subtracção, arredonda-se por forma a que o último dígito a reter, na resposta, corresponda ao último algarismo mais significativo nos números a serem somados ou subtraídos entre si.

3. Na multiplicação e divisão, arredondar por forma a que o número de algarismos significativos no resultado iguale o menor número de algarismos significativos dos números intervenientes na operação em causa.

4. Em combinações das operações aritméticas, as operações dentro dos parentesis são executadas e os resultados respectivos arredondados antes de prosseguir com a outra operação, em vez de arredondar apenas o resultado final.

Por exemplo:

$$x = 123.9346 \approx 123.935 = x^*. \tag{2.18}$$

Consideremos agora um problema cuja solução é um número real. Este valor é designado por **valor exacto** do problema e, no que se segue será representado por x. Designa-se por **valor aproximado** ou **aproximação** do valor exacto x, e representa-se por x^*, qualquer valor que se pretende utilizar como solução do problema. Associado a um dado valor aproximado x^* define-se por **erro de aproximação** à diferença entre o valor exacto e o valor aproximado. Habitualmente, é representado por Δx^* e é dado por

$$\Delta x^* = x - x^*.$$

No caso em que:

- $x^* < x$, a aproximação diz-se **por defeito** e neste caso $\Delta x^* > 0$;
- $x^* > x$, a aproximação diz-se **por excesso** e neste caso $\Delta x^* < 0$.

Exemplo 2.4.1 *É sabido que* $\pi \approx 3.1415926535897932385$. *Indique uma* 3 *aproximações por defeito e por excesso para* π.

Resolução 2.4.1 *Então,*

$$3 \quad 3.1 \quad 3.14 \quad 3.141 \quad \cdots$$

são aproximações de π *por defeito e*

$$4 \quad 3.2 \quad 3.15 \quad 3.142 \quad \cdots$$

são aproximações de π *por excesso.* ■

O valor absoluto do erro de aproximação é designado por

Definição 2.4.1 (Erro absoluto) *O erro absoluto de um número aproximado* x^*, *que substitui um número exacto* x, *é o valor absoluto da diferença entre eles. O número* Δ *que satisfaz a desigualdade*

$$E_a(x^*) = |x - x^*| \leq \Delta \tag{2.19}$$

é designado por limite do erro absoluto.

É habitual indicar o erro máximo absoluto por ε. Então, se x^* for um valor aproximado de x como um erro máximo absoluto ε, verifica-se que

$$x \in [x^* - \varepsilon, x^* + \varepsilon]\,. \tag{2.20}$$

Neste caso é habitual utilizar-se a notação: $x = x^* \pm \varepsilon$.

Exemplo 2.4.2 *O que se pretende dizer ao escerver:* $x = 1.23 \pm 0.02$*?*

Resolução 2.4.2 *Pretende-se dizer que* 1.23 *é uma aproximação de* x *com um erro máximo absoluto de* 0.02*, ou seja, isto significa que* x *se encontra no intervalo* $[1.21, 1.25]$ ■

Exemplo 2.4.3 *Indique qual o valor truncado às décimas dos números* 123.56 *e* 123.51 *e às centenas o número* 7395*.*

Resolução 2.4.3 *Os valores truncados às décimas são, em ambos os casos o número* 123.5*. Ao truncar o número* 7395 *para as centenas obtemos* 7300*, isto é,* 73×10^2*.* ■

A notação $x = x^* \pm \varepsilon$ tende para ser pouco prática. Uma forma de tornar mais simples a representação de aproximações é considerar majorantes do erro absoluto apenas da forma 0.5×10^n e representar apenas a aproximação até à casa decimal 10^n, ficando implicíto qual o majorante do erro absoluto. Quando se utiliza esta notação, os algarismo da mantissa de uma representação, com excepção dos zeros à esquerda são designados por **algarismos significativos**. Devemos salientar que esta simplificação implica a perda de informação, pois o erro máximo absoluto inicial, ε, é substituído por um majorante da forma 0.5×10^n.

Definição 2.4.2 *Seja* x^* *uma aproximação para o valor exacto* x*. Então,*

- *Dizemos que* x^* *tem* k *casas decimais correctas sse* $E_a(x^*) \leq 0.5 \times 10^{-k}$*;*
- *Dizemos que* x^* *tem* k *algoritmos significativos correctos sse* $E_r(x^*) \leq 5 \times 10^{-k}$*.*

Definição 2.4.3 (Número de casas decimais exactas) *Diz-se que o número aproximado* x^* *escrito na forma decimal, tem* n *números decimais exactas em sentido estrito, se o valor absoluto do erro deste número não excede em* $\frac{1}{2}$ *da unidade decimal de ordem* n*. Neste caso, quando* $n > 1$*, pode-se tomar como limite do erro relativo o número*

$$\delta = \frac{1}{2k}\left(\frac{1}{10}\right)^{n-1}, \tag{2.21}$$

onde k *é o primeiro dígito de valor do número* x^**. Ao contrário, se sabemos que*

$$\delta \leq \frac{1}{2(k+1)}\left(\frac{1}{10}\right)^{n-1}, \tag{2.22}$$

o número x^* *tem* n *casas decimais exactas em sentido estrito. Em particular, o número* x^* *tem* n *casas decimais em sentido estrito se*

$$\delta \leq \frac{1}{2}10^{-n}. \tag{2.23}$$

Isto é, se x^* é uma aproximação para x, dizemos que x^* tem k casas decimais correctas se e somente se $|x - x^*| \leq 0.5 \times 10^{-k}$ e dizemos que x^* tem k algarismos significativos se e somente se $E_r(x^*) \leq 5 \times 10^{-k}$.

Em muitas situações, o erro absoluto não é uma boa medida da qualidade da aproximação calculada. Por exemplo, num cálculo de Engenharia Civil para o projecto duma auto-estrada, um erro de 1cm na largura das faixas é perfeitamente aceitável. O mesmo não acontece num cálculo para o projecto duma peça de automóvel, onde um erro de 1cm pode não ser aceitável. Ou seja, quando estamos a trabalhar com números grandes, os erros aceitáveis são maiores do que quando trabalhamos com números pequenos. A noção de erro relativo modela melhor esta ideia.

Definição 2.4.4 (Erro relativo) *O erro relativo de um número aproximado x^* que substitui um número exacto x é a razão entre erro absoluto e o valor exacto. O número δ que satisfaz a desigualdade*

$$E_r(x^*) = \frac{|x - x^*|}{|x|} \leq \delta \tag{2.24}$$

é designado por limite do erro relativo do número aproximado x^*. Como praticamente $x^* \approx x$, habitualmente consideramos que $\delta = \dfrac{\Delta}{x^*}$.

Em relação ao número considerado em (2.18), temos

$$E_a(x^*) = |x - x^*| = 0.0004 < 0.5 \times 10^{-3},$$

$$E_a(x^*) = \frac{|x - x^*|}{|x^*|} \approx 3.23 \times 10^{-6} < 5 \times 10^{-6}.$$

Exemplo 2.4.4 *Considere os valores exactos $x = 1.1$ e $y = 100.1$ Indique os erros que se cometem ao considerar $x^* = 1$ e $y^* = 100$.*

Resolução 2.4.4 *Temos que*

$$E_a(x^*) = |x - x^*| = |1.1 - 1| = 0.1$$

$$E_a(y^*) = |100.1 - 100| = 0.1$$

Devemos salientar que o erro absoluto é igual em ambos os casos mas o primeiro parece ser mais grave atendendo à grandeza dos valores. Portanto, vamos calcular o erro relativo:

$$E_r(x^*) = \frac{x - x^*}{|x^*|} = \frac{|1.1 - 1|}{1.1} = \frac{0.1}{1.1} = 0.0909091$$

$$E_r(y^*) = \frac{y - y^*}{|y^*|} = \frac{|100.1 - 100|}{100.1} = 0.000999$$

Como o erro relativo de y^{ast} é menor que o erro relativo de x^*, verifica-se que y^* é uma melhor aproximação de y do que x^* é de x^*. ■

2.5 O Polinómio de Taylor

Teorema 2.5.1 (Polinómio de Taylor) *Seja $f(x)$ uma função de classe $\mathcal{C}^m(\Omega)$, onde $\Omega \subseteq \mathbb{R}$. Então, o polinómio de Taylor que aproxima f numa vizinhaça de raio δ de x_0 é dado por*

$$\begin{aligned} P_n(x) &= f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2}f''(x_0) \\ &+ \frac{(x-x_0)^3}{3!}f^{(3)}(x_0) + \cdots + \frac{(x-x_0)^n}{n!}f^{(n)}(x_0) \\ &= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + R_n(x); \end{aligned}$$

O resto é dado pela fórmula

$$R_n(x) = \frac{(x-x_0)^{(n+1)}}{(n+1)!}f^{(n+1)}(\xi) \qquad \xi \in]x, x_0[.$$

Nota 2.5.1 *O resto do polinómio de Taylor, da função f em torno do ponto x_0 pode ser apresentado na forma*

$$R_n(x_0) = f((n+1))(x-(x-x_0)\theta)\frac{(x-x_0)^{n+1}}{(n+1)!}, \tag{2.25}$$

para algum $\theta \in [0,1]$.

Exemplo 2.5.1 *Considere a função $f(x) = \sin(x)$. Determine o polinómio de Taylor de 9 grau em torno duma vizinhança do ponto $x = 0$.*

Resolução 2.5.1 *Tendo em conta o teorema 2.5.1, temos que*

$$\begin{aligned} P_9(x) &= f(0) + \frac{f'(0)}{1!}(x-0) + \frac{f''(0)}{2!}(x-0)^2 + \frac{f'''(0)}{3!}(x-0)^3 + \frac{f^{(4)}(0)}{4!}(x-0)^4 + \\ &+ \frac{f^{(5)}(0)}{5!}(x-0)^5 + \frac{f^{(6)}(0)}{6!}(x-0)^6 + \frac{f^{(7)}(0)}{7!}(x-0)^7 + \frac{f^{(8)}(0)}{8!}(x-0)^8 \\ &+ \frac{f^{(9)}(0)}{9!}(x-0)^9 \end{aligned}$$

Tendo em conta que

$f(0)=0$	
$f'(x)=\cos(x)$	$f'(0)=1$
$f''(x)=-\sin(x)$	$f''(0)=0$
$f'''(x)=-\cos(x)$	$f'''(0)=-1$
$f^{(4)}(x)=-\sin(x)$	$f^{(4)}(0)=0$
$f^{(5)}(x)=\cos(x)$	$f^{(5)}(0)=1$
$f^{(6)}(x)=-\sin(x)$	$f^{(6)}(0)=0$
$f^{(7)}(x)=-\cos(x)$	$f^{(7)}(0)=-1$
$f^{(8)}(x)=\sin(x)$	$f^{(8)}(0)=0$
$f^{(9)}(x)=\cos(x)$	$f^{(9)}(0)=1$

O polinómio de Taylor é então:

$$P_9(x)=x-\frac{x^3}{6}+\frac{x^5}{120}-\frac{x^7}{5040}+\frac{x^9}{362880} \tag{2.26}$$

Neste caso, o resto é dado pela fórmula

$$R_n(x)=\frac{-\sin(\xi)}{10!}x^{10}. \tag{2.27}$$

■

Exemplo 2.5.2 *Considere a função $f(x)=e^x$ no intervalo $[-2,2]$. Determine o grau do polinómio de Taylor de modo a que o erro absoluto devido à truncatura da série seja inferior a 5×10^{-5}.*

Resolução 2.5.2 *Facilmente se verifica que*

$$e^x=\sum_{n=0}^{\infty}\frac{x^n}{n!},\ \textit{com } x_0=0. \tag{2.28}$$

Portanto, o polinómio de Taylor de grau n com resto $R_n(x)$ é

$$\begin{aligned} e^x &= 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+R_n(x), \\ &= 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+e^{\theta x}\frac{x^{n+1}}{(n+1)!} \end{aligned}$$

Agora, é necessário majorar a parcela $R_n(x)$, que corresponde ao erro obido por truncatura. Assim,

$$\varepsilon_{trunc}=|R_n(x)|=\left|e^{\theta x}\frac{x^{n+1}}{(n+1)!}\right|\leq 8\frac{2^{n+1}}{(n+1)!},$$

pois $\theta\in[0,1]$ e $x\in[-2,2]$.

Para determinar o número de parcelas que é necessário considerar, vamos atribuir valores a n na expressão que majora o erro, isto é, na expressão

$$\varepsilon_{trunc(n)} = 8\frac{2^{n+1}}{(n+1)!}$$

Efectuando alguns cálculos, obtemos a tabela:

n	$\varepsilon_{trunc(n)}$
2	10.666667
4	2.1333333
6	0.20317460
8	0.01128747
10	0.00041045374
11	$0.000068408957 \approx 6.8 \times 10^{-6}$
12	$0.000010524455 \approx 1.1 \times 10^{-7}$
13	1.5034936×10^{-6}
14	2.0046581×10^{-7}

Portanto, podemos concluir que para $n = 12$*, temos que quando* $n = 12$*, temos que* $\varepsilon_{trunc(12)} \leq 10^{-5}$ *e devemos utilizar um polinómio de grau* 12*.*

■

2.5.1 O método de Horner

Em análise numérica, o método de Horner (também conhecido como *algoritmo de Horner*), em homenagem a *William George Horner*, é um algoritmo eficiente para a avaliação dos polinómios na forma monomial. O método de Horner descreve um processo manual, através da qual pode-se aproximar as raízes de uma equação polinomial. O esquema de Horner também pode ser visto como um algoritmo rápido para dividir um polinómio por um polinómio linear com a regra de Ruffini.

Dado o polinómio

$$p(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n,$$

em que $a_0, \ldots, a_n$ são números reais, e suponha que queremos estimar o polinómio em um específico valor de x, digamos x_0.

Para conseguir isso, vamos definir uma nova seqüência de constantes da seguinte forma:

$$b_n := a_n \tag{2.29}$$
$$b_{n-1} := a_{n-1} + b_n x_0 \tag{2.30}$$
$$\vdots \tag{2.31}$$
$$b_0 := a_0 + b_1 x_0. \tag{2.32}$$

Então, b_0 é o valor de $p(x_0)$.

Para ilustrar o método de Horner, devemos ter em conta que o polinómio pode ser escrito na forma

$$p\left(x\right) = a_0 + x(a_1 + x(a_2 + \cdots + x(a_{n-1} + a_n x)\cdots)).$$

Assim, substituindo iterativamente b_i na expressão,

$$\begin{aligned} p(x_0) &= a_0 + x_0(a_1 + x_0(a_2 + \cdots x_0(a_{n-1} + b_n x_0)\cdots)) && (2.33)\\ &= a_0 + x_0(a_1 + x_0(a_2 + \cdots x_0(b_{n-1})\ldots)) && (2.34)\\ &\vdots && (2.35)\\ &= a_0 + x_0(b_1) && (2.36)\\ &= b_0. && (2.37) \end{aligned}$$

O esquema de Horner é muitas vezes utilizado para converter entre diferentes sistemas numerais posicionais - caso em que x é a base do sistema de números, e os coeficientes ai são os dígitos da base de representação x de um dado número - e também pode ser usado se x é uma matriz, caso em que o ganho de eficiência computacional é ainda maior.

Exemplo 2.5.3 *Considere o polinómio* $p\left(x\right) = 2x^3 + 3x^2 + 4x + 5$. *Determine o valor de* $p\left(3\right)$ *utilizando o método de Horner.*

$$\begin{aligned} p\left(x\right) &= 2x^3 + 3x^2 + 4x + 5\\ &= \left(\left(\left(2x + 3\right)x + 4\right)x + 5\right) \end{aligned}$$

Logo,

$$\begin{aligned} p\left(3\right) &= \left(\left(\left(2 \times 3 + 3\right)x + 4\right)x + 5\right)\\ &= \left(\left(9 \times 3 + 4\right)x + 5\right)\\ &= \left(\left(31\right) \times 3 + 5\right)\\ &= \left(93 + 5\right)\\ &= 98. \end{aligned}$$

Nota 2.5.2 *Ao algoritmo é atribuído o nome de algoritmo de Horner devido ao trabalho de William George Horner, que o descreveu em* 1819*, o método já era conhecido por Isaac Newton em* 1669*, o matemático chinês Qin Jiushao no seu "Treatise Matemathica"em nove secções no século* 13*, e até mesmo antes pelo persa muçulmano matemático Sharaf al-Din al-Tusi no século* 12*. A primeira utilização do algoritmo de Horner foi nos nove capítulos da Arte Matemática, um trabalho chinês da dinastia Han (*202 *aC* - 220 *dC), editado por Liu Hui (fl. século* 3*)*

2.6 Diferenciação numérica

O teorema de Taylor pode ainda ser utilizado para aproximar derivadas. Nos casos, em que a expressão analítica de uma função é desconhecida, ou é complicada de derivar,

é muito útil estimar o valor da derivada conhecendo unicamente o valor de f em alguns pontos próximos do ponto considerado. Considereando $a = x_i$ e aplicando o teorema de Taylor com $x_{i+1} = x_i + h$, obtemos a relação

$$\begin{aligned} f(x_{i+1}) &= f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 + \cdots + \frac{f^{(n)}(x_i)}{n!}(x_{i+1} - x_i)^n \\ &= +\frac{f^{(n+1)}(\xi)}{(n+1)!}(x_{i+1} - x_i)^{n+1}, \ \ \xi \in (x_i, x_{i+1}) \end{aligned} \tag{2.38}$$

Uma vez que $x_{i+1} = x_i + h$, podemos escrever $x_{i+1} - x_i = h$ e a série (2.38) pode ser simplificada na forma

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \cdots + \frac{f^{(n)}(x_i)}{n!}h^n + \cdots + \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}, \tag{2.39}$$

Truncando a série em $n = 1$, obtemos

$$f(x_{i+1}) = f(x_i) + f(x_i)h + \frac{f''(\xi)}{2!}h^2, \ \ \xi \in]x_i\ , x_{i+1}[\,, \tag{2.40}$$

isto é, a primeira derivada de f em x_i pode ser estimada através da fórmula

$$f'(x_i) = \frac{1}{h}(f(x_{i+1}) - f(x_i)) - \frac{f''(\xi)}{2!}h. \tag{2.41}$$

Em geral, não conhecemos o valor exacto da derivada pois o valor de ξ não é conhecido. Sabendo o valor de $f(x_{i+1})$ e $f(x_i)$ podemos aproximar a derivada em x_i e ignorando o último termo. A parcela,

$$\frac{f''(\xi)}{2!}h$$

é designada por **erro de truncatura**. Habitualmente, dizemos que o erro é de ordem de h, $\mathcal{O}(h)$, pois $\frac{f''(\xi)}{2!}$ são constantes e só h varia. Quando x_{i+1} se aproxima de x_i, o passo h diminui e o erro de truncatura tende para zero.

A série de Taylor também pode ser utilizada para aproximar o valor da derivada da função para pontos anteriores a x_i, por exemplo, para $x_{i-1} = x_i - h$.

Temos então,

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(\xi)}{2!}h^2, \ \ \xi \in]x_{i-1}\ , x_i\ [\,,$$

donde resulta a relação

$$f'(x_i) = \frac{1}{h}(f(x_i) - f(x_{i-1})) + \frac{f''(\xi)}{2!}$$

Portanto, uma aproximação para o valor da derivada de primeira ordem utilizando pontos anteriores a x_i é

$$f'(x_i) \approx \frac{1}{h}(f(x_i) - f(x_{i-1})) \tag{2.42}$$

Podemos ainda utilizar o polinómio de Taylor para obter uma aproximação para a derivada de primeira ordem utilizando um ponto anterior a x_i, o ponto x_{i-1} e um ponto porterior, x_{i+1}.

Calculando o polinómio de Taylor, em x_{i+1} e em x_{i-1}, obtemos respectivamente,

$$f(x_{i+1}) = f(x_i) + f'(x_i)\,h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(\xi_1)}{3!}h^3, \quad \xi_1 \in]x_i\,, x_{i+1}\,[\,, \tag{2.43}$$

$$f(x_{i-1}) = f(x_i) - f'(x_i)\,h + \frac{f''(x_i)}{2!}h^2 - \frac{f'''(\xi_2)}{3!}h^3, \quad \xi_2 \in]x_{i-1}\,, x_i\,[\,, \tag{2.44}$$

Subtraindo (2.44) a (2.43) obtemos

$$f(x_{i+1}) - f(x_{i-1}) = 2f'(x_i)\,h + \frac{h^2}{6}f'''(\xi)$$

isto é,

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - \frac{h^2}{6}f'''(\xi)\,.$$

Portanto,

$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_{i-1})}{2h}. \tag{2.45}$$

Neste caso, o erro de truncatura é da ordem de h^2 o que indica que o erro tende mais rapidamente para zero quando x_{i+1} tende para x_i. Do mesmo modo, podemos deduzir a seguinte aproximação

$$f'(x_i) = \frac{1}{2h}\left[-3f(x_i) + 4f(x_{i+1}) - f(x_{i+2})\right] + \frac{h^2}{3}f'''(\xi)\,, \tag{2.46}$$

ou

$$f'(x) = \frac{1}{2h}\left[f(x_{i-2}) + 4f(x_{i-1}) + 3f(x_i)\right] + \frac{h^2}{3}f'''(\xi)\,. \tag{2.47}$$

Para a segunda derivada, podemos obter as seguintes aproximações

$$f''(x_i) = \frac{1}{h^2}\left[f(x_{i-1}) - 2f(x_i) + f(x_{i+1})\right] - \frac{h^2}{12}f^{(4)}(\xi)$$

Exemplo 2.6.1 *Uma partícula move-se com sobre o eixo das abissas, tendo-se registado as seguintes posições para diversos valores do tempo t:*

t	0.0	0.2	0.4	0.6	0.8	1.0
$x(t)$	0.0	0.1987	0.3894	0.5646	0.7174	0.8415

Determine a velocidade e a aceleração da partícula no instante $t = 0.4$.

Da Física sabemos que a velocidade é a derivada do deslocamento e a aceleração corresponde à segunda derivada do deslocamento.

Um vez que temos dois pontos "anteriores"a $t = 0.4$ e dois pontos "posteriores", podemos utilizar as fórmulas com 3 pontos, mais especificamente a centrada, pois é a que dá a melhor aproximação.

Consideremos $x_i = 0.2i$ para $i = 0, 1, \ldots, 5$. Então, queremos calcular as derivadas de primeira e de segunda ordem em $x_i = 0.2$. Vamos utilizar os pontos $x_1 = 0.1$ e $x_3 = 0.5$.

Defina-se $h = 0.4 - 0.2 = 0.2$. Devemos salientar que os pontos estão igualmente espaçados; isto é, $h = x_{i+1} - x_i$ para $0 \leq i \leq 4$. Neste caso, dizemos que a malha é uniforme.

Então, podemos escrever

$$\begin{aligned} \dot{x}(0.4) &= \frac{1}{2 \times 0.2}\left[-x(0.2) + x(0.6)\right] \\ &= \frac{1}{0.4}\left(-0.1987 + 0.5646\right) = 0.91475 \end{aligned}$$

$$\begin{aligned} \ddot{x}(0.4) &= = \frac{1}{0.2^2}\left(x(0.2) - 2x(0.4) + x(0.6)\right) \\ &= \frac{1}{0.04}\left(0.1987 - 2 \times 0.3894 + 0.5646\right) = -0.3875 \end{aligned}$$

2.7 Propagação de erros

Consideremos uma função $f(x)$ real de variável real e, para esta função pretendemos calcular o seu valor não no valor exacto x mas num valor aproximado x^*. Nesta secção vamos ver como é possível estimar o erro que vamos obter, isto é, vamos ver como estimar o erro absoluto,

$$E_a\left(f\left(x^*\right)\right) = \left|f(x) - f\left(x^*\right)\right|.$$

Obviamente, não conhecendo o valor de x, não podemos determinar o valor de $f(x)$. No entanto, este problema tem uma solução, que passa pela utilização do Polinómio de Taylor. Utilizando o Polinómio de Taylor, podemos escrever

$$\begin{aligned} f(x) &= f\left(x^*\right) + f'(x)\left(x - x^*\right) + \frac{f''(\xi)}{2}\left(x - x^*\right)^2 \\ f(x) - f\left(x^*\right) &= f'(x)\left(x - x^*\right) + \frac{f''(\xi)}{2}\left(x - x^*\right)^2 \end{aligned}$$

Tomando módulos e desprezando o último termo, obtemos a relação

$$\left|f(x) - f\left(x^*\right)\right| \leq \left|f'(x)\right|\left|x - x^*\right| \equiv \left|f'(x)\right|\Delta x,$$

isto é,

$$E_a\left(f\left(x^*\right)\right) \leq \left|f'\left(x^*\right)\right| E_a\left(x^*\right). \tag{2.48}$$

Isto é, o erro final depende da derivada da função (se o valor da derivada é muito elevado, o erro aumenta; caso contrário, díminui).

O caso em que a função depende de várias variáveis é semelhante. Consideremos uma uma função de n-variáveis $f(x_1, x_2, \ldots, x_n)$ e que pretendemos calcular no ponto (aproximado) $(x_1, x_2, \ldots, x_n)$ de $\mathbb{R}^n$. Se os erros absolutos de cada uma das variáveis são limiatdos

por Δx_i, para $1 \leq i \leq n$, o limite superior para o erro absoluto de $f(x_1^*, x_2^*, \ldots, x_n^*)$ pode ser obtido por

$$\begin{aligned} E_a\left(f\left(x_1^*, x_2^*, \ldots, x_n^*\right)\right) &= \left|f\left(x_1, x_2, \ldots, x_n\right) - f\left(x_1^*, x_2^*, \ldots, x_n^*\right)\right| \\ &\leq \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\left(x_1^*, x_2^*, \ldots, x_n^*\right)\right| \Delta x_i. \end{aligned} \tag{2.49}$$

Em relação ao erro relativo, podemos, mais uma vez utilizar o polinómio de Taylor

$$f(x) \approx f(x^*)(x - x^*)$$

e portanto,

$$\begin{aligned} \frac{f(x) - f(x^*)}{f(x)} &\approx \frac{f'(x)(x - x^*)}{f(x)} \times \frac{x - x^*}{x} \\ &= \approx \frac{f'(x^*)\, x^*}{f(x^*)} \times \frac{x - x^*}{x}, \end{aligned} \tag{2.50}$$

o que tomando módulos leva à desigualdade

$$E_r\left(f\left(x^*\right)\right) \leq \left|\frac{f'(x^*)\, x^*}{f(x^*)}\right| E_r\left(x^*\right). \tag{2.51}$$

Estamos pois em condições de apresentar a seguinte

Definição 2.7.1 (Número de condição) *O valor* $\left|\frac{f'(x^*)\, x^*}{f(x^*)}\right|$ *designado por "número de condição"de* f *em* x^**. Este número indica a medida na qual o erro de* x^* *se "propaga"por* $f(x^*)$*. Isto é, quando o número de condicição é maior que* 1*, cresce o erro de* x^**. Se* $condf(x^*) < 1$ *então, o erro de* x^* *vai ser atenuado no cálculo de* $f(x^*)$*.*

Este valor pode ser utilizado para avaliar a perda ou ganho de algarismos significativos aquando do cálculo dos valores de uma função, pois este número caracteriza a ampliação ou redução do erro relativo. Quando o número de condição for reduzido, a função diz-se **bem condicionada**. Quando este número for elevado, a função diz-se **mal condicionada** e o erro relativo aumenta.

Exemplo 2.7.1 *Considere a função* $f(x) = \ln(x)$*, a aproximação* $x^* = 1.01$ *do valor exacto* $x = 1.009$*, com um erro relativo* $(E_a(x^*))$ *inferior a* 0.1% *Determine o erro cometido ao calcular* $f(x)$ *por* $f(x^*)$*. Comente o resultado obtido.*

Resolução 2.7.1 *Pretendemos calcular o erro obtido ao calcular* $f(1.009)$ *em vez do valor real* $f(1.01)$ *com um erro relativo inferior a* 0.1%*. Tendo em conta os resultados apresentados anteriormente, podemos escrever*

$$\begin{aligned} E_r\left(f\left(x^*\right)\right) &< \left|\frac{1}{f\left(x^*\right)} \times \left[x^* \frac{\partial f}{\partial x^*}\left(x^*\right)\right]\right| E_r\left(x^*\right) \\ &= \left|\frac{1}{\ln\left(1.01\right)} \times \left[1.01 \times \frac{1}{1.01}\right]\right| \times 0.001 \\ &= 100.4992 \times 0.0001 \\ &= 0.1 = 10\%. \end{aligned}$$

Portanto, um erro relativo de 0.1% *deu origem a um erro final de* 10%*, isto é* 100 *vezes superior. Neste caso, o número de condição é dado por*

$$cond\left(\ln\left(x\right), 1.009\right) = \left|\frac{1.009 \times \frac{1}{1.009}}{\ln\left(1.009\right)}\right| = 111.6 \tag{2.52}$$

Portanto, $cond\left(\ln\left(x\right), 1.009\right) = 111.6$ *o que implica que o cálculo do valor da função é mal condicionada nesse ponto.*

■

No caso de f ter uma função de várias variáveis e o erro relativo de cada variável x_i for majorado por δx_i o majorante do erro relativo de $f\left(x_1^*, x_2^*, \ldots, x_n^*\right)$ é dádo pela fórmula

$$E_r\left(f\left(x_1^*, x_2^*, \ldots, x_n^*\right)\right) \leq \sum_{i=1}^{n} \left|\frac{1}{f\left(x_1^*, x_2^*, \ldots, x_n^*\right)} \times \left[x_i^* \frac{\partial f}{\partial x_i}\right]\left(x_1^*, x_2^*, \ldots, x_n^*\right)\right| \delta x_i \tag{2.53}$$

Exemplo 2.7.2 *Seja* $f\left(x, y\right) = x + y$*. Se* $E_a\left(x^*\right) \leq \Delta x$ *e* $E_a\left(y^*\right) \leq \Delta y$*, então*

$$\left|f\left(x, y\right) - f\left(x^*, y^*\right)\right| \leq \left|\frac{\partial f}{\partial x^*}\left(x^* + y^*\right)\right| \Delta x + \left|\frac{\partial f}{\partial y^*}\left(x^* + y^*\right)\right| \Delta y = \Delta x + \Delta y,$$

ou seja, o erro absoluto da soma é menor que a soma dos erros absolutos das parcelas.

Suponhamos agora que $f\left(x, y\right) = xy$*,*

$$E_r\left(f\left(x^*, y^*\right)\right) \leq \left|\frac{1}{f\left(x^*, y^*\right)} \times \left[x^* \frac{\partial f}{\partial x}\right]\left(x^*, y^*\right)\right| E_r\left(x^*\right) + \left|\frac{1}{f\left(x^*, y^*\right)} \times \left[y^* \frac{\partial f}{\partial y}\right]\left(x^*, y^*\right)\right| E_r\left(y^*\right)$$

Ao resolver um sistema

$$AX = B,$$

podem surgir problemas de condicionamento e de estabilidade numérica. Os problemas de estabilidade numérica estão relacionados com o algoritmo que utilizamos para resolver o sistema. Por exemplo, para evitar os problemas de instabilidade numérica, é habitual considerar o método de eliminação de Gauss com pesquisa de pivot. No entanto, se o problema for mal condicionado, essas técnicas de pesquisa de pivot deixam de ser úteis, já que um problema mal condicionado será sempre numericamente instável. Interessa-nos, portanto, identificar quais os sistemas que nos podem trazer problemas de condicionamento.

Supondo que nos era dado, não o vector B exacto, mas apenas uma aproximação $\tilde{B}$, vamos analisar a influência desse erro nos resultados obtidos, já que em vez do valor exacto, obtemos um valor aproximado $\tilde{x}$, solução do sistema:

$$A\tilde{X} = \tilde{B}.$$

Definição 2.7.2 *Designa-se por número de condição de uma matriz A relativamente à norma $\|\cdot\|$, ao valor :*

$$\kappa = cond(A) = \|A\| \left\|A^{-1}\right\|. \tag{2.54}$$

Um problema diz-se *bem condicionado* se pequenos erros nos dados iniciais originam pequenas alterações na solução do problema. Mas, quando pequenos erros produzem grandes alterações na solução do problema, o problema diz-se *mal condicionado.*

Exemplo 2.7.3 *Consideremos a equação de segundo grau*

$$x^2 - 2.029x + 1.0285 = 0,$$

cujas soluções são

$$x_1 = 0.9878494841, \ e \ x_2 = 1.041150516$$

Arredondando o termo independente para 1.029*, introduzimos um erro relativo inicial de aproximadamente de* 0.05%*, as soluções passam a ser:*

$$x_1^* = 1 \ e \ x_2 = 1.029$$

que sofreram um erro relativo de aproximadamente 1.2% *o que é muito superior ao inicial. Neste caso, o problema é mal condicionado.*

2.8 Cancelamento subtractivo

Uma outra situação que pode ocorrer é a perda de algarismos significativos aquando da subtração de números muito próximos. Este tiop de erros é designado por *cancelamento subtractivo.*

Exemplo 2.8.1 *Calcule a diferença*

$$\sqrt{9876} - \sqrt{9875}.$$

Utilizando uma máquina de calcular com uma mantissa com 10 *dígitos, temos*

$$\sqrt{9876} - \sqrt{9875} = 0.9937806599 \times 10^2 - 0.9937303457 \times 10^2 = 0.502142 \times 10^{-2}$$

O problema reside no facto de que os dados do problema foram fornecidos com 10 *algarismos significativos e o resultado aparece com* 6 *algarismos sendo apenas* 5 *deles significativos.*

Para resolver este problema é tentar substituir a substração por outra operação. Neste caso, multiplicamos e dividimos pelo "conjugado"de $\sqrt{9876}-\sqrt{9875}$*, isto é, por* $\sqrt{9876}+\sqrt{9875}$*. Assim, obtemos*

$$\begin{aligned}\sqrt{9876}-\sqrt{9875} &= \left(\sqrt{9876}-\sqrt{9875}\right)\times\frac{\sqrt{9876}+\sqrt{9875}}{\sqrt{9876}+\sqrt{9875}}\\ &= \frac{1}{\sqrt{9876}+\sqrt{9875}}=\frac{1}{0.9937806599\times10^{2}+0.9937303457\times10^{-2}}\\ &= 0.5031403493\times10^{-2},\end{aligned}$$

que é muito mais exacto.

Em resumo: Seja x^* a aproximação do valor exacto x então

$$E_a(x^*)=|x-x^*| \text{ e } E_r(x^*)=\frac{|x-x^*|}{|x|}$$

$$E_a(f(x^*))\leq\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}(x_1^*,\ldots,x_n^*)\right||x_i-x_i^*|,$$

$$E_r(f(x^*))\leq\sum_{i=1}^{n}\left|\frac{x_i^*\frac{\partial f}{\partial x_i}(x_1^*,\ldots,x_n^*)}{f(x_1^*,\ldots,x_n^*)}\right|\frac{|x_i-x_i^*|}{|x_i|}.$$

Ao determinar a solução da equação $f(x)=0$ usando um método numérico é importante recordar que em geral a função não é representada exactamente no computador/máquina de calcular. Estamos pois interessados em conhecer o efeito das perturbações de f nos seus zeros.

Seja $\tilde{f}(x)=f(x)+\delta(x)$ onde $|\delta(x)|\leq\Delta$. Seja x^* a raíz de $f(x)=0$ e $\tilde{x}^*$ a "correspondente"raíz de $f(x)=0$. Têm-se então que $|f(\tilde{x}^*)|\leq\Delta$.

Se f' não varia muito numa vizinhança de x^* então

$$|\tilde{x}^*-x^*|\approx\frac{\Delta}{|f(\tilde{x}^*)|},$$

se x^* for uma raíz simples.

Assim, se $|f(\tilde{x}^*)|$ é pequeno então $|\tilde{x}^*-x^*|$ é grande.

Podemos afirmar que o número de condicionamento ou número de condição de um problema é uma medida indicando se o problema tem "boas condições"para ser tratado numericamente. Um problema com um número de condição pequeno é chamado de bem condicionado, enquanto os problemas que possuem um número de condição elevado são denominados mal condicionados.

2.9 Exercícios

1. Converta os seguintes números para a sua forma decimal:

a) $a = (101101)_2$,
b) $b = (110101011)_2$,
c) $c = (0.1101)_2$;
d) $d = (0.101010\ldots)_2$
e) $e = (6)_8$;
f) $f = (232)_8$;
g) $g = (15.04)_8$;
h) $h = (2.32)_8$;
i) $k = (9F6)_{16}$;
j) $l = (E5B)_{16}$

2. Converta os seguintes números décimais para binário:

   (a) $m = 23$; (b) $n = 0.81125$ (c) $p = 22.75$

3. Obtenha a representação na base octal de

   (a) $q = 2003$; (b) $r = 0.390625$; (c) $p = 31.375$.

4. Obtenha a representação em base hexadecimal de

   (a) $t = 2003$; (b) $u = 1000$; (c) $v = 0.1$.

5. Represente os valores $a = 0.01234$ e $b = 345.6789$ no sistema $FP(10, 4, 2)$.

6. Quantos números reais de $]0.111 ; 2.30[$. pertencem ao sistema $FP(10, 3, 2)$?

7. Escreva o número $x = 123456.78$ nos sistemas $FP(10, 4, 2)$ e $FP(2, 6, 3)$.

8. Qual é o valor máximo e o valor mínimo que se pode representar no sistema $FP(3, 4, 2)$? Qual a sua representação decimal?

9. Qual é o menor número positivo que se pode representar no sistema $FP(3, 4, 2)$?

10. Os números $x_1 = 0.52 \times 10^{-4}$, $x_2 = 0.61 \times 10^5$, $x_3 = -0.37 \times 10^{-3}$ e $x_4 = -0.61 \times 10^5$ vão ser adicionados em aritmética de ponto flutuante $FP(10, 2, 1)$. Diga, explicando convenientemente qual dos seguintes resultados tem menor erro:

    a) $s_1 = (x_1 + x_2) \, p \, (x_3 + x_4)$;
    b) $s_2 = (x_1 + x_3) \, p \, (x_2 + x_4)$;

11. No sistema $FP(10, 4, 2)$ qual é o número imediatamente a seguir ao 1?

12. Calcule o valor do polinómio

$$P(x) = x^3 - 5x^2 + 6x + 0.55$$

    em $x = 2.73$, utilizando a aritmética de ponto flutuante $FP(10, 3, 2)$. Compare com o valor exacto.

13. Repita o exercício anterior mas reescrevendo o polinómio na forma

$$P(x) = ((x - 5) + 6)\, x + 0.55$$

14. Considere a função $f(x) = \sin(x)$.

a) Determine a aproximação de $f\left(\frac{\pi}{8}\right)$ e de $f\left(\frac{10\pi}{9}\right)$, obtida através do polinómio de Taylor de f de grau 4, em torno de 0.

b) Determine um majorante para o erro absoluto das aproximações, calculadas na alínea anterior, usnado o termo do erro na fórmula de Taylor e compare-o com o valor exacto do erro.

c) Determine um majorante para o erro

$$|f(x) - P_4(x)|, \ x \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right].$$

d) Pretende-se aproximar a função seno pelo seu polinómio de Taylor de grau n em torno de zero. Qual é o valor mínimo de n de modo a que o polinómio aproxime a função com um erro menor que 10^{-6} no intervalo $\left[0, \frac{\pi}{2}\right]$?

e) Determine um valor aproximado de $\sin\left(\frac{\pi}{8}\right)$, com 9 casa decimais correctas utilizando o polinómio de Taylor em torno de 0.

15. O polinómio $P_2(x) = 1 - 0.5x^2$ é utilizado para aproximar a função $\cos(x)$ no intervalo $(-0.5; 0.5)$. Determine um majorante para o erro máximo de qualquer aproximação.

16. Determine a derivada de primeira ordem da função $g(x) = e^{\sin(x)}$ no ponto $x = 0.5$ utilizando diferenças finitas (progressivas, regressivas e centrais) de primeira ordem com $h = 0.01$. De entre as fórmulas utilizadas, qual é a que permite uma melhor estimativa?

17. Utilize as fórmulas de diferenças progressivas e regressivas para determinar as aproximações que completem a tabela:

x	$f(x)$	$f'(x)$
0.0	-1	$\cdots$
0.2	-0.2839867	$\cdots$
0.4	0.2484244	$\cdots$

Sabendo que $f(x) = x\cos(x) - 2x^2 + 3x - 1$, determine o erro efectivamente cometido em cada aproximação e compare-o com o majorante teórico.

18. A distância percorrida em metros por um foguete em cada segundo apresenta os seguintes valores:

t	0	1	2	3	4	5
y	0.0	2.5	7.8	18.2	51.9	80.3

Utilize a diferenciação numérica para aproximar a velocidade e a aceleração em cada valor de tempo.

19. Um rectângulo mede $19\text{cm} \pm 0.2$ de largura e $31\text{cm} \pm 0.5$ de comprimento.

a) Determine a área deste rectêngulo indicando os intervalos possíveis.

b) Determine uma estimativa do valor da área deste rectângulo e calcule um estimativa do limite superior do erro absoluto da área do rectângulo.

20. Suponha que $x^* = 2.475$ tem 3 casas decimais correctas em relação ao valor exacto x.

21. Determine uma estimativa para o limite superior do erro relativo.

22. Dada a função $f(x) = \sqrt{x-2}$, determine uma estimativa para o limite superior do erro relativo de $f(x^*)$.

23. Sejam $a = \pi$, $\bar{a} = 3.14$, $b = \frac{\pi}{2}$ e $\bar{b} = 1.57$.

a) Determine os erros relativos das aproximações $\sin(3.14)$ e $\sin(1.57)$, dos valores $\sin(a)$ e $\sin(b)$ respectivamente.

b) Qual dos números de condicionamento $\text{cond}(\sin, 3.14)$ e $\text{cond}(\sin, 1.57)$ será maior?

c) Pretende-se calcular valores para a expressão

$$\frac{1-\sqrt{1+x^2}}{x},$$

para algum x na vizinhança de zero. Reescreva a expressão de modo a evitar a perda de dígitos significativos devido ao cancelamento subtractivo.

d) Considere a função real de variá

Capítulo 3

Solução de equações

3.1 Algumas definições e resultados importantes

Definição 3.1.1 (Raíz simples de uma equação) *Seja $y = f(x)$ uma função real de variável real. Dizemos que $\alpha \in \mathbb{R}$ é uma raíz da equação $f(x) = 0$ ou que é um zero da equação $f(x) = 0$ se e somente se*

$$f(\alpha) = 0. \tag{3.1}$$

Definição 3.1.2 (Raízes múltiplas) *Diz-se que x^* é uma raíz com multiplicidade m da equação $f(x) = 0$ se*

$$f(x) = (x - x^*)^m g(x), \tag{3.2}$$

sendo $g(x)$ uma função contínua numa vizinhança de x^* e $g(x^*) \neq 0$. Se $m = 1$ a raíz diz-se simples.

Teorema 3.1.1 *Seja $f \in C^m([x^* - \delta, x^* + \delta])$ para $\delta > 0$. Então x^* é raíz de multiplicidade m se e só se*

$$\begin{cases} f(x^*) = 0 \\ f'(x^*) = f''(x^*) = \cdots = f^{(m-1)}(x^*) = 0 \\ f^{(m)}(x^*) \neq 0 \end{cases} \tag{3.3}$$

Para determinar uma solução ou a solução de uma equação $f(x) = 0$ vamos utilizar métodos iterativos. Um processo (ou método) iterativo é um método pelo qual se pode obter uma sucessão $x_1, x_2, \ldots, x_n, \ldots$ de *valores aproximados* ou *iteradas* da solução procurada. A solução das iteradas pode representar-se por

$$\{x_n \ : \ n \geq 0\}. \tag{3.4}$$

Devemos salientar que na prática, cada nova iterada x_{n+1} é calculada recorrendo ao conhecimento de uma ou mais iteradas x_i determinadas nos passos anteriores. Seja agora α uma raíz da equação $f(x) = 0$. Chama-se erro e_n da iterada x_n de um processo iterativo a

$$e_n = \alpha - x_n. \tag{3.5}$$

Um processo iterativo tem interesse prático unicamente quando as iteradas se aproximam da raíz procurada, isto é, a sucessão (3.4) é convergente. Um processo iterativo diz-se convergente quando

$$\lim_{n\to\infty} x_n = \alpha \tag{3.6}$$

ou, de forma equivalente quando,

$$\lim_{n\to\infty} e_n = 0. \tag{3.7}$$

Interessa ainda caracterizar a "velocidade"de convergência da sucessão para a raíz α.

Definição 3.1.3 *Suponha-se que a sucessão de iteradas* $\{x_n : n \geq 0\}$ *é convergente para* α. *Diz-se que o processo iterativo converge com ordem* $p\,(p > 1)$ *para o ponto* α *se*

$$\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|^p} = \kappa_\infty, \tag{3.8}$$

para algum $\kappa_\infty > 0$. *Se* $p = 1$, *a velocidade de convergência diz-se linear e se* $p > 1$ *supralinear. Se* $p = 2$, *a velocidade de convergência diz-se quadrática, etc . . . A constante* κ_∞ *é designada por coeficiente assintótico de convergência.*

A determinação iterativa de zeros de funções passa pela resolução de vários problemas:

1) Determinação dos intervalos com a menor amplitude possível nos quais existe um e um só zero de $f\,(x)$;

2) O cálculo aproximado da raíz utilizando um processo interativo (convergente!). Para tal é necessário escolher de forma adequada os dados iniciais;

3) Determinação do erro cometido aquando da aplicação do critério de paragem.

Um problema matemático cuja solução é muito sensível a variações nos dados e parâmetros diz-se *mal condicionado*. Um problema diz-se *bem condicionado* se pequenas variações nos dados e parâmetros induzem sempre pequenas variações na solução.

Um método numérico diz-se *instável* ou que apresenta instabilidade induzida se a acumulação dos erros durante o cálculo pode ter grande influência no resultado final. Um método estável produz sempre bons resultados (com problemas bem condicionados).

3.2 Critérios de paragem

Os métodos que de seguida são apresentados para o cálculo aproximado de soluções de equações não lineares são métodos iterativos. Assim, é necessário estabelecer critérios de paragem, isto é, os critérios que são utilizados para indicar quando devemos deixar de aplicar o processo. Os critérios habitualmente utilizados são:

1. Número de iterações;

2. Juntamente, associado ao critério anterior é habitual impor a condição

$$|e_n| \leq \varepsilon,$$

3. Diferença entre duas iterações consecutivas é inferior a um determinado valor, isto é, o processo iteractivo pára quando

$$|x_{n+1} - x_n| \leq \varepsilon,$$

4. Podemos também utilizar um critério de paragem baseado na estimativa do erro relativo, isto é,

$$\frac{|x_{n+1} - x_n|}{|x_{n+1}|} \leq \delta,$$

5. Em certos casos, (quando $f'(x) \geq 1$, numa vizinhaça apropriada da raíz α) também se utiliza a condição

$$|f(x_n)| \leq \varepsilon \ll 1,$$

onde ε representa a tolerância apropriada.

6. A aproximação para o erro absoluto é inferior a um valor especificado.

Devemos salientar que não é conveniente utilizar unicamente um dos critérios pois, dependendo da função, podemos obter valores que satisfazem os critérios de paragem mas que não são boas aproximações para a raíz.

A condição 5 é um corolário do teorema de Lagrange. Seja f contínua no intervalo $\mathcal{I} = [a, b]$ e diferenciável em $]a, b[$. Seja α um zero de $f(x)$ em $]a, b[$. Então, do teorema de Lagrange

$$\begin{aligned} f(x_n) - f(\alpha) &= f'(c)(x_n - \alpha) \Rightarrow \\ |x_n - \alpha| &= \left| \frac{f(x_n)}{f(c)} \right| \end{aligned}$$

para algum c entre x_n e α (e portanto em $\mathcal{I}$). Portanto,

$$|e_n| = |x_n - \alpha| \leq \frac{|f(x_n)|}{\min\limits_{x \in \mathcal{I}} |f'(x)|}, \tag{3.9}$$

desde que a derivada de f não se anule em $\mathcal{I}$.

3.3 Localização das raízes

Definição 3.3.1 *Um zero ou raíz de uma função f ou raíz da equação $f(x) = 0$ a qualquer número c tal que $f(c) = 0$.*

Teorema 3.3.1 (Bolzano) *Seja $f(x)$ uma função contínua em $[a, b]$, se $f(a)f(b) < 0$ então existe (pelo menos um) $c \in [a, b]$ tal que $f(c) = 0$.*

Teorema 3.3.2 (Rolle) *Seja $f(x)$ uma função contínua e derivável em $[a, b]$. Se $f(a) = f(b)$ então existe $c \in [a, b]$ tal que $f'(c) = 0$.*

Corolário 3.3.1 *Entre dois zeros de uma função diferenciável num intervalo há pelo menos um zero da sua derivada.*

O corolário 3.3.1 é um resultado muito importante na determinação dos zeros de um função diferenciável. Assim, se f' tem um zero, a função não pode ter mais do que 2 zeros pois se assim não fosse, se a função tivesse 3 zeros, teria que ter pelo menos dois zeros da derivada.

Corolário 3.3.2 *Entre dois zeros consecutivos da derivada de uma função diferenciável num intervalo, não pode haver mais do que um zero dessa função.*

Conjugando estes resultados com o teorema de Bolzano, sobre funções contínuas, podemos afirmar:

- Se a função assumir valores de sinais contrários entre os zeros consecutivos da derivada, entre esses dois zeros existe um zero da função;
- Se o sinal for o mesmo, então, não há zero algum da função entre os dois zeros da derivada.

Um resultado que nos permite majorar o erro entre a solução exacta x e a sua aproximação x^* é o

Teorema 3.3.3 *Seja f uma função continuamente diferenciável no intervalo $[a; b]$ tal que $m_1 = \min_{\xi \in [a,b]} |f'(\xi)| > 0$. Seja $x \in [a, b]$ tal que $f(x) = 0$. Então,*

$$|x - x^*| \leq \frac{|f(x^*)|}{m_1}, \quad \forall x^* \in [a, b]. \tag{3.10}$$

Também é conveniente impor/verificar a monotonia das funções em estudo. Se uma função (contínua e diferenciável) é estritamente crescente/decrescente num intervalo então nesse intervalo ela admite um só zero (nas condições do teorema 3.3.1).

Teorema 3.3.4 (Valor médio para Integrais) *Seja f um função contínua no intervalo $[a, b]$ e $g(x)$ uma função integrável, que não muda de sinal em $[a, b]$. Então, existe pelo menos um valor $\xi \in (a, b)$ tal que*

$$f(\xi) \int_a^b g(x)\, dx = \int_a^b f(x)\, g(x)\, dx. \tag{3.11}$$

Teorema 3.3.5 (Weierstrass) *Seja f uma função contínua e definida no intervalo $[a,b]$. Então, para cada $\varepsilon > 0$ existe um polinómio $p(x)$ definido em $[a,b]$ tal que*

$$\max_{x\in[a,b]} |f(x) - p(x)| < \varepsilon. \tag{3.12}$$

Este teorema indica que por muito pequeno que seja o valor de ε, existe sempre um polinómio $p(x)$ na faixa

$$\mathcal{F} = \left\{(x,y) \in \mathbb{R}^2 \ : \ x \in [a,b] \wedge y \in [f(x) - \varepsilon, f(x) + \varepsilon]\right\}$$

Exemplo 3.3.1 *Determine o número de soluções reais da equação*

$$3x^4 - 2x^3 - 3x^2 + 1 = 0 \tag{3.13}$$

Resolução 3.3.1 *Consideremos a função*

$$f(x) = 3x^4 - 2x^3 - 3x^2 + 1.$$

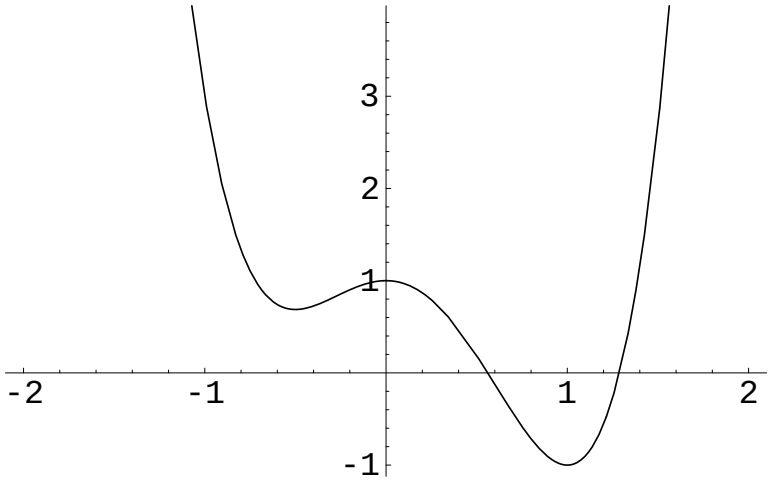

Figura 3.1: Gráfico da função $f(x) = 3x^4 - 2x^3 - 3x^2 + 1$

e o gráfico da sua derivada é
A sua derivada, $f'(x)$ é

$$f'(x) = \frac{df}{dx} = \frac{d}{dx}\left(3x^4 - 2x^3 - 3x^2 + 1\right) = 12x^3 - 6x^2 - 6x.$$

A derivada admite por zeros $-\frac{1}{2}$, 0 e 1.
Como

$$f\left(-\frac{1}{2}\right) = \frac{11}{16} > 0, \ \ f(0) = 1 > 0, \ \ f(1) = -1 < 0,$$

podemos concluir que:

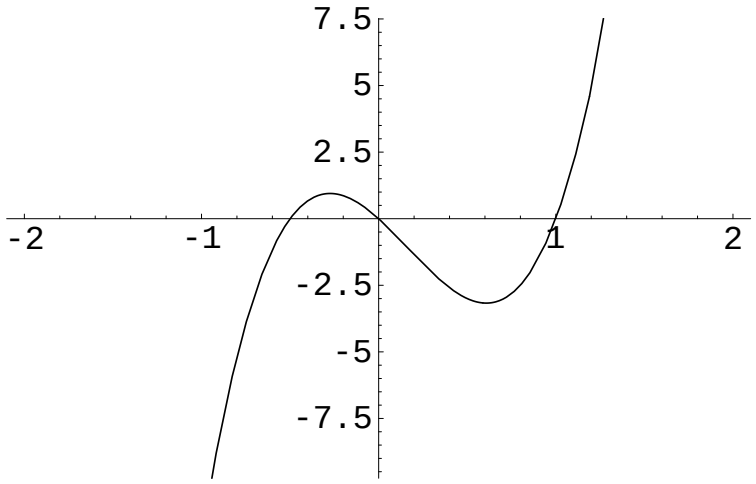

Figura 3.2: Gráfico da função $f'(x) = 12x^3 - 6x^2 - 6x$

- *em* $]0,1[$ *existe um zero da função (solução da equação);*
- *em* $]-\frac{1}{2},0[$ *não há nenhum zero da solução da equação.*

Por outro lado, atendendo a que

$$\lim_{x\to-\infty} f(x) = +\infty \text{ e } \lim_{x\to\infty} f(x) = +\infty,$$

vê-se que existe mais uma solução da equação em $]1,+\infty[$.

A equação tem, em $\mathbb{R}$*, apenas duas soluções* x_1 *e* x_2 *tais que*

$$0 < x_1 < 1 \text{ e } x_2 > 1.$$

■

3.3.1 Zeros de Polinómios

No que se segue, $p_n(x)$ representa um polinómio de grau n, com coeficientes reais definido por

$$p_n(x) = \sum_{k=0}^{n} a_k x^n = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \tag{3.14}$$

Teorema 3.3.6 (Teorema fundamental da Álgebra) *Um polinómio* $p_n(x)$ *de grau* n *tem exactamente* n *raízes reais ou complexas. Se* $z \in \mathbb{C}$ *é raíz de* $p_n(x)$ *então* $\bar{z} \in \mathbb{C}$ *também é raíz de* $p_n(x)$*.*

Teorema 3.3.7 *Todo o polinómio de grau ímpar com coeficientes reais tem pelo menos uma raiz real.*

Teorema 3.3.8 *Considere-se a sequência de sinais dos coeficientes não nulos de um polinómio $p_x(x)$. Seja N o número de mudanças de sinal na sequência. Se k for o número de zeros positivos de p então $k \leq N$ e $N-k$ é um número inteiro par não negativo.*

Exemplo 3.3.1 *Determine o número de zeros positivos e negativos do polinómio*

$$p(x) = x^3 - 2x - 5.$$

Resolução 3.3.2 *A lista de sinais dos coeficientes não nulos é $+,-,-$. O número de mudanças de sinal é $N=1$. Assim, o número de zeros positivos é $k \leq 1$. Sendo $N-k = 1-k$ um número par não negativo deverá ter-se $k=1$.* ■

Teorema 3.3.9 (Fórmulas de Girard-Newton-Viéte) *Seja $p_n(x)$ um polinómio de grau n de coeficientes reais. Sejam $x_1^*, x_2^*, \ldots, x_n^*$ os seus zeros. Então:*

1. $\sum_{i=1}^{n} x_i^* = -\frac{a_{n-1}}{a_n}$,
2. $\prod_{i=1}^{n} x_i^* = (-1)^n \frac{a_0}{a_n}$,
3. *Se* $\rho = 1 + \max_{0 \leq j \leq n-1} \left| \frac{a_j}{a_n} \right|$ *então* $|x_i^*| \leq \rho$, $(i = 1, 2, \ldots, n)$.
4. *Sejam* $R_1 = \frac{n|a_0|}{|a_1|}$, $R_2 = \left[\frac{|a_0|}{|a_1|} \right]^{\frac{1}{n}}$ *e* $R = \min\{R_1, R_2\}$. *Então, para algum i,*
$$|x_i^*| < R.$$
5. *Se* $p_n(\tilde{x}) \neq 0$ *e* $p'(\tilde{x}) \neq 0$, *então existe um zero de* x_i^* *de tal que*
$$|x_i^* - \tilde{x}| \leq n \left(\frac{|p(\tilde{x})|}{|p'(\tilde{x})|} \right).$$

3.4 Aproximação de raízes

Nesta secção vamos apresentar alguns métodos que nos permitem obter aproximações para as raízes de uma equação. Esses métodos são o método da Bissecção, de Newton-Raphson, da Tangente Fixa e o da Secante. O método da Bissecção é um método globalmente convergente no sentido em que a convergência é garantida desde que se conheça um intervalo $\mathcal{I} = [a, b]$ onde a função é contínua, monótona e $f(a) \times f(b) < 0$. Não são impostas condições sobre a amplitude do intervalo $\mathcal{I}$. Os restantes métodos são localmente convergentes pois as iterações iniciais devem estar suficientemente próximas da raíz para haver convergência. Os métodos localmente convergentes, em geral, convergem mais rapidamente que os métodos globalmente convergentes (no caso de raízes simples).

Consequentemente, métodos híbridos têm sido desenvolvidos combinando o método da bissecção com o de Newton ou o da Secante. Usando o método da bissecção calcula-se

um intervalo com um comprimento suficientemente pequeno contendo a raíz e as restantes iterações são obtidas com um método localmente convergente.

Vamos de seguida apresentar métodos que nos permitem encontrar soluções aproximadas para equações da forma $f(x) = 0$ onde $f(x)$ não é necessariamente um polinómio.

Aquando do cálculo de raízes aproximadas pode haver alguns problemas nos seguints casos:

- O critério de convergência $|x_{n+1} - x_n| \leq \varepsilon$, com $\varepsilon \ll 1$ pode falhar quando $f'(x)$ é muito elevado, sendo x o zero da equação;
- O critério de convergência $|f(x_n)| \leq \varepsilon$ pode falhar se $f'(x)$ é muito pequeno;
- Para resolver estes problemas mencionados nos dois itens anteriores, devemos utilizar os dois critérios.

3.4.1 Método da Bissecção

O método da Bissecção é baseado no teorema do valor intermédio (Teorema de Bolzano, ver teorema 3.3.1). A ideia é encontrar um intervalo (inicial) onde a função (contínua, monótona e diferenciável) tem um zero. Vamos supor que $f'(x) > 0$ ou $f'(x) < 0$, $\forall x \in [a, b]$. Depois, vamos "dividir"esse intervalo em subintervalos até encontrar um com comprimento suficientemente pequeno que contém a raíz. O critério para "encontrar"o subintervalo que contêm a raíz baseia-se na continuidade e monotonia da função e no facto de que o produto do valor da função dos extremos desse intervalo é negativo. Se f tem um zero em $\mathcal{I}_n = [a_n, b_n]$, então $f(a_n) \times f(b_n) < 0$. Calculamos

$$c_n = a_n + \frac{b_n - a_n}{2} = \frac{a_n + b_n}{2}.$$

Se $f(a_n) \times f(c_n) < 0$ o zero está em $[a_n, c_n]$. Caso contrário, o zero está em $[c_n, b_n]$.

O método pode ser aplicado seguindo os seguintes passos:

P_1 Encontrar os valores de a e b;

P_2 Fazer $c = a + \dfrac{b-a}{2} = d\frac{a+b}{2}$;

P_3 Se $f(c) = 0$ então $x = c$ é a solução (verificar se o valor satisfaz o critério de paragem);

P_4 Se $f(c) \neq 0,\ f(a) f(c) > 0$ então $a = c$. Caso contrário, $b = c$.

P_5 Repetir o procedimento desde o passo P_2.

O procedimento pode ser apresentado (numa forma condensada) por

$$f(a_k).f(b_k) < 0, \quad c_k = \frac{a_k + b_k}{2}, \quad E_a(x_k) \leq \frac{b-a}{2^{k+1}}, \quad k = 0, 1, \ldots, \tag{3.15}$$

onde $I_0 = [a, b]$ tal que $f(a) \times f(b) < 0$ e f é contínua.

Este procedimento vai gerar uma sucessão de valores c_k tais que

$$|c_k - c| \leq \frac{b-a}{2^{k+1}},\ k \geq 0. \tag{3.16}$$

Escolhendo um valor para o erro ε, o número mínimo de iterações n para garantir a precisão desejada é dada pela fórmula

$$n \geq \frac{\log_{10}(b-a) - \log_{10}\varepsilon}{\log_{10} 2}. \tag{3.17}$$

Uma estimativa para o erro relativo é

$$|E_r| = \left|\frac{x_r^n - x_r^{n-1}}{x_r^n}\right| \times 100\%. \tag{3.18}$$

Exemplo 3.4.1 *Utilizando o método da Bissecção, calcule uma aproximação para* $\sqrt{2}$. *Considere como critério de paragem* $\varepsilon = \frac{b-a}{2^{k+1}} < 5 \times 10^{-2}$.

Resolução 3.4.1 *Consideremos a função* $f(x) = x^2 - 2$ *cujo gráfico no intervalo* $\mathcal{I}[1,2]$ *é:*

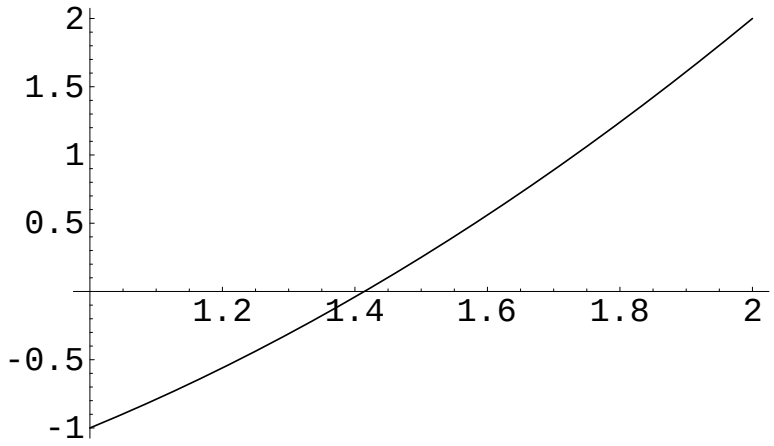

Figura 3.3: Gráfico da função $f(x) = x^2 - 2$

É uma função contínua pois é uma função polinomial. Para $x > 0$, $f'(x) > 0$ *e como* $f(1) \times f(2) < 0$, *a função admite um e um só zero em* $[1,2]$.

$$x_k = \frac{a_k + b_k}{2},\ \ \varepsilon < \frac{b-a}{2^{k+1}}.$$

k	$\mathcal{I}_k$	x_k	$f(a_k)$	$f(x_k)$	$f(c_k)$	ε
0	$\mathcal{I}_0 = [1,2]$	1.5	$f(1) < 0$	$f(1.5) > 0$	$f(2) > 0$	$\varepsilon < 0.5$
1	$\mathcal{I}_1 = [1,1.5]$	1.25	$f(1) < 0$	$f(1.25) < 0$	$f(1.5) > 0$	$\varepsilon < 0.25$
2	$\mathcal{I}_2 = [1.25,1.5]$	1.375	$f(1.25) < 0$	$f(1.375) < 0$	$f(1.5) > 0$	$\varepsilon < 0.125$
3	$\mathcal{I}_3 = [1.375,1.5]$	1.4375	$f(1.375) < 0$	$f(1.4375) > 0$	$f(1.5) > 0$	$\varepsilon < 0.0625$
4	$\mathcal{I}_4 = [1.4375,1.5]$	1.4375	$f(1.4375) < 0$	$f(1.46875) > 0$	$f(1.5) > 0$	$\varepsilon < 0.03125$
5	$\mathcal{I}_5 = [1.4375,1.46875]$	1.453125	$f(1.4375) < 0$	$f(1.453125) > 0$	$f(1.5) > 0$	$\varepsilon < 0.015625$

■

Exemplo 3.4.2 *Utilize o método da Bissecção para calcular a raíz da equação* $e^{-x} - \ln x = 0$. *Considere os seguintes critérios de paragem: número máximo de* 4 *iterações;* $|x_{k+1} - x_k| \leq 10^{-4}$ *e* $|f(x_k)| \leq 10^{-4}$.

Resolução 3.4.2 *A função* $f(x) = e^{-x} - \ln(x)$ *é uma função contínua no intervalo* $I_0 = [1, 2]$ *pois é a soma de duas funções contínuas: a função* $e^{\alpha x}$ *com* $\alpha \in \mathbb{R}$, *e a função* $\ln(x)$. *O domínio de* $e^{\alpha x}$ *é* $\mathbb{R}$ *e o domínio de* $\ln(x)$ *é* $\mathbb{R}^+$. *Portanto, a função está bem definida em* I_0. *A derivada da função é*

$$f'(x) = -e^{-x} - \frac{1}{x} = -\left(e^{-x} + \frac{1}{x}\right) \leq 0, \forall x \in \mathcal{I}_0,$$

que também está bem definida em I_0. *Mais, a função é estritamente decrescente em* I_0. *Temos ainda que*

$$f(1) \times f(2) < 0.$$

Consequentemente há um (único) zero de f *em* $\mathcal{I}_0$.

A representação gráfica de uma função é muito útil aquando da determinação da primeira aproximação. Podemos utilizar a representação gráfica para obter um intervalo onde existe uma só raíz, depois aplicando o método da bissecção obtemos uma melhor aproximação para a raíz e depois, utilizando um método localmente convergente obtemos uma aproximação razoável para a raíz.

A função tem a seguinte representação gráfica para $x \in [1, 2]$ *:*

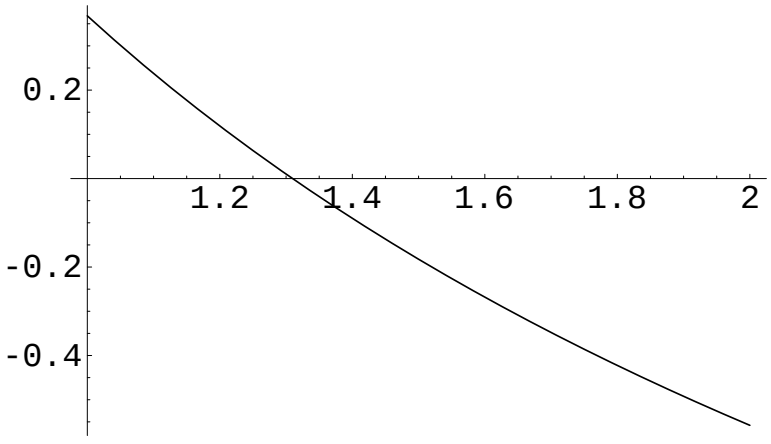

Figura 3.4: Gráfico da função $f(x) = e^{-x} - \ln(x)$

$\longrightarrow_{n \to \infty}$

Observando o gráfico, podemos verificar que no intervalo $I = [1, 2]$ *a função é contínua, monótona decrescente e tem um zero no intervalo* $I = [1.2, 1.4]$*. Logo, estes valores são à partida bons candidatos à condição inicial. No entanto, para dar ao leitor uma ideia da velocidade de convergência dos vários métodos vamos utilizar com valores iniciais* $I = [1, 2]$*,* $x_0 = 1$ *(Newton-Raphson) e* $x_0 = 1$*,* $x_1 = 2$ *(Secante).*

Método da Bissecção

$I_n = [a_n, b_n]$	$c_n = \frac{a_n + b_n}{2}$	$f(a_n)$	$f(c_n)$	$f(b_n)$	I_{n+1}
$I_0 = [1, 2]$	$c_1 = 1.5$	$f(1) > 0$	$f(1.5) < 0$	$f(2) < 0$	$I_1 = [1, 1.5]$
$I_1 = [1, 1.5]$	$c_2 = 1.25$	$f(1) > 0$	$f(1.25) > 0$	$f(1.5) < 0$	$I_2 = [1, 25, 1.5]$
$I_2 = [1.25, 1.5]$	$c_3 = 1.375$	$f(1.25) > 0$	$f(1.375) < 0$	$f(1.5) < 0$	$I_3 = [1.25, 1.375]$
$I_3 = [1.25, 1.375]$	$c_4 = 1.3125$	$f(1.25) > 0$	$f(1.3125) < 0$	$f(1.375) < 0$	$I_4 = [1.25, 1.3125]$
$I_4 = [1.25, 1.3125]$	$c_5 = 1.28125$	$f(1.25) > 0$	$f(1.28125) > 0$	$f(1.3125) < 0$	$I_5 = [1.28125, 1.3125]$
$I_5 = [1.28125, 1.3125]$	$c_6 = 1.296875$	$f(1.28125) > 0$	$f(1.296875) > 0$	$f(1.3125) < 0$	$I_6 = [1.296875, 1.3125]$
$I_6 = [1.296875, 1.3125]$	$c_7 = 1.3045875$	$f(1.296875) > 0$	$f(1.3045875) > 0$	$f(1.3125) < 0$	$I_7 = [1.3045875, 1.3125]$
$I_7 = [1.3045875, 1.3125]$	$c_8 = 1.30854375$	$f(1.3045875) > 0$	$f(1.30854375) > 0$	$f(1.3125) < 0$	$I_8 = [1.30854375, 1.3125]$

e $f(c_8) = f(1.30854375) = 0.001298390420$*. A aproximação não é muito "razoável"para a quantidade de cálculos realizados. Depois, podemos utilizar um método localmente convergente para se obter uma melhor aproximação para a solução.*

Com quatro iterações, a solução (aproximada) seria $x_4 = 1.3125$ *e* $f(x_4) \approx -0.002787366752$*. Para satisfazer o segundo critério, seriam necessárias mais aproximações !*

■

Exercício 3.4.1 *Resolva o mesmo problema considerando* $I_0 = [1.2, 1.4]$.

3.4.2 Método da Falsa Posição(*regula falsi*)

O método da Falsa posição, por vezes também designado por *regula falsi*, permite determinar o zero (que à partida supomos único) de uma função f, contínua num intervalo $[a, b]$, monótona[1] e tal que $f(a) \times f(b) < 0$.

Este método é semelhante ao método da Bissecção mas, em cada iteração o intervalo $[a_n, b_n]$ é dividido em duas partes mas, a divisão é feita no ponto x_{n+1} que corresponde à intersecção com o eixo das abcissas da recta que passa pelos pontos $(a_n, f(a_n))$ e $(b_n, f(b_n))$ tais que $f(a_n) \times f(b_n) < 0$. Seja $f(x)$ uma função contínua no intervalo $\mathcal{I} = [a, b]$, monótona tal que $f(a) \times f(b) < 0$. O método da Bissecção, considerado anteriormente, cada elemento da sucessão é o ponto médio de uma sucessão de intervalo encaixados $[a_0, b_0] = [a, b], \ldots, [a_n, b_n]$ onde se encontra a raíz, isto é,

$$x_{n+1} = \frac{a_n + b_n}{2}.$$

No método da **falsa posição** obtemos a partir da equação da recta que une os pontos $(a_n, f(a_n))$ e $(b_n, f(b_n))$, isto é, da equação

$$y = f(a_n) + \frac{f(b_n) - f(a_n)}{b_n - a_n}(x - a_n), \tag{3.19}$$

a relação

$$x_{n+1} = b_n - \frac{f(b_n)}{\frac{f(b_n)-f(a_n)}{b_n-a_n}}, \text{ com} n = 0, 1, 2, 3, \ldots \tag{3.20}$$

ou seja

$$\begin{aligned} x_{n+1} &= b_n - f(b_n)\frac{b_n - a_n}{f(b_n) - f(a_n)}, \\ &= \frac{a_n f(b_n) - b_n f(a_n)}{f(b_n) - f(a_n)}, \text{ com } n = 0, 1, 2, 3, \ldots \end{aligned}$$

Devemos salientar que uma vez que $f(a_n) f(b_n) < 0$ temos que $x_{n+1} \in]a_n, b_n[$.

Na iteração seguinte, é utilizado o intervalo $[a_n, x_{n+1}]$ ou o subintervalo $[x_{n+1}, b_n]$, consoante se verifique a condição $f(a_n) f(x_{n+1}) < 0$ ou $f(x_{n+1}) f(b_n) < 0$. No caso dificil de ocorrer, $f(x_{n+1}) = 0$, a aplicação do método pára nessa iteração.

Como este método utiliza mais informação para gerar as iterações é, em geral de convergência mais rápida. Prova-se que em certas situações a ordem de convergência do método é linear com um coeficiente assimptótico κ_∞ menor que $\frac{1}{2}$.

O método da falsa posição corresponde a aproximar a função pela recta secante nos extremos do intercalo e a utilizar o zero da recta como estimativa para o zero da função (daí o seu nome). Esta aproximação é tanto mais razaoável quanto mais o gráfico de f se aproxima de uma recta, isto é, $f'(x)$ varia pouco.

[1]Esta condição garante a existência de um único zero de f no intervalo $[a, b]$. Esta condição é empregue unicamente para facilitar a exposição do método.

Teorema 3.4.1 *Se f é contínua e estritamente monátona no intervalo $[a,b]$, $f(a) \times f(b) \leq 0$ então, o método da Falsa posição produz uma sucessão convergente para o único zero de f nesse intervalo.*

Utilizando este método, não é possível, de um modo geral, determinar à partida o número de iterações necessárias para se obter a precisão desejada para a solução aproximada. Mas, a principal desvantagem deste método ocorre quando a sucessão de raízes aproximadas se situam só à esquerda ou só á direita da raíz procurada. Esta situação que ocorre sempre que o gráfico da função apresenta concavidade voltada para cima ou para baixo no intervalo em estudo, a velocidade de convergência pode tornar-se bastante mais lenta. As condições suficientes para a convergência do método da Falsa Posição são iguais às do método da Bissecção.

Os critérios de paragem mais utilizados são:

- $|f(a_n)| < \varepsilon$ ou,
- $|x_{n+1} - x_n| \leq \varepsilon$.

O resultado seguinte indica uma forma de determinar um majorante para o erro da aproximação.

Teorema 3.4.2 *Seja f uma função continuamente diferenciável no intervalo $[a,b]$ tal que $f(a) \times f(b) \leq 0$. Sejam*

$$0 < m_1 = \min_{\xi \in [a,b]} \left|f'(\xi)\right| > 0 \text{ e } M_1 = \max_{\xi \in [a,b]} \left|f'(\xi)\right|.$$

Então, o erro de aproximação se x, a única solução de $f(x) = 0$ em $[a,b]$ pela estimativa x_{n+1} é dada por

$$|x - x_{n+1}| \leq \frac{M_1 - m_1}{m_1} |x_{n+1} - x_n|. \tag{3.21}$$

Isto é, o erro ε na iteração $n+1$ é então dado por

$$\varepsilon_{n+1} = \frac{M_1 - m_1}{m_1} |x_{n+1} - x_n| \leq \delta \Rightarrow |x - x_{n+1}| \leq \delta$$

Este método pode ser aplicado efectuando os seguintes passos:

- Para $n = 0, \ldots, ITMAX$ (número máximo de iterações),

  1)
  $$x_{n+1} = \frac{f(b_n)\, a_n - f(a_n)\, b_n}{f(b_n) - f(a_n)};$$

  2) se $f(a_n) \times f(x_{n+1}) < 0$ então
  $$a_{n+1} = a_n \text{ e } b_{n+1} = x_{n+1};$$

  3) caso contrário, faça-se
  $$a_{n+1} = x_{n+1} \text{ e } b_{n+1} = b_n;$$

4) Parar se algum critério de paragem for satisfeito antes de $ITMAX$

Podemos, em relação a este método apresentar o seguinte caso especial. Quando

- $f(x)$ é concava ou convexa em $[a, b]$;
- a segunda derivada existe em $[a, b]$;
- e $f''(x)$ não muda de sinal nesse intervalo.

Tem-se sempre uma das extremidades fixas. Este caso especial chama-se **Método das Cordas**.

Exemplo 3.4.1 *Utilizando o método da Corda Falsa, determine uma aproximação para* $\sqrt{2}$.

Resolução 3.4.3 *Consideremos a função* $f(x) = x^2 - 2$ *cuja raízes são* $x = \pm\sqrt{2}$. *O gráfico da função* $f(x)$ *no intervalo* $\mathcal{I} = [0, 2]$ *é*

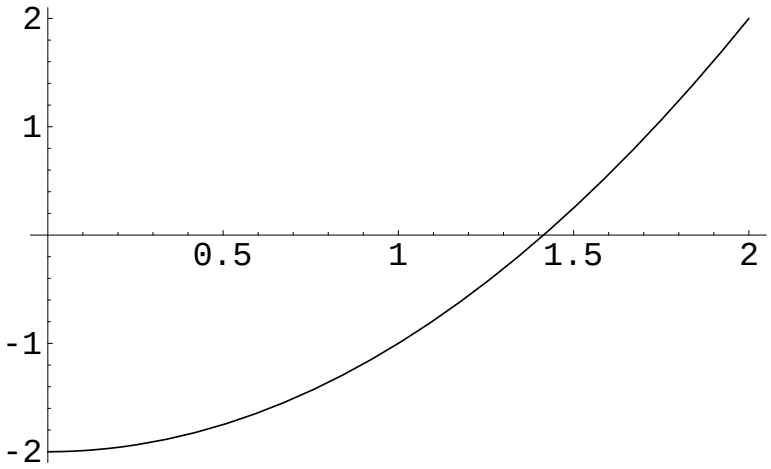

Figura 3.5: Gráfico da função $f(x) = x^2 - 2$.

Observando o gráfico verificamos que o zero de $f(x)$ *se encontra no intervalo* $\mathcal{I} = [1, 2]$. *Aplicando a fórmula da recorrência,*

$$c_n = \frac{a_n\left(b_n^2 - 2\right) - b_n\left(a_n^2 - 2\right)}{b_n^2 - a_n^2}. \tag{3.22}$$

Depois, é necessário comparar os sinais de $f(a_n)$, $f(c_n)$ *e* $f(b_n)$.
Então, se

- *Se* $f(a_n) \times f(c_n) < 0$ *então fazemos* $b_n = c_n$ *e o novo intervalo é* $\mathcal{I}_n = [a_n, b_n]$;
- *Se* $f(c_n) \times f(b_n) < 0$ *então fazemos* $a_n = c_n$ *e o novo intervalo é* $\mathcal{I}_n = [a_n, b_n]$.

n	$\mathcal{I}_n$	a_n	b_n	c_n	$f(a_n)$	$f(c_n)$	$f(b_n)$
0	$\mathcal{I}_0 = [1, 2]$	1	2	$1.3333333\,(3)$	< 0	< 0	> 0
1	$\mathcal{I}_1 = [1.33333\,(3), 2]$	$1.33333\,(3)$	2	1.4	< 0	< 0	> 0
2	$\mathcal{I}_2 = [1.4, 2]$	1.4	2	1.41176	< 0	< 0	> 0
3	$\mathcal{I}_3 = [1.41176, 2]$	1.41176	2	1.41379	< 0	< 0	> 0
4	$\mathcal{I}_4 = [1.41379, 2]$	1.41379	2	1.41414	< 0	< 0	> 0
5	$\mathcal{I}_5 = [1.41414, 2]$	1.41414	2	1.4142	< 0	< 0	> 0
6	$\mathcal{I}_6 = [1.4142, 2]$	1.4142	2	1.41421	< 0	< 0	> 0
6	$\mathcal{I}_7 = [1.41421, 2]$	1.41421	2	1.41421	< 0	< 0	> 0

■

Nota 3.4.1 *Muitos dos* 110 *problemas encontrados nos Papiros de Rhindi e de Moscovo*[2] *são de origem prática. Para muitos desses problemas a resolução não exigia mais do que uma equação linear simples e o método utilizado pelos egípcios para a resolução de equações era o Método da Falsa Posição. Por exemplo, para resolver a seguinte equação:*

$$x + \frac{x}{7} = 24. \tag{3.23}$$

Primeiramente, assume-se um valor conveniente para x, de modo a eliminar o denominador da fracção. Neste caso, fazemos $x = 7$. Então temos:

$$\begin{aligned} 7 + \frac{7}{7} = y & \\ 7 + 1 &= y \\ y &= 8 \end{aligned} \tag{3.24}$$

Dividimos o valor real da equação pelo valor encontrado na equação falsa (3.24):

$$\frac{24}{8} = 3 \tag{3.25}$$

Agora, multiplicamos o resultado obtido em (3.25) pelo valor que foi assumido para x:

$$7 \times 3 = 21.$$

Encontrando, assim, o valor real de x:

$$x + \frac{x}{7} = 24 \tag{3.26}$$

$$\begin{aligned} 21 + \frac{21}{7} &= 24 \\ 24 &= 24 \end{aligned} \tag{3.27}$$

[2]Também conhecido por Papiro *Golonishev* em referência ao seu proprietário Vladimir Golenishchev: É um papiro egípcio em forma de uma estreita tira de 5,5m de comprimento por 8cm de largura, com 25 problemas matemáticos

3.4.3 Método de Newton-Raphson

O Método de Newton (ou método Newton-Raphson[3]) é um método localmente convergente para estimar as raízes de uma função. Para isso, considera-se um ponto do domínio da função; calcula-se a equação da tangente (derivada) da função nesse ponto; calcula-se o ponto de intercepção da tangente ao eixo das abcissas; calcula-se o valor da função nesse ponto, e repete-se o processo, que deve tender a uma das raízes da função rapidamente, ou não tender a nada, deixando isso claro logo.

Seja $f \in \mathcal{C}^2([a,b])$ tal que $f(x) = 0$ para algum x pertencente a $[a,b]$. Seja x^* uma aproximação para x de tal forma que

a) $f'(x) \neq 0$,

b) $|x - x^*| \ll 1$.

Então a sucessão $\{x_n\}_{n=0}^{\infty}$ dada pela fórmula recursiva

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \ldots \tag{3.28}$$

converge para a solução do problema $f(x) = 0$.

Teorema 3.4.3 *O método de Newton-Raphson converge para uma solução se as seguintes condições são válidas:*

1. $f \in \mathcal{C}^2([a,b])$;
2. $f(a) \times f(b) < 0$;
3. $f'(x)$ *e* $f''(x)$ *não mudam de sinal em* $[a,b]$;
4. $f(x_0) \times f''(x_0) > 0$.

ou,

1. $f \in \mathcal{C}^2([a,b])$;
2. $f(a) \times f(b) < 0$;
3. $f'(x) \neq 0, \forall x \in [a,b]$;
4. $f''(x) \leq 0$ *ou* $f''(x) \geq 0, \forall x \in [a,b]$;
5. $\left|\frac{f(a)}{f'(a)}\right| \leq b - a$ *e* $\left|\frac{f(b)}{f'(b)}\right| \leq b - a$
6. $f(x_0) \times f''(x_0) > 0$.

Devemos salientar que o método de Newton-Raphson pode falhar quando $f'(x)$ assum valores muito pequenos.

O método pode ser aplicado seguindo os seguintes passos:

[3]Joseph Raphson (1648-1715)

P_1 Fazer $x = x_0 - \dfrac{f(x_0)}{f'(x_0)}$;

P_2 Verificar se o valor obtido satisfaz o critério de paragem utilizado;

P_3 Fazer $x = x_0$.

P_4 Repetir o procedimento desde o passo P_1.

É importante escolher uma aproximação inicial muito próxima da raíz. Este facto pode determinar a velocidade de convergência do método. Portanto, se a aproximação inicial está 'àfastada"da raíz, a convergência pode ser lenta.

Exercício 3.4.2 *Consegue agora deduzir a fórmula (1.1), apresentada por Heron?*

Teorema 3.4.4 *Seja $\mathcal{I} = [a, b]$ uma vizinhaça do zero α de uma função $f \in \mathcal{C}^2([a, b])$ e suponha-se válidas as condições:*

1) $0 < m_1 \leq |f'(x)|$, $\forall x \in \mathcal{I}$;

2) $0 < |f''(x)| \leq m_2$, $\forall x \in \mathcal{I}$;

3) $\frac{m_2}{2m_1}(b - a) < 1$.

Então, escolhendo para x_0 o extremo (a ou b) do intervalo $\mathcal{I}$ em que a função f tem o mesmo sinal que a sua segunda derivada, isto é, de forma a que $f(x) f''(x_0) > 0$, o método de Newton converge para a raíz α e o erro das iteradas consecutivas satisfaz a relação

$$|e_{n+1}| \leq M |e_n|^2 . \tag{3.29}$$

Exemplo 3.4.3 *Considere a equação*

$$2x + \ln x = 1.$$

1) Mostre que a equação dada tem uma única raíz.

2) Pretende-se usar o método de Newton-Raphson para obter a raíz da equação dada.

   a) Indique um intervalo de comprimento $\frac{1}{2}$ que contenha a raíz da equação dada.

   b) Verifique se no intervalo indicado são satisfeitas todas as condições de convergência e escolha uma aproximação inicial adequada.

   c) Efecture duas iterações e indique majorantes para os erros das aproximações obtidas;

   d) Determine uma solução aproximada da raíz da equação de modo a garantir um erro absoluto não superior a 10^{-6}.

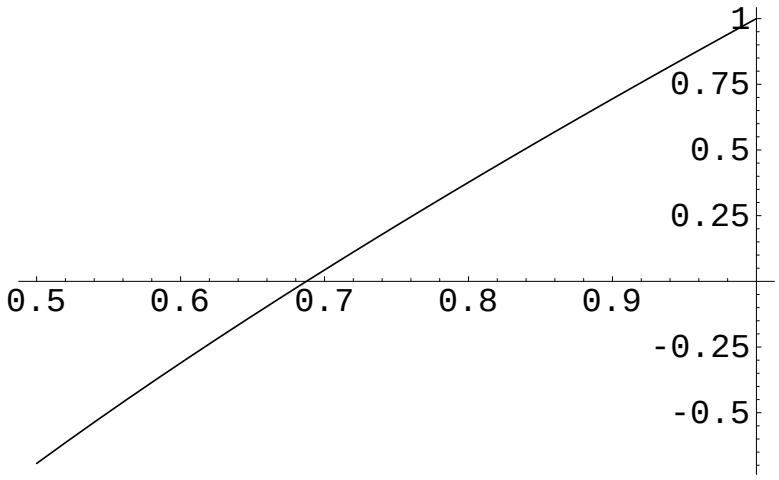

Figura 3.6: Gráfico da função $f(x) = 2x + \ln(x) - 1$

Resolução 3.4.4 *1) Seja* $f(x) = 2x + \ln(x) - 1$. *Os zeros de* f *são as raízes da equação* $f(x) = 0$. *O domínio da função* $f(x)$ *é*

$$D_f = \{x \in \mathbb{R} : \ x > 0\} =]0, +\infty[\, ,$$

e é contínua e diferenciável nesse intervalo.

Tendo em conta a representação gráfica da função,

verificamos que $f\left(\frac{1}{2}\right) = \ln \frac{1}{2} = -\ln 2 < 0$ *e que* $f(1) > 1$. *A sua derivada é*

$$f'(x) = 2 + \frac{1}{x} > 0, \ \ \forall x \in]0, +\infty[\, .$$

Isto é, a função é estritamente crescente. Consequentemente, a função admite um e um só zero no intervalo $]0, +\infty[$, *em particular no intervalo* $\left[\frac{1}{2}, 1\right]$.

2) a) Podemos apresentar o intervalo $\left[\frac{1}{2}, 1\right]$.

b) Neste caso facilmente verificamos que $f \in \mathcal{C}^2\left(\left[\frac{1}{2}, 1\right]\right)$, $0 < m_1 = 3 \leq |f'(x)| = \left|2 + \frac{1}{x}\right|$, $\forall x \in \left[\frac{1}{2}, 1\right]$, $0 < |f''(x)| = \left|-\frac{1}{x^2}\right| \leq 4$, $\forall x \in \left[\frac{1}{2}, 1\right]$, *e* $M(b-a) < 1$ *com* $M = \frac{m_2}{2m_1} = \frac{2}{3}$. *Assim, para aproximação inicial devemos escolher o extremo do intervalo* $\left[\frac{1}{2}, 1\right]$ *para o qual a função tem o mesmo sinal que a segunda derivada, isto é,* $x_0 = \frac{1}{2}$.

c) Neste caso,

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \\ &= x_n - \frac{2x_n + \ln x_n - 1}{2 + \frac{1}{x_n}} \end{aligned}$$

O que leva a

n	x_n	*majorante de* $\|e_n\|$
0	$\frac{1}{2}$	$\frac{1}{2}$
1	0.67329	$\left(\frac{1}{2}\right)^2 \times \frac{2}{3} \approx 0.16667$
2	0.68735	$0.16667^2 \times \frac{2}{3} \approx 1.8518 \times 10^{-2}$

d) Na tabela seguinte apresentamos as 4 primeiras iterações utilizando o método de Newton:

n	x_n	*majorante de* $\|e_n\|$
0	$\frac{1}{2}$	$\frac{1}{2}$
1	0.67329	1.666667×10^{-1}
2	0.68735	1.8518519×10^{-2}
3	6.8741126×10^{-1}	2.2862368×10^{-4}
4	6.8741126×10^{-1}	3.4845860×10^{-8}

a resposta é $x_4 = 0.68741126$.

■

Exemplo 3.4.4 *Utilizando o método de Newton-Raphson, determine uma aproximação para* $\sqrt{2}$ *com erro* $\varepsilon = |x_{n+1} - x_n| < 10^{-8}$.

Resolução 3.4.5 *Como vimos anteriormente,* $\sqrt{2}$ *é uma zero da função* $f(x) = x^2 - 2$. *A existência e unicidade do zero no intervalo* $\mathcal{I} = [1, 2]$ *foi provada anteriormente. Vamos em primeiro lugar, verificar se está garantida a convergência do método de Newton-Raphson no intervalo* $\mathcal{I} = [1, 2]$.

Se $f(x) = x^2 - 2$ *então* $f'(x) = 2x$. *Vamos, em primeiro lugar verificar se as condições do teorema 3.4.4 são satisfeitas em* $\mathcal{I}$.

Neste caso,

$$m_1 = \max_{\xi \in [1,2]} \left| f'(x) \right| = \max_{\xi \in [1,2]} |2x| = 4$$

e

$$m_2 = \max_{\xi \in [1,2]} \left| f'0(x) \right| = \max_{\xi \in [1,2]} |2| = 2.$$

E portanto,

$$\frac{m_2}{2m_1}(b - a) = \frac{2}{2 \times 4}(2 - 1) = \frac{1}{4} < 1.$$

Logo, as condições do teorema 3.4.4 são válidas.

Utilizando o teorema 3.4.3, facilmente verificamos que as condições $1 - 4$ *são válidas. Em relação à condição* 5, *temos que para* $x = 1$,

$$\left| \frac{f(1)}{f'(1)} \right| = \left| \frac{1}{2} \right| < 1$$

e quando $x = 2$,

$$\left| \frac{f(2)}{f'(2)} \right| = \left| \frac{1}{2} \right| < 1.$$

em relação à última condição, considerando $x_0 = 2$, temos que

$$f(2) \times f''(2) > 0.$$

Portanto, podemos aplicar o método de Newton com, por exemplo, $x_0 = 2$.
A fórmula de iterativa é dada por

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n}, n \geq 0,\ x_0 = 2. \tag{3.30}$$

Temos então

n	x_n	ε_n
0	2	0.5
1	1.5	0.08333333333
2	1.41666666 (6)	0.002450980392
3	1.414215686	$2.123625731 \times 10^{-6}$
4	1.41421356	2.37309×10^{-9}

■

Exemplo 3.4.5 *Utilize o método de Newton-Raphson para calcular a raíz da equação*

$$e^{-x} - \ln x = 0.$$

Considere os seguintes critérios de paragem: número máximo de 4 iterações; $|x_{n+1} - x_n| \leq 10^{-4}$ e $|f(x_n)| \leq 10^{-4}$.

Resolução 3.4.6 *Vamos considerar que $x_0 = 1$. Neste caso $f(x) = e^{-x} - \ln x$ e portanto,*

$$f'(x) = -e^{-x} - \frac{1}{x}.$$

Logo, tendo em conta (3.28) temos o seguinte esquema de recorrência:

$$x_{n+1} = x_n + \frac{e^{-x_n} - \ln(x_n)}{e^{-x_n} + \frac{1}{x_n}},\ \ n \geq 0. \tag{3.31}$$

Temos então

$$\begin{aligned}
x_0 &= 1 && (3.32)\\
x_1 &= x_0 + \frac{e^{-x_0} - \ln(x_0)}{e^{-x_0} + \frac{1}{x_0}} = 1.268941421\\
x_2 &= x_1 + \frac{e^{-x_1} - \ln(x_1)}{e^{-x_1} + \frac{1}{x_1}} = 1.309108403\\
x_3 &= 1.309799389\\
x_4 &= 1.309799586\\
x_5 &= 1.309799586 && (3.33)
\end{aligned}$$

E,

$$\begin{aligned} f(x_0) &= e^{-1} \approx 0.3678794411 && (3.34)\\ f(x_1) &= 0.04294603551 \\ f(x_2) &= 0.0007144373195 \\ f(x_3) &= 2.033673160 \times 10^{-7} \\ f(x_4) &= -2.018246585 \times 10^{-10} \\ f(x_5) &= -2.018246585 \times 10^{-10} && (3.35) \end{aligned}$$

■

3.4.4 Método da Tangente fixa

O método da Tangente fixa é um caso particular do método de Newton-Raphson. No denominador em vez de calcular $f'(x_n)$ consideramos (sempre !) $f'(x_0)$. A vantagem desta alteração é que pode diminuir o esforço computacional. A relação de recorrência é então:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}, \quad k = 0, 1, \ldots \tag{3.36}$$

Exemplo 3.4.6 *Utilize o método da Tangente Fixa para calcular a raíz da equação* $e^{-x} - \ln x = 0$. *Considere os seguintes critérios de paragem: número máximo de* 4 *iterações;* $|x_{n+1} - x_n| \leq 10^{-4}$ *e* $|f(x_n)| \leq 10^{-4}$.

Resolução 3.4.7 *Vamos considerar que* $x_0 = 1$. *Neste caso* $f(x) = e^{-x} - \ln x$ *e portanto,*

$$f'(x) = -e^{-x} - \frac{1}{x} \text{ e } f'(1) \approx -1.367879441.$$

Logo, tendo em conta (3.36) temos o seguinte esquema de recorrência:

$$x_{n+1} = x_n + \frac{e^{-x_n} - \ln(x_n)}{-1.367879441}, \quad n \geq 0. \tag{3.37}$$

Temos então

$$\begin{aligned} x_0 &= 1 && (3.38)\\ x_1 &= x_0 + \frac{e^{-x_0} - \ln(x_0)}{-1.367879441} = 1.268941421 \\ x_2 &= x_1 + \frac{e^{-x_1} - \ln(x_1)}{-1.367879441} = 1.300337488 \\ x_3 &= 1.307513555 \\ x_4 &= 1.309242142 \\ x_5 &= 1.309663353 \\ x_6 &= 1.309766274 \\ x_7 &= 1.309791439 && (3.39)\\ &\vdots \end{aligned}$$

Devemos salientar que aplicando este método, a velocidade de convergência é menor. ■

Exercício 3.4.3 *Resolva o mesmo problema considerando* $x_0 = 1.3$.

3.4.5 Método da Secante

Este método é uma variação do método de Newton-Raphson que é aplicado a funções cujas derivadas são "complicadas"de calcular. A ideia é substituir a derivada de $f(x)$, $f'(x)$ por uma sua aproximação, isto é,

$$f'(x_{n-1}) = \lim_{x_n \to x_{n-1}} \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \tag{3.40}$$

o que nos permite escrever a aproximação

$$f'(x_{n-2}) \approx \frac{f(x_{n-1}) - f(x_{n-2})}{x_{n-1} - x_{n-2}}. \tag{3.41}$$

Portanto,

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}, \quad k = 1, 2, \ldots \tag{3.42}$$

Este método, ao contrário dos anteriores requer duas aproximações iniciais.
O método pode ser aplicado seguindo os seguintes passos:

P_1 Fazer $y_0 = f(x_0)$ e $y_1 = f(x_1)$ (x_0 e x_1 são as duas aproximações iniciais).

P_2 Calcular

$$x = \frac{y_1 - y_0(x_1 - x_0)}{y_1 - y_0}.$$

P_3 Verificar se o valor obtido satisfaz o critério de paragem utilizado;

P_4 Fazer $x_0 = x_1$, $y_0 = y_1$, $x_1 = x$ e $y_1 = f(x_1)$.

P_5 Repetir o procedimento desde o passo P_2.

Os critérios de convergência para o Método da Secante são idênticos aos do Método de Newton-Raphson.

Teorema 3.4.5 *Seja* $\mathcal{I} = [a, b]$ *uma vizinhaça do zero* α *de uma função* $f \in C^2([a, b])$ *e suponha-se válidas as condições:*

1) $0 < m_1 \leq |f'(x)|$, $\forall x \in \mathcal{I}$;

2) $0 < |f''(x)| \leq m_2$, $\forall x \in \mathcal{I}$;

3) $\frac{m_2}{2m_1}(b - a) < 1$.

Então, escolhendo para x_0 e x_1 dois pontos do intervalo $\mathcal{I}$ tais que

$$f(x_0) f''(x_0) > 0 \text{ e } f(x_1) f''(x_1) > 0,$$

o método da secante converge para a raíz α.

Teorema 3.4.6 *Nas condições do teorema 3.4.5, o erro do método da secante satisfaz a relação*

$$|e_{n+1}| \leq M |e_n| |e_{n-1}|. \tag{3.43}$$

Prova-se que a ordem de convergência do método da secante (supondo que α é uma raíz simples) é $p = \dfrac{1+\sqrt{5}}{2} \approx 1.618$.

Exemplo 3.4.7 *Utilize o método da Secante para calcular a raíz da equação*

$$e^{-x} - \ln x = 0.$$

Resolução 3.4.8 *A existência e unicidade de um zero já foi previamente estudada no exemplo 3.4.2. Neste caso é necessário apresentar duas aproximações iniciais. Vamos considerar $x_0 = 1$ e $x_1 = 2$. Portanto, temos*

$$\begin{aligned}
x_2 &= x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1) - f(x_0)} \approx 1.397410482, \qquad (3.44)\\
x_3 &= x_2 - \frac{f(x_2)(x_2 - x_1)}{f(x_2) - f(x_1)} \approx 1.285476120,\\
x_4 &= x_3 - \frac{f(x_3)(x_3 - x_2)}{f(x_3) - f(x_2)} \approx 1.310676758,\\
x_5 &= x_4 - \frac{f(x_4)(x_4 - x_3)}{f(x_4) - f(x_3)} \approx 1.309808398,,\\
x_6 &\approx 1.309799582,\\
x_7 &\approx 1.309799585. \qquad (3.45)
\end{aligned}$$

■

Exercício 3.4.4 *Resolva o mesmo problema considerando $x_0 = 1.2$ e $x_1 = 1.4$.*

Exemplo 3.4.8 *Localize a raíz da equação $x \ln(x) - 1 = 0$ e determine com um erro absoluto inferior a 5×10^{-2} pelos métodos*

1. *Bissecção;*
2. *Newton;*
3. *Secante.*

Resolução 3.4.9 *Seja* $f(x) = x \ln x - 1$, *para* $x > 0$. *A função* f *é contínua em* $\mathbb{R}^+$ *pois é o produto de duas funções contínuas,* $f_1(x) = x$ *e* $f_1(x) = \ln(x)$ *ao qual subtraímos a constante* 1. *A derivada da função é*

$$f'(x) = \ln x + 1 > 0, \quad x \geq 1.$$

Mais, temos que $f(1) \times f(2) < 0$. *Consequentemente, a função admite um e um só zero em* $\mathcal{I}_0 = [1, 2]$. *A representação gráfica da função em* $\mathcal{I}_0 = [1, 2]$ *é*

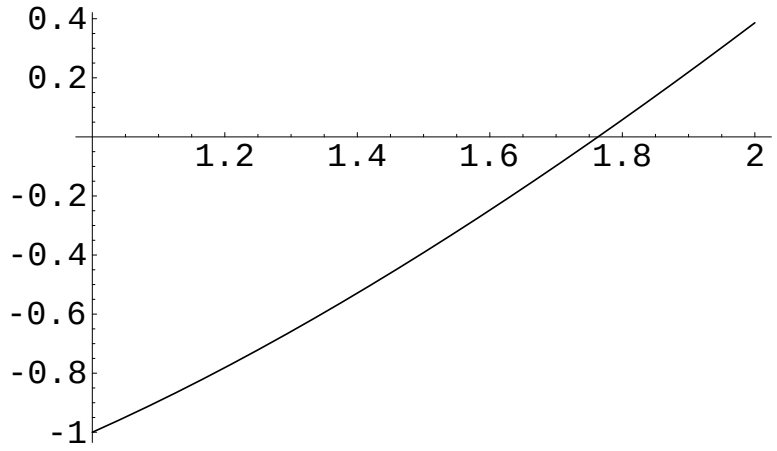

Figura 3.7: Gráfico da função $f(x) = x \ln(x) - 1$

1. *Método da Bissecção:* $x_k = \dfrac{a_k + b_k}{2}$, $\varepsilon < \dfrac{b-a}{2^{k+1}}$.

k	$\mathcal{I}_k = [a_k, b_k]$	x_k	$f(a_k)$	$f(x_k)$	$f(b_k)$	$\varepsilon <$
0	$[1, 2]$	1.5	$-$	$-$	$+$	0.5
1	$[1.5, 2]$	1.75	$-$	$-$	$+$	0.25
2	$[1.75, 2]$	1.875	$-$	$+$	$+$	0.125
3	$[1.75, 1.875]$	1.8125	$-$	$+$	$+$	0.0625
4	$[1.75, 1.8125]$	1.7813	$-$	$+$	$+$	0.0313
5	$[1.75, 1, 7813]$	1.76657	$-$	$+$	$+$	0.0156

portanto,

$$\tilde{x} = 1.7813, \quad k = 4$$

2. *Método de Newton:*

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad \varepsilon < \left| \frac{f(x_k)}{f'(x_k)} \right|$$

k	x_k	$f(x_k)$	$f'(x_k)$	$\varepsilon <$
0	1	-1	1	1
2	2	0.3863	1.6932	0.2282
2	1.7719	0.0136	1.5721	0.0087
3	1.7633	0.0001	1.5672	0.0
4	1.7632	0		

Portanto,

$$\tilde{x} = 1.7719, \quad k = 2$$

3. *Método da Secante:*

$$x_{k+1} = x_k - \frac{f(x_k)}{\dfrac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}}, \quad \varepsilon_{k+1} \approx |x_{k+1} - x_k|.$$

Vamos considerar que $x_0 = 2$ *e* $x_1 = 1$. *Então,*

$$\begin{aligned}
x_0 &= 2 \\
x_1 &= 1 \\
x_2 &= x_1 - \frac{f(x_1)}{\frac{f(x_1)-f(x_0)}{x_1-x_0}} = 1.7213 \\
x_3 &= 1.7716, \quad \varepsilon_3 \approx 0.05 \\
x_4 &= 1.7631 \\
x_5 &= 1.7632, \quad \varepsilon_5 \approx |1.7632 - 1.7631| = 0.0001 = 10^{-4}.
\end{aligned}$$

Logo, $\tilde{x} = 1.7716, \quad k = 3.$

■

Exemplo 3.4.9 *Determine a menor raíz positiva da equação* $e^{-x} - \sin(x) = 0$ *com um erro absoluto inferior a* $5 \times 10^{-2} = 0.05$ *pelos métodos da bissecção, Newton-Raphson e Secante.*

Resolução 3.4.10 *Vamos em primeiro apresentar a representação gráfica da função que nos fornece uma estimativa quer, para o método da bissecção como para os métodos de Newton-Raphson e da Secante.*

Seja $f(x) = e^{-x} - \sin(x)$, *para* $x > 0$. *A função* f *é contínua em* $\mathbb{R}^+$ *pois é a soma de duas funções contínuas,* $f_1(x) = e^{-x}$ *e* $f_2(x) = -\sin(x)$. *A derivada da função é*

$$f'(x) = e^{-x} - \cos(x),$$

Mais, temos que $f(0) \times f\left(\frac{\pi}{2}\right) < 0$. *Consequentemente, a função admite um e um só zero em* $\mathcal{I}_0 = [1, 2]$.

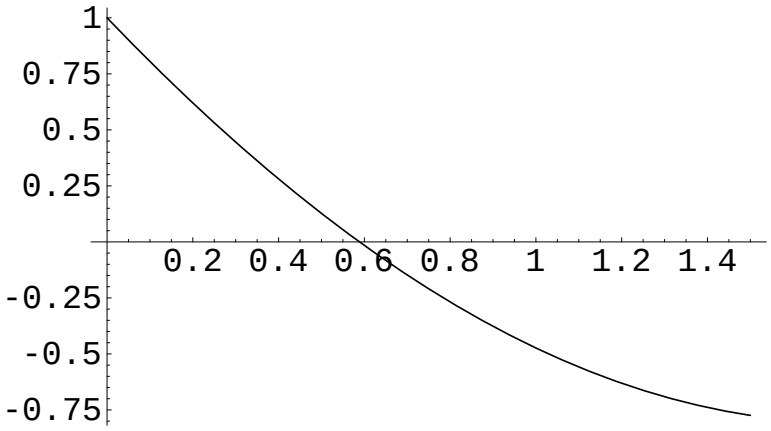

Figura 3.8: Gráfico da função $f(x) = e^{-x} - \sin(x)$

1. Método da Bissecção

$$x_k = \frac{a_k + b_k}{2}, \quad \varepsilon_k < \frac{b-a}{2^{k+1}}$$

k	$\mathcal{I}_k = [a_k, b_k]$	x_k	$f(a_k)$	$f(x_k)$	$f(b_k)$	$\varepsilon <$
0	$\left[0, \frac{\pi}{2}\right]$	0.7854	+	−	−	0.7854
1	$[0, 0.7854]$	0.3927	+	+	−	0.3927
2	$[0.3927, 0.7854]$	0.5891	+	−	−	0.1964
3	$[0.3927, 0.5891]$	0.4909	+	+	−	0.0982
4	$[0.4909, 0.5891]$	0.5400	+	+	−	0.0491
5	$[0.5400, 0.5891]$	0.5646	+	+	−	0.0246
6	$[0.5646, 0.5891]$	0.5769	+	+	−	0.0123
7	$[0.5769, 0, 5891]$	0.5830	+	+	−	0.0062
8	$[0.5830, 0.5891]$	0.6861	+	+	−	0.0031
9	$[0.5861, 0.5891]$	0.5861	+	+	−	0.0012
10	$[0.5876, 0.5891]$					

2. Método de Newton

$$f(x) = e^{-x} - \sin(x), \quad f'(x) = -e^{-x} - \cos(x),$$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad \varepsilon < \left|\frac{f(x_k)}{f'(x_k)}\right|$$

k	x_k	$f(x_k)$	$f'(x_k)$	$\varepsilon <$
0	0	1	−2	0.5
1	0.5	0.1271	−1.41841	0.0856
2	0.5856	0.0041	−1.3902	0.0029
3	0.5885	0.00005	−1.3869	0.00004

3. Método da Secante:

$$x_{k+1} = x_k - \frac{f(x_k)}{\frac{f(x_k)-f(x_{k-1})}{x_k - x_{k-1}}}, \quad \varepsilon_{k+1} \approx |x_{k+1} - x_k| .$$

Vamos considerar $x_0 = 0$ *e* $x_1 = \frac{\pi}{2}$. *Aplicando a fórmula temos*

$$\begin{aligned}
x_0 &= 0 \\
x_1 &= \frac{\pi}{2} \\
x_2 &= 0.8765, \quad \varepsilon \approx 0.6943 \\
x_3 &= 0.3203, \quad \varepsilon \approx 0.5562 \\
x_4 &= 0.6198, \quad \varepsilon \approx 0.2995 \\
x_5 &= 0.5916, \quad \varepsilon \approx 0.0282 \\
x_6 &= 0.5885, \quad \varepsilon \approx 0.0031 \\
x_7 &= 0.5885
\end{aligned}$$

Nota 3.4.2 *Considerando* $x_0 = \frac{\pi}{2}$ *e* $x_1 = 0$, *obtemos* $x_2 = 0.8765$, $x_3 = 0.6482$, $x_4 = 0.5803$, $x_5 = 0.5887$, $x_6 = 0.5885$ *e* $x_7 = 0.5885$.

■

3.5 Exercícios

1. Utilizando os teoremas fundamentais indique o número de raízes de cada uma das seguinte equações:

   a) $x^3 - x + 1 = 0$
   b) $x^2 - x + 1 = 0$
   c) $x^4 - \frac{1}{2}x - 1.55 = 0$
   d) $x^2 - 4x - 1 = 0$
   e) $x^2 - 4x + 4 = 0$
   f) $x^2 - 4x + 8 = 0$

2. Localize graficamente os zeros das equações indicadas na alínea anterior.

3. Localize a menor raíz positiva da função

$$f(x) = x - \tan(x + 1)$$

4. Considere a função

$$f(x) = x^3 - x - 1.$$

   Prove que f admite um e um só zero no intervalo $\mathcal{I} = [1, 2]$.

5. Localize graficamente os zeros de cada uma das seguintes equações

   a) $\cos(x) = x^3$;
   b) $\ln(x) + x^3 = 0$;
   c) $|x| - e^x = 0$;
   d) $x^{2.1} - 4x = 0$;
   e) $(x-2)^2 - \ln(x) = 0$ nos intervalos $[1, 2]$ e $[e, 4]$;
   f) $2x\cos(2x) - (x-2)^2 = 0$ nos intervalos $(2, 3)$ e $(3, 4)$.

6. Utilizando a alínea anterior e o método da bissecção determine todas as raízes (reais) de cada uma das seguintes equações:

   a) $x^3 - x + 1 = 0$
   b) $x^2 - x + 1 = 0$
   c) $x^4 - \frac{1}{2}x - 1.55 = 0$
   d) $x^2 - 4x - 1 = 0$
   e) $x^2 - 4x + 4 = 0$
   f) $x^2 - 4x + 8 = 0$

   Utilize o seguinte critério de paragem: o número máximo de 10 iterações ou $|f(x_k)| < 10^{-5}$.

7. Resolva a equação $x - e^{-x}$ no intervalo $[0, 1]$, utilizando o método da Bissecção ($\varepsilon < 0.05$).

8. Utilize o método da Bissecção para aproximar a solução, com erro inferior a 10^{-1}, da equação

$$x + \frac{1}{2} + 2\cos(\pi x) = 0,$$

   no intervalo $\left[\frac{1}{2}, 1\right]$.

9. Quantas vezes teríamos que aplicar o método da Bissecção na questão anterior se quiséssemos determinar x_k de tal forma que $|x^* - x_k| < 10^{-5}$, sendo x^* a solução exacta.

10. Localize graficamente as raízes da equação

$$\ln(x) - 3 + \frac{3}{2} = 0,$$

e determine um valor aproximado de uma delas com uma casa decimal correcta, utilizando o método da Bissecção.

11. Pretende-se construir um tanque cúbico com uma capacidade de $25000l$. Determina uma aproximação para o comprimento do lado do tanque utilizando o método da Bissecção 4 vezes e indique a precisão do resultado obtido.

12. Utilizando o método de Newton-Raphson, determine uma solução de cada uma das seguintes equações:

a) $3x^2 - 4\cos(x) = 0$, $x_0 = 1$ (convergência rápida)

b) $1 - 10x + 25x^2$, $x_0 = 1$ (convergência lenta)

c) $\arctan(x) = 0$, $x_0 = 1.35$ (Converge, Oscila)

d) $x^3 - x + 3 = 0$, $x_0 = 0$ (**Não converge**; Cíclico)

e) $xe^{-x} = 0$, $x_0 = 2$ (**Não converge**; Diverge para infinito)

f) $Arctang(x) = 0$, $x_0 = 1.4$ (**Não converge**; Diverge e oscila)

13. Calcule uma aproximação para a raíz real

$$f(x) \equiv (x-1)e^x - 2 = 0,$$

usando o método de Newton. Utilize como critério de paragem a condição $|f(x_k)| \leq 10^{-5}$.

14. Pretende-se determinar o zero da função

$$f(x) = \ln(x) - 10\sin(x),$$

que se encontra no intervalo $\mathcal{I} = [3, \pi]$.

a) Prove que no intervalo $\mathcal{I}$ existe uma e uma só solução da equação $f(x) = 0$;

b) Utilizando o método da Falsa Posição, encontre uma solução aproximada da raíz pretendida com um erro absoluto inferior a 0.01.

15. Considere a função

$$f(x) = \ln\left(4 - x^2\right) - x.$$

a) Indique o seu domínio.

b) Localize graficamente as raízes da equação $f(x) = 0$.

c) Verifique se a função satisfaz as condições de convergência do método de Newton-Raphson nos intervalos $\mathcal{I}_1 = [-1.99, -1.9]$ e no intervalo $\mathcal{I}_1 = [1, 1.5]$

d) Utilizando o método de Newton-Raphson, determine uma aproximação para a maior raíz real da equação $f(x) = 0$ com um erro inferior a $10-2$.

e) Resolva as questões anteriores com

$$f(x) = x\cos(x) - e^x,$$

no intervalo $\mathcal{I} = [-1, -0.2]$.

16. Localize graficamente a raíz real posiitva da equação

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

e obtenha um valor aproximado da raíz efectuando duas iterações com o método de Newton.

17. Considere a função $g(x) = \cos\left(x^2\right) - \dfrac{x}{10}$.

(a) Verifique que as condições de convergência do método de Newton-Raphson são satisfeitas no intervalo $[1.2, 1.22]$.

(b) Utilizando o método de Newton-Raphson indique uma aproximação para o zero no intervalo indicado apresentando um limite superior para o erro.

18. Verifique que a equação

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

tem uma raíz em $[1, 2]$. Calcule uma aproximação para essa raíz com quatro algarismos significativos.

19. a) Para que valores de γ a equação $xe^{-x} = \gamma$ não tem raízes reais ?

b) Localize as raízes da equação

$$xe^{-x} - 0.25 = 0.$$

c) Use o método da Bissecção para calcular a menor raíz da equação com um erro inferior a $\frac{1}{2} \times 10^{-3}$.

d) Para obter eficientemente uma aproximação com maior precisão para a raíz referida, tome como aproximação inicial a calculada na alínea anterior (com o método da Bissecção) e use o método de Newton.

20. Compare o funcionamento dos métodos da Bissecção, Secante e Newton no cálculo da raíz real de cada uma das equações

a) $x^3 - x - 1 = 0$.

b) $x - e^{-x} = 0$.

21. Partindo das aproximações iniciais encontradas graficamente, determine, utilizando o método de Newton-Raphson as raízes das equações

    a) $e^{-2x} - x^2 = 0$, no intervalo $\mathcal{I} = [0, 1]$;

    b) $-e^{-2x} + \cos(x) = 0$, no intervalo $\mathcal{I} = [1, 2]$

22. Considere a equação
$$x^2 - 4x + 4 = 0.$$

    a) Resolva-a utilizando o método da bissecção. Como critério de paragem considere que $|f(x^*)| \leq 10^{-6}$.

    b) Resolva-a utilizando o método de Newton considerando $x_0 = 1$.

    c) Justifique o comportamento "anormal"dos dois métodos.

23. Determine a raíz positiva mínima da equação $\tan(x) = x$. Utilize como critério a condição $|f(x_k)| < 10^{-5}$.

24. Considere a função $f(x) = x^4 - 16$.

    a) Indique um intervalo em que a equação $f(x) = 0$ admite uma e uma só solução. Justifique devidamente a sua resposta.

    b) A equação admite raízes complexas. Utilizando o método de Newton-Raphson, com $x_0 = i$, $i^2 = -1$, determine as restantes. [4]

25. Considere a equação $x^3 - 3x^2 + 2x - 6 = 0$. Utilizando o método de Newton-Raphson, determine todas as raízes da equação. Considere $x_0 = 2.5$ e $x_0 = 1 + 2i$ para as raízes complexas.

[4] Tenha em atenção que num polinómio de coeficientes reais, as raízes complexas aparecem sempre aos pares, isto é, se z_1 é raíz de $p_n(x)$ então $\bar{z}_1$ também é raíz de $p_n(x)$.

Capítulo 4

Sistemas de equações

4.1 Breve referência histórica

No século I da era cristã, foi publicado na China um livro intitulado Jiuzhang Suanshu (*Os Nove Capítulos da Arte Matemática*).

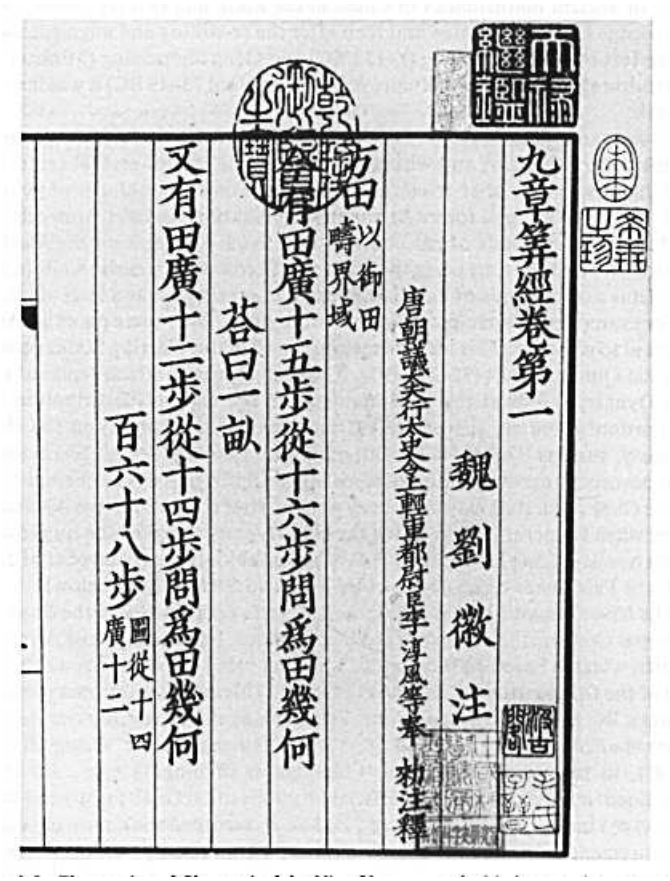

Figura 4.1: Primeira página do primeiro capitulo do livro Nove Capítulos de Arte Matemática

Pela influência que causou em toda a matemática oriental, essa obra é muitas veses comparada aos Elementos de Euclides. Os "Nove Capítulos"são constituídos por 246 problemas de aritmética e geometria mas a sua referência neste capítulo, dedicado à resolução de sistemas lineares, tem a ver com o facto de aí ter sido descrita uma forma sistemática de resolver sistemas lineares com coeficientes positivos. Como curiosidade refira-se que as operações eram efectuadas com o auxílio de pequenos paus dispostos numa folha de papel. Através de manipulações sobre esses paus a técnica proposta era em todo semelhante ao método da decomposição de Gauss (apresentado, somente, no século XIX!). As operações eram efectuadas sobre os coeficientes do sistema o que confere à técnica um estatuto de uma proto-álgebra linear. De notar que os chineses, já desde essa altura, usavam um sistema de numeração de posição, com recurso ao uso de um quadrado em branco para representar o zero. Para cada questão apresentada é apenas apresentada uma solução e não há qualquer referência ao método de resolução. Uma explicação para este facto é que o livro poderá ter sido utilizado apenas como livro de texto.

O livro trata da resolução de problemas utilizando o *Teorema de Pitágoras*, proporcionalidade, cálculo de área do círculo e volume da esfera, determinação de raízes quadradas e cúbicas e resolução de problemas pelo método que é designado por *falsa posição*.

No século XIII o matemático chinês Zhu Shijie publicou uma obra intitulada Suanxue Quimeng ("Introdução à Ciência do Cálculo") onde aperfeiçoou o método de resolução de sistemas lineares proposto nos "Nove Capítulos", não se libertando, contudo, do recurso aos pauzinhos para efectuar as contas. Este livro teve também o mérito de ter sido estudado, muito mais tarde, pelo matemático japonês Seki Takakazu (1642-1708).

Esta página contém tabelas dos coeficientes binomiais e dos números de Bernoulli.

Inspirado por esta obra, Takakazu generalizou a álgebra chinesa libertando-a do recurso aos paus. O cálculo proposto por este ilustre matemático (na obra Kaikendai no Hô) não restringe o número de incógitas e estabelece regras gerais, em vez de resolver casos particulares. Outra contribuição importante de Takakazu foi a introdução da noção de determinante no seu livro Kaifukudai no Hô. A noção de determinante, cuja teoria foi estudada de forma sistemática por Charles Jacobi (1804-1851), precedeu a de matriz que só foi considerada como um ente matemático por William Rowan Hamilton (1805-1856) no seu livro "Lectures on Quaternions". No entanto, a resolução de sistemas de equações lineares já tinha sido considerada por vários autores no ocidente, sendo de destacar o contributo de Carl Friedrich Gauss (1777-1875). Na sua obra "Theoria Motus"(1809) Gauss apresentou uma técnica de resolução de sistemas lineares (surgidos no contexto de um problema de mínimos quadrados) que não é mais do que o método de eliminação que todos conhecemos e que hoje tem o seu nome. Mais tarde, o mesmo Gauss apresentou um processo iterativo para resolver sistemas lineares de grande dimensão, antecipando o procedimento conhecido por método de Gauss-Seidel. Além dos contributos de Gauss, são de salientar os trabalhos de Joseph Louis Lagrange (1736-1813), Pierre Simon Laplace (1749-1827) e Charles Jacobi (1804-1851). Refira-se que Jacobi, baseado nos trabalhos de Gauss sobre mínimos quadrados, efectuou vários trabalhos sobre sistemas lineares tendo também influenciado um seu aluno alem ao Ludwig Seidel. Estes dois matemáticos resolveram vários problemas usando métodos iterativos que foram baptizados com os seus nomes. A teoria de matrizes foi posteriormente aperfeiçoada por Arthur Cayley (1821-1895) e James Joseph Sylvester (1814-1899) (a quem se deve a introdução do termo "matriz"), tendo progredido de forma espectacular até ao início do século XX. Foi para a teoria de matrizes

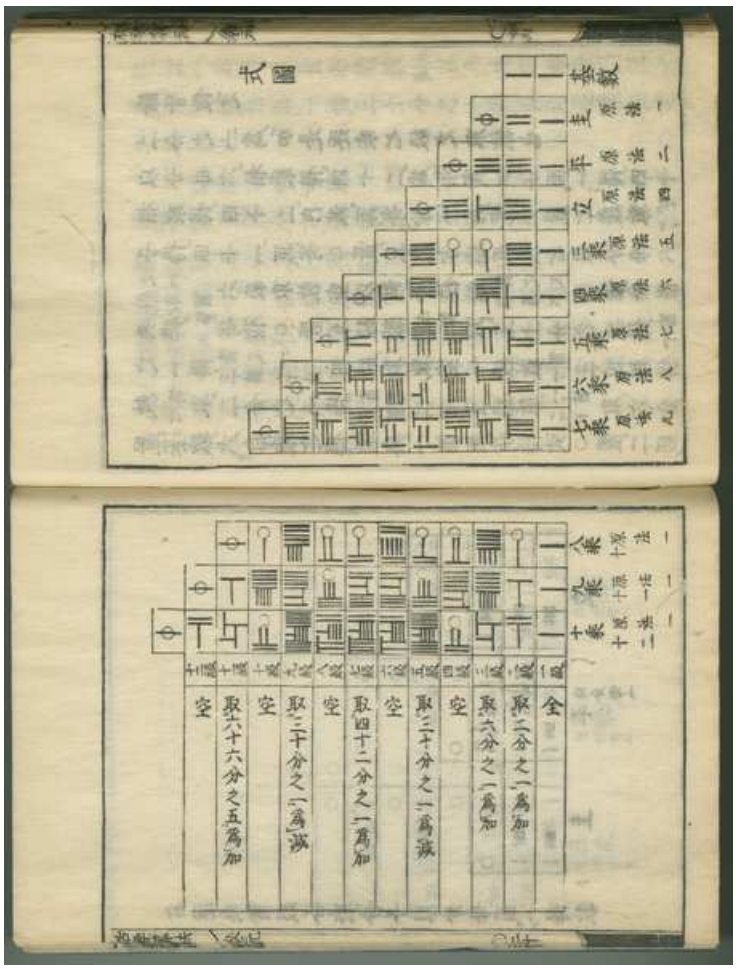

Figura 4.2: Página do livro de Seki

que o físico Heisenberg apelou, em 1925, quando criou a mecânica quântica.

4.2 Resolução de Sistemas de equações

Duas questões fundamentais em Análise Numérica são a resolução de sistemas de equações lineares e o cálculo de valores próprios. Os sistemas de equações lineares ocorrem nos mais diversificados domínios da matemática aplicada. Em problemas concretos podem ocorrer sistemas de grandes dimensões cuja resolução obriga à utilização de meios computacionais. É portanto essencial conhecer algoritmos eficientes de resolução de sistemas lineares.

Nesta secção vamos apresentar métodos que permitem resolver sistemas de equações lineares na forma

$$AX = B, \tag{4.1}$$

onde A é a matriz dos coeficientes, X a coluna com as incógnitas e B a coluna com os

termos independentes. Há dois tipos de métodos a considerar:

1. Os métodos directos, que nos permitem obter a solução do sistema após um número finito de operações. Como exemplo deste tipo de métodos, temos o método de Gauss, o método "explícito", que envolve o cálculo da inversa da matriz A e a aplicação da teoria dos determinantes, via matriz adjunta.

2. Os métodos indirectos são métodos que nos fornecem unicamente aproximações para a solução. Este tipo de métodos é utilizado para a resolução de sistemas de grandes dimensões (por exemplo em que o número de incógnitas é superior a 40) que aparecem em problemas de Engenharia.

   Para resolver o sistema $AX = B$ os métodos iterativos partem de uma aproximação inicial, que será designada por $x^{(0)}$ e constroem uma sucessão de aproximações sucessivas

   $$x^{(1)},\ x^{(2)},\ x^{(3)},\ x^{(4)}, \ldots, x^{(k)}, \ldots$$

   que esperamos seja convergente para a solução exacta x do sistema. No que se segue, suponhamos que A é regular e ainda que $a_{ii} \neq 0$ para $i = \overline{1, n}$.

É possível resolver sistemas de equações utilizando a Regra de Cramer onde, a solução de cada componente do vector de incógnitas é dado pelo quociente entre dois determinantes:

$$x_i = \frac{\left|A^i\right|}{|A|},$$

onde:

a) $|A|$ é o determinante da matriz A;

b) $\left|A^i\right|$ determinante da matriz A em que a $i-$ésima coluna é substituída pelo vector dos termos independentes, B.

Consideremos agora o problema de resolver um sistema de equações de ordem 20, isto é, com 20 equações e com 20 incógnitas. Utilizando a regra de Cramer é necessário:

a) Cálculo de 21 determinantes, cada um de ordem 20;

b) Como vimos no início do semestre, o determinante de uma matriz é definido como a diferença entre as permutações pares e as permutacoes impares. No exemplo a soma tem 20! termos cada qual requerendo 19 multiplicações. Assim, a solução do sistema requer $21 \times 20! \times 19$ multiplicações, além de um número $21 \times 20!$ de somas que será ser considerado.

c) Utilizando um computador com capacidade de 2000 Mflops (2.000.000.000 operações por segundo) o tempo necessário para resolver o problema será de

   $$\frac{21 \times 20! \times 19}{2000000000 \times 3600 \times 24 \times 360} = 15604.55 \text{ anos}$$

   o que é impraticável! O método de Cramer também possui pouca estabilidade numérica (erros de arredondamento excessivos).

4.3 Normas de matrizes e condicionamento

Para além da instabilidade numérica que decorre da propagação de erros de arredondamento, podem também surgir problemas de mau condicionamento. Para podermos analisar convenientemente o problema do condicionamento é necessário introduzir o conceito de norma de um vector e de norma de uma matriz.

Seja E um espaço vectorial real ou complexo, isto é, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$.

Definição 4.3.1 *Chama-se produto interno ou escalar em E a uma aplicação de*

$$\begin{array}{rcl} E \times E & \rightarrow & \mathbb{K} \\ (x, y) & \rightarrow & (x|y) \end{array}$$

tal que

$\mathbf{P}_1$) $(x|x) \geq 0$ *e* $(x|x) = 0$ *sse* $x = 0$.

$\mathbf{P}_2$) $(\alpha x + \beta y|z) = \alpha(x|z) + \beta(y|z)$, *linearidade no primeiro argumento.*

$\mathbf{P}_3$) *Se* $\mathbb{K} = \mathbb{R}$, $(y|x) = (x|y)$ *mas se* $\mathbb{K} = \mathbb{C}$, $(y|x) = \overline{(x|y)}$.

Definição 4.3.2 *Chama-se espaço euclidiano (ou unitário) a um espaço linear de dimensão finita, real (ou complexo) onde se define um produto interno.*

Exemplo 4.3.1 *1. Em $\mathbb{R}^n$, usando as coordenadas dos vectores na base canónica, $x = (x_1, x_2, \ldots, x_n)$ e $y = (y_1, y_2, \ldots, y_n)$, temos um produto interno,*

$$(x|y) := \sum_{i=1}^{n} x_i y_i$$

Em $\mathbb{R}^3$, temos

$$(x|y) = x_1 y_1 + x_2 y_2 + x_3 y_3$$

2. Em $\mathbb{C}^n$ define-se analogamente

$$(x|y) := \sum_{i=1}^{n} x_i \overline{y_i}$$

Definição 4.3.3 *Num espaço com produto interno chama-se norma do vector x ao escalar*

$$||x|| = (x|x)^{\frac{1}{2}}$$

Exemplo 4.3.2 *1. Em $\mathbb{R}^n$, $||x|| = \sqrt{x_1^2 + \cdots + x_n^2}$*

2. Em $\mathbb{C}^n$, $\| \boldsymbol{x} \| = \sqrt{|\boldsymbol{x}_1|^2 + \cdots + |\boldsymbol{x}_n|^2}$

Proposição 4.3.1 *1. Se E é real, o produto interno é uma forma bilinear. Se E é complexo, o produto interno é linear no primeiro argumento e anti-linear (ou conjugado-linear) no segundo argumento:*

$$(x|\alpha y + \beta z) = \bar{\alpha}(x|y) + \bar{\beta}(x|z)$$

2. $(x|0) = (0|x) = 0$

Basta ter em conta que

$$(x|0) = (x|x - x) = (x|x) - (x|x) = 0$$

3. Se

$$(x|y) = 0, \ \ \forall y \in E$$

então $x = 0$. *Da mesma forma se* $(y|x) = 0, \ \ \forall y \in E$ *então* $x = 0$ *Basta ter em conta que* $(x|x) = 0 \rightarrow x = 0$

Proposição 4.3.2 $\mathbf{N_1)}$ $||x|| = 0$ *sse* $x = 0$ *e* $||x|| \geq 0, \ \ x \in E$.

$\mathbf{N_2)}$ $||\lambda x|| = |\lambda|||x||, \ \ \lambda \in \mathbb{K}, x \in E$.

basta ter em conta que $(\lambda x|\lambda x) = \lambda^2(x|x)$.

$\mathbf{N_3)}$ *A desigualdade triangular,*

$$||x + y|| \leq ||x|| + ||y||$$

$\mathbf{N_4)}$ *A desigualdade de Cauchy-Schwarz*

$$|(x|y)| \leq ||x||||y||$$

Definição 4.3.4 *Seja* $\| \cdot \|$ *uma norma vectorial em* $\mathbb{R}^n$. *A norma da matriz A induzida pela norma vectorial é:*

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{4.2}$$

Vamos de seguida apresentar algumas normas de vectores e matrizes.

1. Normas de vectores
   a) $\|x\|_1 = |x_1| + \cdots + |x_n|$,
   b) $\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$,
   c) $\|x\|_\infty = \max\limits_{i=1,\ldots,n} \{|x_1|, \ldots, |x_n|\}$

2. Como exemplo de normas de matrizes podemos apresentar as seguintes:
   a) $\|A\|_1 = \max\limits_{j=1,\ldots,m} \sum\limits_{i=1}^{n} |a_{ij}| \Leftrightarrow$ A maior soma absoluta das colunas;
   b) $\|A\|_\infty = \max\limits_{i=1,\ldots,n} \sum\limits_{j=1}^{m} |a_{ij}| \Leftrightarrow$ A maior soma absoluta das linhas;
   c) $\|A\|_E = \sqrt{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} a_{ij}^2} \Leftrightarrow$ Norma Euclidiana (também é designada por norma de Frobenius).

Exemplo 4.3.1 *Seja* $A = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$. *Determine* $\|A\|_1$, $\|A\|_E$ *e* $\|A\|_\infty$.

Resolução 4.3.1 *Se* $A = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$ *então:*

a) $\|A\|_1 = \{3 + |-1|\,; 2 + 0\} = 4$;

b) $\|A\|_\infty = \{3 + 2; |-1| + 0\} = 5$;

c) $\|A\|_E = \sqrt{3^2 + 2^2 + (-1)^2 + 0^2} = \sqrt{9 + 4 + 1} = \sqrt{14}$.

■

Definição 4.3.5 *Designamos por raio espectral,* ρ *de uma matriz* A *ao valor*

$$\rho(A) = \max_{i=1,2,\ldots,n} |\lambda_i| \tag{4.3}$$

onde λ_i, $(i = 1, 2, \ldots, n)$ *são os valores próprios de* A.

Teorema 4.3.1 *Seja* A *uma matriz quadrada. Então,*

a) Para qualquer norma matricial $\|\cdot\|$ *temos*

$$\rho(A) \leq \|A\|,$$

b) Para qualquer $\varepsilon > 0$, *existe sempre uma norma induzida* $\|\cdot\|$, *tal que :*

$$\|A\| \leq \rho(A) + \varepsilon.$$

Ou seja, o raio espectral é o ínfimo do conjunto das normas induzidas de uma matriz.

Ao resolver um sistema linear

$$AX = B$$

podemos ter problemas de condicionamento e de estabilidade numérica. Os problemas de estabilidade numérica estão relacionados com o algoritmo que utilizamos para resolver o sistema. Por exemplo, para evitar os problemas de instabilidade numérica, é habitual considerar o método de Gauss com pesquisa de pivot. No entanto, se o problema for mal condicionado, essas técnicas de pesquisa de pivot deixam de ser úteis, já que um problema mal condicionado será sempre numericamente instável. Portanto, interessa identificar quais os sistemas que nos podem trazer problemas de condicionamento.

Supondo que nos era dado, não o vector B mas uma sua aproximação $\tilde{B}$. Vamos analisar a influência desse erro nos resultados obtidos, uma vez que em vez do valor exacto obtemos um valor aproximado $\tilde{X}$, que é a solução do sistema aproximado

$$A\tilde{X} = \tilde{B}. \tag{4.4}$$

As normas previamente apresentadas permitem-nos estabelecer uma medida de comparação entre os erros vectoriais, definindo-se os erros do mesmo modo que no caso escalar,

a) Erro Absoluto de $\tilde{X}$: $\|E_{aX}\| = \left\|X - \tilde{X}\right\|$;

b) Erro Relativo de E_{rX}: $\|E_{rX}\| = \dfrac{\left\|X - \tilde{X}\right\|}{\|X\|}$,

relativamente a uma norma induzida.

Para estabelecermos a relação entre os erros relativos dos dados e os erros relativos dos resultados é importante estabelecer um conceito que está relacionado com a norma de matrizes: o número de condição.

Definição 4.3.6 (Ver definição 2.54) *Designa-se por número de condição de uma matriz A relativamente à norma $\|\cdot\|$ ao número:*

$$\kappa = CondA = \|A\| \left\|A^{-1}\right\| \geq 1. \tag{4.5}$$

Teorema 4.3.2 *Seja A uma matriz invertível e seja δA uma "perturbação"da matriz A, tal que*

$$\|\delta A\| < \frac{1}{\|A^{-1}\|}.$$

Então,

$$\frac{\|E_r x\|}{\|x\|} \leq \frac{\kappa}{1 - \kappa \frac{\|E_{rA}\|}{\|A\|}} \left(\frac{\|E_{rA}\|}{\|A\|} + \frac{\|E_{rB}\|}{\|B\|} \right), \tag{4.6}$$

onde $\kappa = cond(A) = \|A\| \times \left\|A^{-1}\right\|$ é chamado o número de condição da matriz A.

Nota 4.3.1 *Quando o número de condição é grande, o problema de resolução do sistema $AX = B$ é mal condicionado. Pois pequenas variações nos elementos de A e B podem causar grande variação relativa na solução.*

4.4 Métodos directos

Como exemplo de métodos directos, podemos referir o método de Gauss, os métodos que envolvem a decomposição da matriz dos coeficientes como um produto de várias matrizes, (por exemplo a decomposição em LU), a utilização da matriz inversa. Estes métodos são pouco práticos para a resolução de sistemas de equações lineares que aparecem em problemas de Engenharia, pois o número de equações é muito elevado ($n \geq 40$). Por exemplo, quando o determinante da matriz dos coeficientes é um número muito próximo de zero, os métodos directos **não funcionam bem** (o facto de se dividir um número por outro muito próximo de zero, pode provocar erros de arredondamento ...).

O método de Gauss é baseado numa "redução"do sistema dado a um sistema equivalente cuja matriz dos coeficientes é triangular superior. Tal sistema pode então ser resolvido por substituição inversa. Tal redução é feita efectuando operações sobre as filas da matriz dos coeficientes. Para evitar a propagação de erros, aquando da aplicação deste método é habitual escolher-se o pivot (elementos da diagonal principal). As técnicas para a escolha do pivot são as seguintes:

a) Na chamada escolha parcial de pivot, no início do passo k, é escolhido o pivot a_{pk} tal que

$$|a_{pk}| = \max_{k \leq i \leq n} (|a_{ij}|) .$$

Se $p \neq k$ as linhas p e k são trocadas entre si.

b) Na chamada escolha total de pivot, no início do passo k, é escolhido o pivot a_{pk} tal que

$$|a_{pk}| = \max_{k \leq i, k \leq j \leq n} (|a_{ij}|) .$$

Se $p \neq k$ as linhas p e k são trocadas entre si. Se $r \neq k$ as colunas r e k são trocadas entre si. (mas de modo a obter um sistema equivalente). A acumulação de erros de arredondamento é menor quando e utiliza a escolha total de pivot. No entanto, o esforço computacional é superior.

Exemplo 4.4.1 *Considere o sistema $AX = B$ com*

$$A = \begin{bmatrix} -10^{-5} & 1 \\ 2 & 1 \end{bmatrix} \quad e \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

cuja solução exacta é $x = -0.4999975$ e $y = 0.999995$. Resolva o sistema pelo método de Gauss numa máquina de calcular com 4 casas decimais, escrevendo os números na forma normalizada, $0.****\times 10^{\pm e}$ e:

a) Sem escolha parcial de pivot;

b) Com escolha parcial de pivot.

Resolução 4.4.1 *Sem escolha parcial de pivot, temos*

$$\left[\begin{array}{cc|c} -0.1 \times 10^{-4} & 0.1 \times 10 & 0.1 \times 10 \\ 0.2 \times 20 & 0.1 \times 10 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} -0.1 \times 10^{-4} & 0.1 \times 10 & 0.1 \times 10 \\ 0.2 \times 20 & 0.2 \times 10^{6} & 0.2 \times 10^{6} \end{array}\right]$$

$$\Rightarrow \left\{\begin{array}{rcl} x & = & 0 \\ y & = & 1 \end{array}\right.$$

Com escolha parcial de pivot,

$$\left[\begin{array}{cc|c} 0.2 \times 20 & 0.1 \times 10 & 0 \\ -0.1 \times 10^{-4} & 0.1 \times 10 & 0.1 \times 10 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 0.2 \times 20 & 0.1 \times 10 & 0 \\ & 0.1 \times 10 & 0.1 \times 10 \end{array}\right]$$

$$\Rightarrow \left\{\begin{array}{rcl} x & = & -0.5 \\ y & = & 1 \end{array}\right.$$

Verifica-se que com escolha parcial de pivot, os resultados são mais precisos.

■

Definição 4.4.1 *Uma matriz $A \in M_{(n,n)}(\mathbb{R})$ diz-se diagonal dominante se*

$$|a_{ii}| \geq \sum_{1\leq j\leq, j\neq i} |a_{ij}|\,, \quad j = \overline{1,n} \tag{4.7}$$

e diz-se estritamente diagonal dominante se

$$|a_{ii}| > \sum_{1\leq j\leq, j\neq i} |a_{ij}|\,, \quad j = \overline{1,n}. \tag{4.8}$$

Nota 4.4.1 *Num sistema $AX = B$ em que A é estritamente diagonal dominante, não é necessário efectuar a escolha parcial de pivot pois ela está garantida à partida.*

O método de eliminação de Gauss consiste em condensar a matriz dos coeficientes até se obter uma matriz triangular superior e depois, procedendo à substituição inversa, determinar os valores das incógnitas pela ordem x_n, x_{n-1}, ..., x_1.

4.4.1 Sistemas triangulares

Consideremos um sistema de n equações com n incógnitas cuja matriz dos coeficientes é triangular superior invertível.

$$\begin{cases} a_{xx}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{nn}x_n &= b_n \end{cases} \tag{4.9}$$

Da última equação de (4.28) obtemos

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

Substituindo o valor de x_n na penúltima equação segue-se que

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

e assim sucessivamente até se obter

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

a) $x_n = \dfrac{b_n}{a_{nn}}$,

b) $x_i = \dfrac{1}{a_{ii}}\left(b_i - \displaystyle\sum_{j=i+1}^{n} a_{ij}x_j\right)$

Este método é designado por método de substituição inversa (a matriz dos coeficientes é triangular superior). Quando a matriz dos coeficientes é triangular inferior invertível, o método é designado por método de substituição directa. (em primeiro lugar obtemos o valor de x_1).

4.4.2 Sistemas tridiagonais

Consideremos o sistema $AX = B$ em que $A = [a_{ij}]$ com $a_{ij} = 0$ para $|i - j| > 2$, isto é,

$$\begin{bmatrix} d_1 & u_1 & & & & \\ l_2 & d_2 & u_2 & & & \\ & l_3 & d_3 & u_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & & & u_{n-1} \\ & & & & l_n & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \tag{4.10}$$

Gauss, desenvolveu um algoritmo para calcular a solução de sistemas triangulares, que passamos a indicar:

para k de 1 até $n-1$ fazer

$m \leftarrow \dfrac{l_{k+1}}{d_k}$

$d_{k+1} \leftarrow d_{k+1} - mu_k$

$b_{k+1} \rightarrow b_{k+1} - mb_k$

$x_n \rightarrow \dfrac{b_n}{d_n}$

para k de $n-1$ até 1 fazer

$x_k \leftarrow \dfrac{(b_k - u_k x_{k+1})}{d_k}$

Exemplo 4.4.2 *Aplique o algoritmo de Gauss ao sistema tridiagonal $AX = B$ com*

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 2 \end{bmatrix} \text{ e } B = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

Resolução 4.4.2 *Como $n = 3$, $k = 1, 2$.*

Para $k = 1$ temos

$$\begin{aligned} m &= \frac{l_2}{d_1} = \frac{1}{2} \\ d_2 &= d_2 - mu_1 = 4 - \frac{1}{2} \times 1 = \frac{7}{2} \\ b_2 &= b_2 - mb_1 = 1 - \frac{1}{2} \times 0 = 1 \end{aligned}$$

Para $k = 2$,

$$\begin{aligned} m &= \frac{l_3}{d_2} = \frac{1}{\frac{7}{2}} \\ d_2 &= d_2 - mu_1 = 4 - \frac{1}{2} \times 1 = \frac{7}{2} \\ b_2 &= b_2 - mb_1 = 1 - \frac{1}{2} \times 0 = 1 \end{aligned}$$

Portanto,

$$x_3 = \frac{b_3}{d_3} = 1, \;\; x_2 = \frac{b_2 - u_2 x_3}{d_2} = \frac{1 - 1 \times 1}{4} = 0, \; x_1 = \frac{b_1 - u_1 x_2}{d_1} = \frac{0 - 1 \times 0}{2} = 0.$$

■

4.4.3 Métodos de eliminação compacta

Em muitos problemas surge o problema de resolver vários sistemas de equações lineares em que a matriz dos coeficientes é comum. A única diferença reside na coluna dos termos independentes. Os métodos que de seguida descrevemos têm especial aplicação para resolve este tipo de problemas.

Vamos supor que no sistema de equações lineares

$$Ax = B \tag{4.11}$$

a matriz dos coeficientes se pode escrever como o produto de duas matrizes,

$$A = LU, \tag{4.12}$$

onde L é uma matiz triangular inferior, cujos elementos da diagonal principal são todos iguais a 1 e U uma matriz triangular superior. Assim, substituindo (4.12) em (4.11) temos

$$LUx = B. \tag{4.13}$$

Designando

$$Ux = y, \tag{4.14}$$

vem, substituindo em (4.13), obtemos

$$Ly = B. \tag{4.15}$$

Assim, calculada a decomposição LU da matriz A, o sistema (4.15) é resolvido por substituição directa e, calculado y, o sistema (4.15) é resolvido por substituição inversa para obter x.

Para resolver um novo sistema $Ax = C$ é apenas necessário resolver $Ly = C$ por substituição directa e $Uz = y$ por substituição inversa.

Teorema 4.4.1 *Seja A uma matriz de ordem n e A_k o bloco formado pelas primeiras k linhas e pelas primeiras k colunas de A. Suponha-se que* $\det A_k \neq 0$*, para $k = 1, 2, \ldots, n-1$. Então, existe uma e uma única matriz triangular inferior L, cujos elementos da diagonal são todos iguais a 1, e uma matriz triangular U tal que $A = LU$.*

Os elementos das matrizes L e U são dados por

- $u_{1j} = a_{1j}, \;\; (j = 1, 2, \ldots, n)$
- $l_{i1} = \dfrac{a_{i1}}{u_{11}}, \;\; (j = 2, 3, \ldots, n)$

- $u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}, \ \ (i = 2, 3, \ldots, n; j = i, i+1, \ldots, n)$
- $l_{ji} = \dfrac{a_{ji} - \sum_{k=1}^{i-1} l_{jk}u_{ki}}{u_{ii}}, \ \ (i = 2, 3, \ldots, n-1; j = i+1, i+2, \ldots, n)$

Nota 4.4.2 *Em resumo, o sistema $AX = B$ pode ser resolvido do seguinte modo:*

$$AX = B \Leftrightarrow LUX = B \leftrightarrow \left\{ \begin{array}{rcl} UX & = & Y \\ LY & = & B \end{array} \right. ,$$

depois,

i) Resolvendo em primeiro lugar o sistema $LY = B$, calcula-se Y (por substituição directa);

ii) Resolvendo em seguida o sistema $UX = Y$, calcula-se X (por substituição inversa).

Exemplo 4.4.3 *Ao calcular a decomposição LU da matriz*

$$A = \left[\begin{array}{cccc} 2 & 2 & 1 & 4 \\ 1 & -3 & 2 & 3 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 2 \end{array} \right] \tag{4.16}$$

obtem-se

$$L = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ -0.5 & -0.5 & 1 & 0 \\ 0.5 & 0.5 & -1 & 1 \end{array} \right] \ e \ U = \left[\begin{array}{cccc} 2 & 2 & 1 & 4 \\ 0 & -4 & 1.5 & 1 \\ 0 & 0 & 0.25 & 1.5 \\ 0 & 0 & 0 & 1 \end{array} \right]$$

pois,

1*) primeira linha de U;*

$$u_{11} = 2, \ \ u_{12} = 2, \ \ u_{13} = 1, \ \ u_{14} = 4$$

2*) primeira coluna de L;*

$$l_{21} = \frac{1}{2}, \ \ l_{31} = -\frac{1}{2}, \ \ l_{41} = \frac{1}{2}, \ \ u_{14} = 4$$

3*) segunda linha de U;*

$$u_{22} = -3 - \frac{1}{2} \times 2 = -4, \ \ u_{23} = 2 - \frac{1}{2} \times 1 = \frac{3}{2}, \ \ u_{24} = 3 - \frac{1}{2} \times 4 = 1$$

4*) segunda coluna de L,*

$$l_{32} = \frac{1 - \left(-\frac{1}{2}\right) \times 2}{-4} = -\frac{1}{2}, \quad l_{42} = \frac{-1 - \frac{1}{2} \times 2}{-4} = \frac{1}{2}$$

5) *terceira linha de* U,

$$u_{33} = -1 - \left(-\frac{1}{2} \times 1 - \frac{1}{2} \times \frac{3}{2}\right) = \frac{1}{4}, \quad u_{34} = -1 - \left(-\frac{1}{2} \times 4 - \frac{1}{2} \times 1\right) = \frac{3}{2}$$

6) *terceira coluna de* L,

$$l_{43} = \frac{1 - \left(\frac{1}{2} \times 1 + \frac{1}{2} \times \frac{3}{2}\right)}{\frac{1}{4}} = -1$$

7) *última linha de* U;

$$u_{44} = 2 - \left(\frac{1}{2} \times 4 + \frac{1}{2} \times 1 - \frac{1}{2} \times \frac{3}{2}\right) = 1.$$

Exemplo 4.4.4 *Resolva, utilizando a decomposição em* LU *o sistema* $AX = B$ *onde* A *é a matriz (4.16) e* $B = [2 \ 1 \ 0 \ 0]^T$.

Resolução 4.4.3 *De acordo com (4.4.2) temos*

1)

$$LY = B \Leftrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 2 \\ 0.5 & 1 & 0 & 0 & 1 \\ -0.5 & -0.5 & 1 & 0 & 0 \\ 0.5 & 0.5 & -1 & 1 & 0 \end{array}\right] \Leftrightarrow \left\{\begin{array}{rcl} y_1 & = & 2 \\ y_2 & = & 1 - 1 = 0 \\ y_3 & = & 1 + 0 = 1 \\ y_4 & = & -1 + 0 + 1 = 0 \end{array}\right.$$

2)

$$UX = Y \Leftrightarrow U = \left[\begin{array}{cccc|c} 2 & 2 & 1 & 4 & 2 \\ 0 & -4 & 1.5 & 1 & 0 \\ 0 & 0 & 0.25 & 1.5 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

$$\Leftrightarrow \left\{\begin{array}{rcl} 2x_1 + 3 + 4 & = & 2 \Leftrightarrow x_1 = -\frac{5}{2} \\ -4x_2 + 6 & = & 0 \Leftrightarrow x_2 = \frac{3}{2} \\ \frac{1}{4}x_3 + 0 & = & 1 \Leftrightarrow x_3 = 4 \\ x_4 & = & 0 \end{array}\right.$$

■

Nota 4.4.3 *Utilizando a decomposição em* $A = LU$, *podemos facilmente calcular o determinante de* A *e a inversa de* A. *Assim, tendo em conta que*

$$\det(AB) = \det(A) \det(B),$$

temos que

$$\begin{aligned}\det(A) &= \det(LU) = \det(L)\det(U) = \\ &= 1 \times \det(U) = 2 \times (-4) \times \frac{1}{4} \times 1 = -2. \end{aligned} \tag{4.17}$$

Para calcular a inversa, A^{-1}, procedemos do seguinte modo

1) Resolver a equação matricial, $LY = I_4$,

$$LY = I \Rightarrow Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.5 & 1 & 0 & 0 \\ 0.75 & -0.5 & 1 & 0 \\ 0.5 & -1 & 1 & 1 \end{bmatrix}$$

2) Resolver a equação $UX = Y$,

$$UX = Y \Rightarrow X = \begin{bmatrix} -0.75 & -1 & -0.5 & 3 \\ 0.25 & 1 & -0.5 & -2 \\ 0 & 4 & -2 & -6 \\ 0.5 & -1 & 1 & 1 \end{bmatrix} = A^{-1}$$

Exemplo 4.4.5 *Resolva o sistema de equações lineares*

$$\begin{cases} -x_1 + 2x_2 + x_3 &= 0 \\ 8x_2 + 6x_3 &= 10 \\ -2x_1 + 5x_3 &= -11 \end{cases} \tag{4.18}$$

utilizando a decomposição em LU.

Resolução 4.4.4 *A decomposição em LU da matriz A é*

$$A \equiv \begin{bmatrix} -1 & 2 & 1 \\ 0 & 8 & 6 \\ -2 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 & 1 \\ 0 & 8 & 6 \\ 0 & 0 & 6 \end{bmatrix} \equiv LU$$

Logo, o sistema $Ax = B$ fica

$$\begin{aligned} Ax &\equiv \begin{bmatrix} -1 & 2 & 1 \\ 0 & 8 & 6 \\ -2 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= (LU)\, x \\ &= \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 & 1 \\ 0 & 8 & 6 \\ 0 & 0 & 6 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ -11 \end{bmatrix}. \end{aligned}$$

Então,

i) Seja $y = Ux$ e vamos resolver o sistema $Ly = b$, isto é,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ -11 \end{bmatrix}$$

cuja solução é $[y_1 \ y_2 \ y_3]^T = [0 \ 10 \ -6]^T$. A solução deste sistema foi obtida utilizando a substituição directa.

ii) Depois, é necessário resolver o sistema $Ux = y$ sendo $y = [0 \ 10 \ -6]^T$, isto é,

$$\begin{bmatrix} -1 & 2 & 1 \\ 0 & 8 & 6 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ -6 \end{bmatrix}$$

que utilizando a substituição inversa permite obter a solução

$$x = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} \tag{4.19}$$

■

Há outro modo de decompor uma matriz. A chamada decomposição de Cholesky. Neste caso, a matriz A é decomposta como produto de uma matriz triangular superior U com a sua transposta, isto é, $A = U^T U$.

$$u_{11} = \sqrt{a_{11}}, \quad u_{1j} = \frac{a_{1j}}{u_{11}}, \quad u_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} u_{kj}^2}, \qquad j = 2, \ldots, n$$

$$u_{ij} = \frac{1}{u_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} u_{ki} u_{kj} \right), \quad i = 2, \ldots, \ n, \quad j = i+1, \ldots, \ n \qquad u_{ij} = 0, \ i > j$$

4.4.4 Eliminação de Gauss e Decomposição LU

A eliminação de Gauss pode ser usada para decompor uma matriz dos coeficientes A, em duas matrizes L e U onde L é uma matriz triangular inferior e U é uma matriz triangular superior.

Consideremos uma matriz quadra de ordem 3,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Através de passos de eliminação, podemos reduzir a matriz original dos coeficientes, a matriz A, numa matriz U,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Leftrightarrow U = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix}$$

O primeiro passo da eleiminação de Gauss consisite em multiplicar a primeira linha por $f_{21} = \frac{a_{21}}{a_{11}}$ e subtrair o resultado à segunda linha, eliminando o elemento a_{21}. Igualmente, multiplicamos a primeira linha por $f_{31} = \frac{a_{31}}{a_{11}}$ de modo a eliminar a_{31}. O passo final consiste em multiplicar o elemento que está na posição $(2,2)$ por $f_{32} = \frac{a'_{32}}{a'_{22}}$ e subtrai-lo à terceira linha de modo a elimiar o elemento a'_{32}. Do ponto de vista matricial obtemos

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \longrightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & a'_{32} & a'_{33} \end{bmatrix} \longrightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix} \tag{4.20}$$

A matriz L é uma matriz triângular inferior, cujos elementos da diagonal principal são iguais a 1 e os restantes factores são os factores f_{21}, f_{31} e f_{32}, isto é,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ f_{21} & 1 & 0 \\ f_{31} & f_{32} & 1 \end{bmatrix}$$

Multiplicando as matrizes L e U, obtemos a matriz original A.

A eliminação de Gauss representa a decomposição em LU da matriz A. Vamos de seguida apresentar um exemplo com um sistema de equações com três equações a três incógnitas.

Exemplo 4.4.6 *Utilizando a decomposição em $A = LU$, resolva o sistema de equações lineares*

$$\begin{cases} 2x + y + 4z &= 2 \\ 6x + y &= -10 \\ -x + 2y - 10z &= -4 \end{cases}$$

Resolução 4.4.5 *Representando o sistema na forma matricial $AX = B$, na forma*

$$\begin{bmatrix} 2 & 1 & 4 \\ 6 & 1 & 0 \\ -1 & 2 & -10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -10 \\ -4 \end{bmatrix}$$

Aplicando o método de eliminação de Gauss, obtemos o sistema $UX = D$, da forma

$$\begin{bmatrix} 2 & 1 & 4 \\ 0 & -2 & -12 \\ 0 & 0 & -23 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -16 \\ -23 \end{bmatrix}$$

O primeiro passo da eliminação consistiu em subtrair à segunda linha a primeira multiplicada por 3; no segundo passo, multiplicou-se a segunda linha por $-\frac{1}{2}$ e subtraiu-se à

terceira linha: o terceiro passo consiste em subtrair à terceira linha a segunda multiplicada por $-\frac{5}{4}$*. Portanto, a matriz* L *é*

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -\frac{1}{2} & -\frac{5}{4} & 1 \end{bmatrix}$$

Note-se que os elementos abaixo da diagonal principal são exactamente os multiplicadores $3, -\frac{1}{2}$ *e* $-\frac{5}{4}$ *utilizados no processo de eliminação de Gauss. Facilmente se mostra que* $LU = A$*, isto é,*

$$LU = A \Leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -\frac{1}{2} & -\frac{5}{4} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 4 \\ 0 & -2 & -12 \\ 0 & 0 & -23 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 4 \\ 6 & 1 & 0 \\ -1 & 2 & -10 \end{bmatrix} \tag{4.21}$$

e que

$$LD = B \Leftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -\frac{1}{2} & -\frac{5}{4} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -16 \\ -23 \end{bmatrix} = \begin{bmatrix} 2 \\ -10 \\ -4 \end{bmatrix} \tag{4.22}$$

■

4.5 Métodos iterativos

Os métodos a seguir descritos são apropriados para sistemas de grande dimensão, cuja matriz dos coeficientes é esparsa, isto é, que tem muitos zeros.

Seja

$$AX = B \tag{4.23}$$

um sistema de Cramer com n equações e n incógnitas. A matriz A pode ser escrita na forma

$$A = M - N, \tag{4.24}$$

sendo M e N matrizes de ordem n e M é invertível. Substituindo no sistema temos

$$\begin{aligned} (M - N)\, X &= B && (4.25) \\ &\text{ou} \\ Mx &= Nx + B && (4.26) \\ &\text{donde} \\ x &= M^{-1}(Nx + B) && (4.27) \end{aligned}$$

Assim, a solução do sistema é o ponto fixo de (4.27).

Teorema 4.5.1 *Se* $\left\|M^{-1}N\right\| < 1$ *a sucessão definida pela iteração*

$$x_{k+1} = M^{-1}(Nx_k + B), \quad (k = 0, 1, \ldots), \tag{4.28}$$

converge para o ponto fixo de (4.27) qualquer se seja $x_0 \in \mathbb{R}^n$*.*

Os critérios de paragem habitualmente utilizados são

i) Critério do erro absoluto: $\left\|x^{(n)} - x^{n-1}\right\| \leq \varepsilon$

ii) Critério do erro relativo: $\left\|x^{(n)} - x^{n-1}\right\| \leq \varepsilon \left\|x^n\right\|$

iii) Critério do valor da função: $\|F(x^n)\| \leq \varepsilon_1 < \varepsilon$

iv) Critério do número máximo de iteracções: $n = nmax$.

4.5.1 Jacobi $A = D + (L + U)$

Este método foi desenvolvido por Carl Jacobi (1804-1851). Neste método, a partir da matriz dos coeficientes $A = [a_{ij}]_{i,j=\overline{1,n}}$ vamos definir três matrizes $D = [d_{ij}]$ uma matriz diagonal, $L = [l_{ij}]$ uma matriz estritamente triangular inferior e $L = [l_{ij}]$ uma matriz estritamente triangular superior tais que

$$d_{ij} = \begin{cases} a_{ij} & \text{se} \quad i = j \\ \\ 0 & \text{se} \quad i \neq j \end{cases}, \quad l_{ij} = \begin{cases} a_{ij} & \text{se} \quad i > j \\ \\ 0 & \text{se} \quad i \leq j \end{cases}, \tag{4.29}$$

$$u_{ij} = \begin{cases} a_{ij} & \text{se} \quad i < j \\ \\ 0 & \text{se} \quad i \geq j \end{cases} \tag{4.30}$$

Então

$$A = D + (L + U). \tag{4.31}$$

$$x_i^{(k+1)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)}, \qquad i = 1, \ldots, \ n \quad k = 0, 1, \ldots$$

$$x^{(k+1)} = -D^{-1}(L + U)x^{(k)} + D^{-1}B \qquad M = -D^{-1}(L + U)$$

Naturalmente, temos que supor que a matriz D é invertível. Pelo teorema 4.5.1 basta garantir que

$$\left\|D^{-1}(L + U)\right\| < 1 \tag{4.32}$$

para que o limite da sucessão

$$x_{k+1} = D^{-1}\left[-(L + U)x^k + B\right], \quad (k = 0, 1, \ldots) \tag{4.33}$$

seja a solução do sistema $AX = B$.

Exemplo 4.5.1 *Consideremos o sistema de equações lineares $AX = B$ onde*

$$A = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & 0 \\ 1 & 0 & -2 \end{bmatrix}, \ X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \ \ e \ B = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}.$$

Resolução 4.5.1 *Para este problema temos as matrizes*

$$L = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \; U = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \; e \; D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \tag{4.34}$$

e

$$-D^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \; e \; -D^{-1}(L+U) = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}. \tag{4.35}$$

Portanto, obtemos a formula recursiva

$$x^{k+1} = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} x^k + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}. \tag{4.36}$$

Vamos considerar como aproximação inicial $x_0 = [0 \;\; 0 \;\; 0]^T$. *Temos então:*

$$X_1 = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.6667 \\ 1.5 \end{bmatrix}.$$

$$X_2 = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0.6667 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0.83333 \\ 0.6667 \\ 1.5 \end{bmatrix}.$$

$$X_3 = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.83333 \\ 0.6667 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0.83333 \\ 0.94444 \\ 1.9167 \end{bmatrix}.$$

$$X_4 = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.83333 \\ 0.94444 \\ 1.9167 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0.97226 \\ 0.94444 \\ 1.9167 \end{bmatrix}.$$

$$X_5 = \begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.97226 \\ 0.94444 \\ 1.9167 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0.97226 \\ 0.99075 \\ 1.9861 \end{bmatrix}.$$

Assim, continuando podemos chegar a solução exacta da equação dada por:

$$X = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

■

Nota 4.5.1 *Devemos salientar que o método converge apesar de que a matriz dos coeficientes não é diagonal dominante. No entanto, a matriz* M, *a matriz de recorrência, verifica a condição:*

$$\lim_{n\to\infty}\left(\begin{bmatrix} 0 & -1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}\right)^n = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.37}$$

Esta condição garante a convergência do método. Para mais detalhes ver secção 4.5.3 e teorema 4.5.2.

Exemplo 4.5.2 *Utilizando o método de Jacobi, determine a solução do sistema de equações lineares $AX = B$ onde*

$$A = \begin{bmatrix} 7 & -2 & 1 & 2 \\ 2 & 8 & 3 & 1 \\ -1 & 0 & 5 & 2 \\ 0 & 2 & -1 & 4 \end{bmatrix} \quad e\ B = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 4 \end{bmatrix} \tag{4.38}$$

Resolução 4.5.2 *Em primeiro lugar, facilmente se verifica que a matriz é diagonal dominante, pois*

$$7 > |-2| + 1 + 2,\ \ 8 > 2 + 3 + 1,\ \ 5 > |-1| + 0 + 2 \quad e\ 4 > 0 + 2 + |-1|.$$

Portanto, o método de Jacobi converge.

As matrizes L, D e U são respectivamente

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 \end{bmatrix},\ \ D = \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

e

$$U = \begin{bmatrix} 0 & -2 & 1 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Temos portanto,

$$D^{-1} = \begin{bmatrix} \frac{1}{7} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}\ e\ L + U = \begin{bmatrix} 0 & -2 & 1 & 2 \\ 2 & 0 & 3 & 1 \\ -1 & 0 & 0 & 2 \\ 0 & 2 & -1 & 0 \end{bmatrix}.$$

Logo, obtemos o processo iterativo

$$\begin{bmatrix} x^{(n+1)} \\ y^{(n+1)} \\ z^{(n+1)} \\ t^{(n+1)} \end{bmatrix} = \begin{bmatrix} 0 & \frac{2}{7} & -\frac{1}{7} & -\frac{2}{7} \\ -\frac{1}{4} & 0 & -\frac{3}{8} & -\frac{1}{8} \\ \frac{1}{5} & 0 & 0 & -\frac{2}{5} \\ 0 & -\frac{1}{2} & +\frac{1}{4} & 0 \end{bmatrix} \begin{bmatrix} x^{(n)} \\ y^{(n)} \\ z^{(n)} \\ t^{(n)} \end{bmatrix} + \begin{bmatrix} -\frac{3}{7} \\ \frac{1}{4} \\ -1 \\ -1 \end{bmatrix} \tag{4.39}$$

Supondo $X^{(0)} = [0 \ \ -1 \ \ 1 \ \ 1]^T$ *temos,*

$$\begin{aligned} X^{(0)} &= [0 \ \ -1 \ \ 1 \ \ 1]^T \\ X^{(1)} &= [-0.285714 \ \ -0.75 \ \ 0.6 \ \ 1.75]^T \\ X^{(2)} &= [-0.371429 \ \ -0.622321 \ \ 0.242857 \ \ 1.525]^T \\ X^{(3)} &= [-0.219643 \ \ -0.438839 \ \ 0.315714 \ \ 1.37188]^T \\ X^{(4)} &= [-0.133878 \ \ -0.484967 \ \ 0.407321 \ \ 1.29835]^T \\ X^{(5)} &= [-0.139136 \ \ -0.53157 \ \ 0.453885 \ \ 1.34431]^T \\ X^{(6)} &= [-0.172236 \ \ -0.553462 \ \ 0.434447 \ \ 1.37926]^T \\ &\vdots \\ X^{(20)} &= [-0.175171 \ \ -0.533783 \ \ 0.416545 \ \ 1.37103]^T \end{aligned}$$

■

Exercício 4.5.1 *Utilizando o método de Jacobi, determine a solução do sistema de equações lineares* $AX = B$ *onde*

$$A = \begin{bmatrix} 2 & 8 & 3 & 1 \\ 7 & -2 & 1 & 2 \\ -1 & 0 & 5 & 2 \\ 0 & 2 & -1 & 4 \end{bmatrix} \quad e \ B = \begin{bmatrix} -2 \\ 3 \\ 5 \\ 4 \end{bmatrix} \tag{4.40}$$

4.5.2 Gauss-Seidel $A = (D + L) + U$

Este método é atribuído a Johann Carl Friedrich Gauss (1777-1855) e Philipp Ludwig von Seidel (1821-1896).

Tal como anteriormente, vamos considerar as matrizes D, L e U. Tem-se

$$A = (D + L) + U.$$

Fazendo a escolha

$$M = D + L, \ \ N = -U,$$

obtemos o chamado método de Gauss Seidel.

Sendo

$$(D+L)\,x^{k+1} = -Ux^k + B, \;\; (k=0,1,\ldots)$$

obtemos

$$x^{k+1} = D^{-1}\left[-Lx^{k+1} - Ux^k + B\right], \;\; (k=0,1,\ldots)\,. \tag{4.41}$$

Se

$$\left\|(D+L)^{-1}\,U\right\| < 1 \tag{4.42}$$

a sucessão obtida converge para a solução

$$x_i^{(k+1)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{(k+1)} - \sum_{j=i+1}^{n}\frac{a_{ij}}{a_{ii}}x_j^{(k)}, \qquad i=1,\ldots\ n \quad k=0,1,\ldots$$

$$x^{(k+1)} = -(D+L)^{-1}Ux^{(k)} + (D+L)^{-1}B, \qquad M = -(D+L)^{-1}U$$

4.5.3 Condições de Convergência

Tanto o método de Jacobi como o método de Gauss-Seidel, são métodos do tipo

$$x^{(n+1)} = Mx^{(n)} + c.$$

A matriz M é designada por matriz de iteração do método. Em ambos os métodos, como vimos anteriormente, a partir de um vector inicial $x^{(0)}$, obtemos as aproximações para a solução, x.

Deste modo, se x é a solução exacta e $x^{(n+1)}$ uma sua aproximação, obtemos da relação

$$x - x^{(n+1)} = M\left(x - x^{(n+1)}\right),$$

ou seja

$$\begin{aligned} e^{(n+1)} &= Me^{(n)} \\ &= M^2e^{(n-1)} \\ &\vdots \\ &= M^{n+1}e^{(0)} \end{aligned}$$

que exprime o vector erro $e^{(n+1)}$ de iteração $(n+1)$ em função do vector erro da aproximação inicial, $e^{(0)}$. Portanto,

$$e^{(n)} = M^ne^{(0)}, n=1,2,\ldots \tag{4.43}$$

Podemos então apresentar a

Definição 4.5.1 *O método iterativo diz-se convergente se*

$$\lim_n e^{(n)} = 0.$$

Teorema 4.5.2 *É condição necessária e suficiente para que o método iterativo, de matriz de iteração M seja convergente que*

$$\lim_n M^{(n)} = 0. \tag{4.44}$$

Teorema 4.5.3 *Dada uma matriz M, as suas sucessivas potências M^n, $n = 1, 2, \ldots$ convergem para a matriz nula se e só se os valores próprios de M forem, em módulo, menores que* 1.

Um outro critério para verificar a convergência do método iterativo é o seguinte: a partir da matriz A definamos as quantidades R_i, por

$$R_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \left| \frac{a_{ij}}{a_{ii}} \right|, \quad i = 1, 2, \ldots, n.$$

Se, $R = \max_i \ R_i < 1$ então o método de Jacobi ou de Gauss-Seidel é convergente.
Em resumo,

1. Erro

$$E_a(x^{(k)}) \leq \frac{\|M\|}{1 - \|M\|} \|x^{(k)} - x^{(k-1)}\| \quad \text{ou} \quad E_a(x^{(k)}) \leq \frac{\|M\|^k}{1 - \|M\|} \|x^{(1)} - x^{(0)}\|$$

2. Condições de convergência: A matriz A do sistema ser estritamente diagonal dominante ou a matriz de iteração M ter $\|M\| < 1$.

4.6 Sistemas de equações não lineares

Um dos métodos que mais facilmente se generalizam para o caso $n-$dimensional é o Método de Newton, que se descreve de seguida.

Pretende-se determinar a solução da equação (não linear)

$$f(x) = 0, \tag{4.45}$$

onde $f \in \mathcal{C}^2([a, b])$. Então, a solução de é o limite da sucessão

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n \in \mathbb{N}_0, \tag{4.46}$$

onde $x_0 \in [a, b]$.

Consideremos agora o sistema de equações não lineares

$$\mathbf{f}(x) = 0, \tag{4.47}$$

que o qual se pode escrever na notação matricial na forma

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ \vdots & \vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

Em primeiro lugar, convém salientar que para os sistemas não lineares, em geral, é complicado provar a existência e da solução.

No que se segue, $\mathbf{z}$ designa a solução do problema (4.47). Utilizando o desenvolvimento de Taylor para aproximar a função $\mathbf{f}$, temos

$$f_i(\mathbf{z}) \approx f_i\left(\mathbf{x}^{(k)}\right) + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\left(\mathbf{x}^{(k)}\right)\left(z_j - x_j^{(k)}\right). \tag{4.48}$$

Utilizando a notação matricial e, designando por $\mathbf{J}$ a matriz Jacobiana do sistema, e por $\mathbf{h}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, o incremento na iteração k, e considerando $\mathbf{f}\left(x^{(k)}\right) = \mathbf{f}^{(k)}$, podemos escrever:

$$\mathbf{J}^{(k)}\mathbf{h}^{(k)} = -\mathbf{f}^{(k)}. \tag{4.49}$$

O sistema (4.49) é um sistema linear de equações algébricas que, na hipótese da matriz Jacobiana ser invertível, pode ser resolvido. A expressão anterior é equivalente à expressão

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left(\mathbf{J}^{(k)}\right)^{-1}\mathbf{f}^{(k)}. \tag{4.50}$$

Algumas dificuldades inerentes a este método são:

- A necessidade de calcular n^2 derivadas para formar a matriz Jacobiana, tarefa que se pode tornar muito complicada ...
- Para cada iteração, é necessário resolver um sistema de equações algébricas.
- A matriz Jacobiana pode, para alguns valores de $x \in D$, D representa o domínio, ser uma matriz singular.

Um dos métodos para resolver alguns destes problemas passa por um cálculo da matriz Jacobiana por diferenças finitas. A ideia é aproximar $\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}^{(k)}\right)$ da seguinte forma:

$$\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}^{(k)}\right) \approx \frac{1}{\eta}\left[f_i(\mathbf{x} + \eta\mathbf{e_j}) - f_i(\mathbf{x})\right] \tag{4.51}$$

em que $\mathbf{e}_j$ representa o $j-$ésimo versor de $\mathbb{R}^n$.

Notemos que podemos evitar, em cada iteração, o cálculo da matriz inversa, recorrendo ao sistema

$$\begin{cases} \mathbf{J}\left(x^n\right)\delta^{(k)} &= -\mathbf{f}\left(x^{(k)}\right) \\ \\ \delta^{(k)} &= x^{(k+1)} - x^{(k)} \end{cases} \tag{4.52}$$

Um dos métodos para obter uma boa aproximação inicial consiste na representação gráfica de cada uma das funções $F_i\left(x\right)$.

Exemplo 4.6.1 *Determine uma aproximação para a solução de*

$$\mathbf{f}\left(x\right) = 0 \Leftrightarrow \begin{cases} x^2 + y^2 - 1 &= 0 \\ \\ xy + x - 1 &= 0 \end{cases}, \tag{4.53}$$

efectuando duas iterações do método de Newton. Indique uma estimativa para o erro cometido.

Resolução 4.6.1 *Seja $x^* = (x_1^*, x_2^*)$ uma solução do problema. Vamos considerar como aproximação inicial o vector $(x,y)^{(0)} = (1,1)$ (obtida graficamente). Para não carregar a notação, vamos considerar que $(x,y)^{(0)} = (x_k, y_k)$, para $k = 0, 1, 2, \ldots$*

Neste caso, temos que

$$\mathbf{f}\left(x\right) = \begin{bmatrix} f_1\left(x,y\right) \\ f_2\left(x,y\right) \end{bmatrix} = \begin{bmatrix} x^2 + y^2 - 1 \\ xy - x - 1 \end{bmatrix}. \tag{4.54}$$

Portanto, a matriz Jacobiana é:

$$\mathbf{J} = \begin{bmatrix} 2x_k & 2y_k \\ y_k + 1 & x_k \end{bmatrix} \tag{4.55}$$

e

$$\det\left(\mathbf{J}\right) \neq 0 \Leftrightarrow x_k^2 - y_k^2 - y_k \neq 0. \tag{4.56}$$

Vamos aplicar o Método de Newton utilizando (4.52).

- *Primeira iteração:*

  Como $x_0^2 - y_0^2 - y_0 = -1 \neq 0$ podemos efectuar a primeira iteração do método. Assim,

$$\begin{bmatrix} 2x_0 & 2y_0 \\ y_0 + 1 & x_0 \end{bmatrix} \begin{bmatrix} \delta x_0 \\ \delta y_0 \end{bmatrix} = - \begin{bmatrix} x_0^2 + y_0^2 - 1 \\ x_0 y_0 + x_0 - 1 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \delta x_0 \\ \delta y_0 \end{bmatrix} = \begin{bmatrix} -1 \\ x_0 - 1 \end{bmatrix}$$

  Daqui retiramos que

$$\begin{bmatrix} \delta x_0 \\ \delta y_0 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$$

  e, como tal,

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$$

- *Segunda iteração:*

  Como $x_1^2 - y_1^2 - y_1 = -1.75 \neq 0$ *podemos efectuar a segunda iteração do método. Assim,*

$$\begin{bmatrix} 1 & 2 \\ 2 & 0.5 \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta y_1 \end{bmatrix} = \begin{bmatrix} -0.25 \\ 0-1 \end{bmatrix}$$

  Daqui retiramos que

$$\begin{bmatrix} \delta x_1 \\ \\ \delta y_1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{28} \\ \\ -\dfrac{1}{7} \end{bmatrix}$$

  e, como tal,

$$\mathbf{x}^{(2)} = \begin{bmatrix} \dfrac{15}{28} \\ \\ \dfrac{6}{7} \end{bmatrix}.$$

  Temos assim que

$$\mathbf{x}^* \approx \left(\frac{15}{28}, \frac{6}{7}\right) \approx (0.5357, 0.8571)\,, \tag{4.57}$$

  sendo uma estimativa para o erro cometido dada por

$$\left\| x^* - \mathbf{x}^{(2)} \right\|_\infty \approx \left\| \mathbf{x}^{(2)} - \mathbf{x}^{(1)} \right\|_\infty = \max\left\{\frac{1}{28}, \frac{1}{7}\right\} = \frac{1}{7} = 0.1429. \tag{4.58}$$

■

Nota 4.6.1 *Utilizando a fórmula de recorrência (4.50) temos que:*

1*) O Jacobiano (determinante da matriz Jacobiana) é*

$$\det \begin{bmatrix} 2x_k & 2y_k \\ y_k + 1 & x_k \end{bmatrix} = 2\left(x_k^2 - y_k\left(y_k + 1\right)\right).$$

Portanto, antes de efectuarmos cada iteracção é necessário garantir que $\det J^{(k)} \neq 0$

2*) A inversa da matriz Jacobiana (calculada na aproximação* x_k*) é*

$$\begin{bmatrix} 2x_k & 2y_k \\ \\ y_k + 1 & x_k \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{x_k}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} & \dfrac{y_k}{y_k\left(y_k+1\right) - x^2} \\ \\ \dfrac{y_k + 1}{2\left(y_k\left(y_k+1\right) - x^2\right)} & \dfrac{x_k}{y_k\left(y_k+1\right) - x^2} \end{bmatrix}$$

3) *Tendo novamente em conta (4.50) é necessário calcular* $\mathbf{J}^{(k)}\mathbf{f}^{(k)}$, *o que dá,*

$$\begin{bmatrix} \dfrac{x_k}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} & \dfrac{y_k}{y_k\left(y_k+1\right) - x^2} \\ \dfrac{y_k+1}{2\left(y_k\left(y_k+1\right) - x^2\right)} & \dfrac{x_k}{y_k\left(y_k+1\right) - x^2} \end{bmatrix} \begin{bmatrix} x_k^2 + y_k^2 - 1 \\ x_k y_k + x_k - 1 \end{bmatrix} =$$

$$= \begin{bmatrix} \dfrac{x_k^3 - x_k\left(y_k^2 + 2y_k + 1\right) + 2y_k}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} \\ \dfrac{x_k^2\left(y_k+1\right) - 2x_k - \left(y_k+1\right)\left(y_k^2 - 1\right)}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} \end{bmatrix}$$

4) *Consequentemente, obtemos para (4.50) a expressão:*

$$\mathbf{x}^{(k+1)} \equiv \begin{bmatrix} x_{(k+1)} \\ y_{(k+1)} \end{bmatrix} = \begin{bmatrix} \dfrac{x_k}{2} + \dfrac{x_k\left(y_k+1\right) - 2y_k}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} \\ \dfrac{y_k - 1}{2} + \dfrac{2x_k - y_k^2 - 2y_k - 1}{2\left(x_k^2 - y_k\left(y_k+1\right)\right)} \end{bmatrix}, \quad k = 0, 1, 2, \ldots$$

Exemplo 4.6.2 *Utilizando o método de Newton-Raphson, resolva o sistema não linear*

$$\begin{cases} x^3 - 2x - 4 &= 0 \\ x - y &= 0 \end{cases}$$

com aproximação inicial $x^{(0)} = y^{(0)} = 1$.

Resolução 4.6.2 *A matriz Jacobiana é:*

$$\mathbf{J}\left(x, y\right) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 3x^2 - 2 & 0 \\ 1 & -1 \end{bmatrix}$$

Consequentemente, em cada iteração é necessário resolver o sistema

$$\mathbf{J}^{(k)}\delta^{(k)} = -F\left(\mathbf{x}^{(k)}\right) \tag{4.59}$$

onde $\mathbf{J}^{(k)} = \mathbf{J}\left(\mathbf{x}^{(k)}\right)$ *e* $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta^{(k)}$. *O sistema (4.59) fica então*

$$\begin{bmatrix} 3x^2 - 2 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = - \begin{bmatrix} x^3 - 2x - 4 \\ x - y \end{bmatrix}$$

Temos então quando

1. *Para k=0*

$$\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = - \begin{bmatrix} 5 \\ 0 \end{bmatrix} \Leftrightarrow \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 1 & -1 & 0 \end{array}\right] \Leftrightarrow \begin{cases} \delta^{(0)} x &= 5 \\ \delta^{(0)} y &= 5 \end{cases}$$

o que implica que

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta^{(k)} \Leftrightarrow \mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \delta^{(0)} \Leftrightarrow \begin{cases} x^{(1)} &= x^{(0)} + \delta^{(0)} x = 1 + 5 = 6 \\ y^{(1)} &= y^{(0)} + \delta^{(0)} y = 1 + 5 = 6 \end{cases}$$

2. *Para k=1*

$$\left[\begin{array}{cc|c} 106 & 0 & -200 \\ 1 & -1 & 0 \end{array}\right] \Leftrightarrow \begin{cases} \delta^{(1)} x &= -\frac{100}{53} \approx -1.8868 \\ \delta^{(1)} y &= -\frac{100}{53} \approx -1.8868 \end{cases}$$

o que implica que

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \delta^{(1)} \Leftrightarrow \begin{cases} x^{(2)} &= x^{(1)} + \delta^{(1)} x = 4.1132 \\ y^{(1)} &= y^{(0)} + \delta^{(1)} y = 4.1132 \end{cases}$$

3. *As iterações estão apresentadas no quadro:*

k	$\delta^{(k)} x = \delta^{(k)} y$	$x = y$
0	5	6
1	-1.8868	4.1132
2	-1.1765	2.9367
3	-0.6473	2.2894
4	$-0,2492$	2.0401
5	-0.0392	2.0009
6	-0.0009	2.0000
7	0.0000	2

■

Exemplo 4.6.3 *Resolva o sistema*

$$\begin{cases} x^2 + y^2 &= 1 \\ e^x - y &= \frac{3}{2} \end{cases}$$

pelo método de Newton considerando como aproximação inicial $\mathbf{x}^{(0)} = [x_0 \;\; y_0]^T = \left[\frac{1}{2} \;\; \frac{1}{2}\right]^T$.

Resolução 4.6.3 *Sejam* $f_1(x,y) = x^2+y^2-1$ *e* $f_2(x,y) = e^x - y - \frac{3}{2}$. *A matriz Jacobiana,* $\mathbf{J}$ *é*

$$\mathbf{J}(x,y) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & 2y \\ e^x & -1 \end{bmatrix}$$

e na iteração k *é dada por*

$$\mathbf{J}\left(\mathbf{x}^{(k)}\right) = \mathbf{J}(x_k, y_k) = \begin{bmatrix} 2x_k & 2y_k \\ e^{x_k} & -1 \end{bmatrix}.$$

Em cada iteração é necessário resolver o sistema linear

$$\begin{bmatrix} 2x_k & 2y_k \\ e^{x_k} & -1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = - \begin{bmatrix} x^2+y^2-1 \\ e^x - y - \frac{3}{2} \end{bmatrix}$$

1. *Para* $k = 0$,

$$\left[\begin{array}{cc|c} 1 & 1 & \frac{1}{2} \\ 1.64872 & -1 & 0.351279 \end{array}\right] \Leftrightarrow \begin{cases} \delta_x = 0.321393 \\ \delta y = 0.178607 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1}{2} + 0.321393 = 0.821293 \\ y_1 = \frac{1}{2} + .178607 = 0.678607 \end{cases}$$

2. *Para* $k = 1$, $x_1 = 0.821293$ *e* $y_1 = 0.678607$ *e portanto*

$$\left[\begin{array}{cc|c} 1.64279 & 1.35721 & -0.135194 \\ 2.27366 & -1 & -0.0950578 \end{array}\right] \Leftrightarrow \begin{cases} \delta_x = -0.0558741 \\ \delta y = -0.0319808 \end{cases} \Rightarrow \begin{cases} x_2 = 0.765519 \\ y_1 = 0.646626 \end{cases}$$

3. *Para* $k = 2$, $x_2 = 0.765519$ *e* $y_2 = 0.646626$ *e portanto*

$$\left[\begin{array}{cc|c} 1.53104 & 1.29325 & -0.00414452 \\ 2.15011 & -1 & -0.00348399 \end{array}\right] \Leftrightarrow \begin{cases} \delta_x = -0.00200623 \\ \delta y = -0.00082962 \end{cases} \Rightarrow \begin{cases} x_3 = 0.763513 \\ y_3 = 0.645796 \end{cases}$$

4. *Para* $k = 3$, $x_2 = 0.763513$ *e* $y_2 = 0.645796$ *e portanto*

$$\left[\begin{array}{cc|c} 1.52703 & 1.29159 & -0.000004575 \\ 2.1458 & -1 & -0.000005195 \end{array}\right] \Leftrightarrow \begin{cases} \delta_x = -0.000002625 \\ \delta y = -0.000000438 \end{cases} \Rightarrow \begin{cases} x_4 = 0.763510 \\ y_4 = 0.645796 \end{cases}$$

■

4.7 Valores e vectores próprios

Recordemos que λ é valor próprio de $\mathcal{A}$ se existe $x \neq 0$ tal que $\mathcal{A}x = \lambda x$. Sendo A uma matriz representando $\mathcal{A}$, temos a equação matricial $AX = \lambda X$ ou $(A - \lambda I)X = 0$, isto é, um sistema homogéneo que deve ter soluções não nulas, isto é, deve ser indeterminado.

Logo, a condição

$$|A - \lambda I| = 0 \tag{4.60}$$

é equivalente a dizer que a matriz $A - \lambda I$ não é regular. Podemos ainda definir o espaço dos vectores próprios associados ao valor próprio λ_i como sendo o núcleo da transformação linear definida pela matriz $(A - \lambda_i I_n)$.

Definição 4.7.1 (Vector próprio dominante) *Sejam $\lambda_1, \lambda_2, \ldots, \lambda_n$ os valores próprios de uma matriz $A \in M_{(n,n)}(\mathbb{R})$. O valor próprio λ_1 é designado por valor próprio dominante se e somente se*

$$|\lambda_1| > |\lambda_i|, \quad i = 2, 3, \ldots, n. \tag{4.61}$$

Nem todas as matriz possuem um valor próprio dominante. Por exemplo, a matriz

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

que admite os valores próprios $\lambda_1 = 1$ e $\lambda_2 = -1$.

4.7.1 Localização

Seja $A \in M_{\mathbb{K}}(n)$ e seja $||A||$ sua norma. Então $||A||$ é um majorante para o raio espectral, independentemente da norma utilizada. Define-se o raio espectral de A como sendo

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|. \tag{4.62}$$

Teorema 4.7.1 *Seja $A \in M_{\mathbb{C}}(n)$. Então, o raio espectral desta matriz verifica a desigualdade*

$$\rho(A) \leq ||A||. \tag{4.63}$$

É claro que, para efeitos de localização de valores próprios de uma matriz A interessa utilizar normas fáceis de calcular, como por exemplo $||\cdot||_\infty$, $||\cdot||_1$ ou $||\cdot||_F$.

Mais, se $\lambda_1, \lambda_2, \ldots, \lambda_n$ são os valores próprios da matriz A então,

$$\frac{1}{\|A\|} \leq |\lambda| \leq \|A\|. \tag{4.64}$$

Teorema 4.7.2 *Seja $A \in M_{\mathbb{C}}(n)$. Então, o raio espectral desta matriz verifica a desigualdade*

$$(\rho(A))^2 \leq ||A||_1 ||A||_\infty. \tag{4.65}$$

Teorema 4.7.3 (Gershgorin) *Sejam $A \in M_{\mathbb{C}}(n)$, $\mathcal{G}_i$ o círculo (no plano de Argand) centrado em a_{ii} e de raio*

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ii}| \tag{4.66}$$

por vezes designado círculo de Gershgorin, e $\mathcal{G}(A) = \cup_{i=1}^{n} \mathcal{G}_i$, a região de Gershgorin. Então, todos os valores próprios de A estão contidos no domínio $\mathcal{G}(A)$.

Teorema 4.7.4 *Se k círculos de Gershgorin forem disjuntos dos restantes, então existem exactamente k valores próprios na sua união.*

Sejam $A \in M_{(n,n)}(\mathbb{C})$ e $R_i = \sum_{j \neq i} |a_{ij}|$ para $1 \leq i \leq n$. Consideremos as ovais (habitualmente designadas por *ovais de Cassini*):

$$\mathcal{O}_{ij} = \{z \in \mathbb{C} : |z - a_{ii}|\,|z - a_{jj}| \leq R_i R_j\}, \quad \forall i \neq j. \tag{4.67}$$

É válido o seguinte teorema:

Teorema 4.7.5 (A. Brauer) *Todos os valores próprios da matriz A encontram-se na reunião dos $\frac{n(n-1)}{2}$ ovais de Cassini, isto é,*

$$\sigma(A) \subseteq \cup_{i \neq j} \mathcal{O}_{ij}. \tag{4.68}$$

Sejam $A = [a_{ij}] \in M_{n,n}(\mathbb{C})$, $B = [b_{ij}] = \frac{A+A^*}{2} \in M_{n,n}(\mathbb{C})$ e $C = [c_{ij}] = \frac{A-A^*}{2} \in M_{n,n}(\mathbb{C})$. Sejam $\lambda_1, \ldots, \lambda_n$, $(|\lambda_1| > \ldots > |\lambda_n|)$, $\mu_1, \ldots, \mu_n$, $\nu_1, \ldots, \nu_n$, os valores próprios de A, B e C respectivamente. Sejam ainda $g = \max_{i,j} |a_{ij}|$, $g' = \max_{i,j} |b_{ij}|$ e $g'' = \max_{i,j} |c_{ij}|$.

Teorema 4.7.6 (Bendixson) *Se $A \in M_{(n,n)}(R)$ então*

$$|Im(\lambda_i)| \leq g'' \sqrt{\frac{n(n-1)}{2}}, \tag{4.69}$$

e

$$\mu_n \leq Re(\lambda_i) \leq \mu_1. \tag{4.70}$$

Teorema 4.7.7 (Hirsch) *Se $A \in M_{(n,n)}(\mathbb{C})$ então,*

$$|\lambda_i| \leq ng, \;\; |Re(\lambda_i)| \leq ng' \;\; e \;\; |Im(\lambda_i)| \leq ng''. \tag{4.71}$$

Se a matriz $A + A^*$ é uma matriz real a última desigualdade pode ser substituida por

$$|Im(\lambda_i)| \leq g'' \sqrt{\frac{n(n-1)}{2}}. \tag{4.72}$$

Teorema 4.7.8 (Desigualdade de Schur(1909)) *Seja* $A = [a_{ij}] \in M_{(n,n)}(\mathbb{C})$ *tem valores próprios* λ_i *para* $i = \overline{1,n}$ *então:*

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i,j=1}^{n} |a_{ij}|^2 . \tag{4.73}$$

Quando a matriz A *é normal, a desigualdade é transformada numa igualdade.*

Teorema 4.7.9 (Georg Pick) *Se* λ_i *é um valor próprio de uma matriz* $A \in M_{(n,n)}(matr)$ *então*

$$|Im(\lambda_i)| \leq g'' cotg\left(\frac{2\pi}{n}\right) \tag{4.74}$$

Sejam R_i a soma dos módulos dos elementos da i-ésima linha da matriz A e T_j a soma dos módulos dos elemenetos da j-ésima coluna da matriz A, então

$$|\lambda_i| \leq \min(R,T), |\lambda_i| \leq \frac{(R+T)}{2}, \tag{4.75}$$

onde $R = \max_i R_i$ e $T = \max_j T_j$.

A última desigualdade foi melhorada por E. T. Browne em 1930 para

$$|\lambda_i| \leq \max_i \frac{(R_i + T_i)}{2} \tag{4.76}$$

e, em 1944 A.B. Farnel obteve o seguinte resultado:

$$|\lambda_i| \leq \sqrt{RT}. \tag{4.77}$$

4.7.2 Alguns processos para o cálculo de valores próprios

Vamos de seguida apresentar alguns métodos que nos permitem calcular aproximações para os valores próprios e os respectivos vectores próprios.

1. Método da Potência:

   Permite-nos o cálculo do valor próprio de maior módulo. É um dos métodos mais simples, mas só se pode aplicar no caso em que A é tem n vectores próprios linearmente independentes e um valor próprio λ de maior módulo, isto é, se por exemplo,

   $$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n| . \tag{4.78}$$

   Seja Z_0 um vector qualquer de dimensão n. Formemos a sucessão de vectores:

   $$Z_{n+1} = AZ_k \equiv A^{k+1} Z_0, \quad k = 0, 1, 2, \ldots \tag{4.79}$$

   Se considerarmos Z_0 nas suas componentes, isto é,

$$Z_0 = \sum_{i=1}^{n} \alpha_i e_i, \tag{4.80}$$

então

$$Z_k = \sum_{i=1}^{n} \lambda_i^k \alpha_i e_i = \lambda_1^k \left[\alpha_1 e_1 + \sum_{i=2}^{n} \left(\frac{\lambda_j}{\lambda_1} \right) \alpha_i e_i \right], \quad k = 0, 1, 2, \ldots \tag{4.81}$$

Como os coeficientes $\left| \frac{\lambda_i}{\lambda_1} \right| < 1$, $i \geq 2$, então $\lim_{k \to \infty} Z_k = \lambda_1 \alpha_1 e_1$ terá a direcção de e_1, desde que $\alpha_1 \neq 0$. Formando o quociente $\frac{Z_{k+1}}{Z_k}$ componente a componente, teremos uma aproximação para λ_1.

Este método pode ser esquematizado na seguinte forma:

1) Considerar um vector inicial x_0 não nulo. Considere-se $i = 0$;
2) Obter a iteração seguinte mediante a formula de recorrência $x_{i+1} = Ax_i$;
3) dividir cada termo do vector x_{i+1} pelo último elemento do vector e este vector passa aser designado por x'_{i+1};
4) Repetir os passos 2 e 3 até que os vectores x'_i e x'_{i+1} concordem no número de dígitos;
5) O vector obtido em 4 é um vector aproximado para o vector próprio correspondente ao vector próprio dominante. Este vector será designado por x.
6) Um valor aproximado para o valor próprio dominante é então dado pela formula

$$\lambda \approx \frac{x^T A x}{x^T x}.$$

Nota 4.7.1 *A escolha do último elemento do vector para efectuar a divisão é arbitrária. Caso o último elemento seja nulo, podemos escolher um outro elemento qualquer do vector mas, é necessário manter esse critério ao longo de todo a resolução. Uma outra possibilidade é dividir todos os elementos do vector por*

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Se utilizarmos o método da potência para a matriz inversa A^{-1}, então, de modo análogo, obteremos o valor próprio de A de menor módulo desde que

$$|\lambda_n| > |\lambda_{n-1}| \geq |\lambda_{n-2}| \geq \ldots \geq |\lambda_1| . \tag{4.82}$$

O método das potência pode falhar pelas seguintes razões:

a) x_0 não ter componente na direcção do primeiro vector (aproximação inicial) - na prática não há problemas porque os erros de arredondamento acabam por introduzir essa componente;

b) Pode haver mais do que um valor próprio que tenha a mesma magnitude (máxima) em módulo, neste caso as iterações vão convergir para um vector que é combinação linear dos vectores próprios associados aos valores próprios dominantes;

c) Para matriz e vector inicial reais as iterações podem nunca convergir para vectores próprios complexos.

2. Método do quociente de Rayleigh

O *quociente de Rayleigh* é definido (para x e para A)

$$\lambda = \frac{x^T Ax}{x^T x}. \tag{4.83}$$

Neste método, também o quociente de Rayleigh de Z_k é gerado por (4.79), tenderá para λ_1, ou seja:

$$\lambda_1 \approx \tau_k = \frac{Z_k^T AZ_k}{Z_k^T Z_k}, \quad k = 0, 1, 2, \ldots \tag{4.84}$$

Assim, após estar determinado um vector próprio utilizando, por exemlo, o método das potências podemos utilizar a expressão (4.83) para determinar o seu valor próprio correspondente.

3. Método da iteração inversa

Uma variante do método da potência é o chamado *método da iteração inversa.* Nele escolhe-se para além de um vector de partida Z_0 também um número p diferente de um valor próprio valor próprio de A, e depois, se formarmos a seguinte sucessão

$$y_k = \frac{Z_k}{\|Z_k\|^2} \tag{4.85}$$

$$(A - pI_n)\, Z_{k+1} = y_k, \quad k = 0, 1, 2, \ldots \tag{4.86}$$

teremos o equivalente ao método da potência com a matriz $(A - pI)^{-1}$. Como esta matriz tem valores próprios $(\lambda_i - p)^{-1} = \dfrac{1}{\lambda_i - p}$, sendo λ_i os valores próprios de A, este método dá o valor próprio mais próximo de p e o respectivo valor próprio.

Exemplo 4.7.1 *Utilizando o método da Potência, obtenha o maior valor próprio (em módulo)de*

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

Resolução 4.7.1 *Os valores próprios da matriz A são* $\lambda_1 \approx 0.3003718517$, $\lambda_2 \approx 2.239123278$ *e* $\lambda_3 \approx 4.460504870$.

Vamos utilizar o método das potências. Seja Z_0 *o vector* $[\,1\ \ 1\ \ 1\,]^T$ *e formemos a seguinte tabela (construída a utilizando (4.79):*

Z_0	AZ_0	A^2Z_0	A^3Z_0	A^4Z_0	A^5Z_0	A^6Z_0	A^7Z_0	A^8Z_0	A^9Z_0	$A^{10}Z_0$	$\frac{A^{10}Z_0}{A^9Z_0}$
1	5	24	111	504	2268	10161	45423	202833	905238	4038939	4.462
1	4	15	60	252	1089	4779	21141	93906	417987	1862460	4.456
1	2	6	21	81	333	1422	6201	27342	121248	539235	4.447

formando os quocientes $\dfrac{Z_{10}}{Z_9}$ *componente a componente, obtemos a última coluna da tabela anterior. Tomando a média, verificamos que*

$$\lambda_1 = \frac{1}{3}\,(4.462 + 4.456 + 4.447) = 4.455.$$

Portanto, uma aproximação para o valor próprio dominante é $\lambda_3 \approx 4.455$ *Este valor é uma boa aproximação para o valor próprio* λ_3.

■

Exemplo 4.7.2 *Determine o valor próprio dominante da matriz*

$$A = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}.$$

Resolução 4.7.2 *Os valores próprios da mariz A são* $\lambda_1 = 6$ *e* $\lambda_2 = 1$. *Vamos em primeiro lugar, utilizando o método da Potência, determinar o vector próprio. Consideremos* $x_0 = [1 \quad 1]^T$. *Temos então*

$$\begin{aligned}
x_1 &= Ax_0 = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 5 \end{bmatrix} \Longrightarrow x_1' = \begin{bmatrix} 1.8 \\ 1 \end{bmatrix} \\
x_2 &= Ax_1 = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1.8 \\ 1 \end{bmatrix} = \begin{bmatrix} 11.4 \\ 5.8 \end{bmatrix} \Longrightarrow x_2' = \begin{bmatrix} 1.965517241 \\ 1 \end{bmatrix} \\
x_3 &= Ax_2 = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1.965517241 \\ 1 \end{bmatrix} = \begin{bmatrix} 11.89655172 \\ 5.96551724 \end{bmatrix} \Longrightarrow x_3' = \begin{bmatrix} 1.994219653 \\ 1 \end{bmatrix} \\
x_4 &= Ax_3 = \begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1.994219653 \\ 1 \end{bmatrix} = \begin{bmatrix} 11.98265896 \\ 5.99421965 \end{bmatrix} \Longrightarrow x_4' = \begin{bmatrix} 1.99903568 \\ 1 \end{bmatrix}
\end{aligned}$$

Podemos verificar que a sucessão converge para o vector $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. *O valor próprio correspondente é então dado pelo quociente de Rayleigh:*

$$\lambda = \frac{x^TAx}{x^Tx} = \frac{[2 \quad 1]\begin{bmatrix} 3 & 6 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{[2 \quad 1]\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \frac{[30]}{[5]} = 6. \tag{4.87}$$

■

Exemplo 4.7.3 *Determine o valor próprio dominante da matriz*

$$A = \begin{bmatrix} 2 & 6 \\ 2 & -2 \end{bmatrix}.$$

Resolução 4.7.3 *Vamos em primeiro lugar, utilizando o método da Potência, determinar o vector próprio. Consideremos* $x_0 = [1 \quad 1]^T$. *Temos então*

$$\begin{aligned} x_1 &= Ax_0 = \begin{bmatrix} 2 & 6 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \end{bmatrix} \Longrightarrow x_1' \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ x_2 &= Ax_1 = \begin{bmatrix} 2 & 6 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Longrightarrow x_2' \begin{bmatrix} 1 \\ 1 \end{bmatrix} = x_0! \end{aligned}$$

Devemos salientar que $x_0 = x_2$. *Isto vai implicar que o os vectores vão entrar num "ciclo". Esta situação ocorre pois não há um valor próprio dominante. Os valores próprios da matriz* A *são* $\lambda_1 = 4$ *e* $\lambda_2 = -4$.

Neste caso, dividimos o vector por 8 *para obter o valor de* x_1'.

■

4.8 Exercícios

1. Considere o vector $u = (-1, 2, 3)$. Determine $\|u\|_1$, $\|u\|_2$ e $\|u\|_\infty$.

2. Considere as matrizes

$$A_1 = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix}, \; A_2 = \begin{bmatrix} 1 & 2 & -1 \\ -2 & -3 & 4 \\ 0 & -2 & 3 \end{bmatrix} \text{ e } A_3 = \begin{bmatrix} 1 & 0 & 2 & -1 \\ -1 & -2 & -3 & 4 \\ 0 & -2 & 3 & -3 \\ 0 & \pi & \sqrt{2} & 2 \end{bmatrix}$$

   Para cada uma delas determine $\|A\|_1$, $\|A\|_\infty$ e $\|A\|_E$.

3. Considere o sistema de equações lineares

$$\begin{cases} 3x + 2y &= 0 \\ 2y - z &= 0 \\ y + 7z &= 0 \end{cases}$$

   a) Prove que o método de Jacobi é convergente

4. Resolva o sistema de equações lineares $AX = B$ onde

$$A = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & 0 \\ 1 & 0 & -2 \end{bmatrix}, \; X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ e } B = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix},$$

   utilizando os métodos Cramer, de Jacobi e o método de Gauss-Seidel. Utilizando a matriz adjunta resolva o sistema $AX = B$.

5. Utilizando o método de Jacobi, determine a solução do sistema de equações lineares $AX = B$ onde

$$A = \begin{bmatrix} 7 & -2 & 1 & 2 \\ 2 & 8 & 3 & 1 \\ -1 & 0 & 5 & 2 \\ 0 & 2 & -1 & 4 \end{bmatrix} \quad \text{e } B = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 4 \end{bmatrix} \tag{4.88}$$

6. Utilizando o método de Jacobi, determine a solução do sistema de equações lineares $AX = B$ onde

$$A = \begin{bmatrix} 2 & 8 & 3 & 1 \\ 7 & -2 & 1 & 2 \\ -1 & 0 & 5 & 2 \\ 0 & 2 & -1 & 4 \end{bmatrix} \quad \text{e } B = \begin{bmatrix} -2 \\ 3 \\ 5 \\ 4 \end{bmatrix} \tag{4.89}$$

7. Sejam

$$A = \begin{bmatrix} 4 & 3 & 0 \\ -1 & 3 & 0 \\ 0 & 1 & 2 \end{bmatrix} \text{ e } B = \begin{bmatrix} 0 \\ 3 \\ 8 \end{bmatrix}$$

   a) Prove que os métodos de Jacobi e de Gauss-Seidel convergem.

   b) Resolva o sistema $AX = B$ utilizando a decomposição em $A = LU$.

   c) Resolva o sistema $AX = B$ utilizando o método de Jacobi, considerando $X = [\,0\ \ 0\ \ 0\,]^T$.

   d) Resolva o sistema $AX = B$ utilizando o método de Gauss-Seidel, considerando $X = [\,0\ \ 0\ \ 0\,]^T$.

8. Resolva o mesmo exercício com as matrizes

$$A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 3 & 0 \\ 4 & 3 & 0 \end{bmatrix} \text{ e } B = \begin{bmatrix} 8 \\ 3 \\ 0 \end{bmatrix}$$

9. Resolva o sistema

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 &= 1 \\ x_1 - x_2 - 3x_3 &= 4 \\ 3x_1 + 2x_2 + (1+\varepsilon)\, x_3 &= 5 \end{cases},$$

   para $\varepsilon = 0$ e para $\varepsilon = 1$.

10. Determine a decomposição em LU da matriz de Hilbert de ordem n ($n = 3, 4, \ldots, 10$).

11. Utilize a decomposição em LU para calcular a inversa da matriz

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

12. Considere a matriz

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 5 & 5 & 5 \\ 1 & 5 & 14 & 14 \\ 1 & 5 & 14 & 30 \end{bmatrix}$$

a) Determine a decomposição de Cholesky da matriz A.

b) Determine a decomposição em LU da matriz A.

c) Utilize o método de Jacobi para resolver o sistema $AX = B$ onde

$$B = [-4 \quad -4 \quad 23 \quad 39]^T.$$

d) Utilize o método de Gauss-Seidel para resolver o sistema $AX = B$ onde

$$B = [-4 \quad -4 \quad 23 \quad 39]^T.$$

13. Considere a matriz

$$A = \begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 255 \\ 55 & 225 & 979 \end{bmatrix}$$

a) Determine a decomposição em LU da matriz A.

b) Determine a decomposição de Cholesky da matriz A.

c) Suponha que pretende resolve o sistema $AX = B$ onde $B = [5.9 \quad 14.9 \quad 254.5]^T$

14. Considere o sistema de equações lineares

$$\begin{cases} x_1 + 10x_2 + x_3 = 12 \\ x_1 + x_2 + 10x_3 = 12 \\ 10x_1 + x_2 + x_3 = 12 \end{cases}$$

a) Reordene as linhas da matriz de modo a que o novo sistema tenha a diagonal estritamente dominantes;

b) Aplique o método de Jacobi ao novo sistema e efectue 4 iterações. Calcule um majorante para o erro da 4° iteração. Considere $\mathbf{x}^{(0)} = [-4 \quad -4 \quad -4]^T$

c) Aplique o método de Gauss-Seidel até que $\left\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\right\|_\infty < 10^{-2}$

15. Considere o sistema de equações não lineares

$$\begin{cases} x^2 + y^2 &= 1 \\ \frac{y}{2} + \sin(x) &= 0 \end{cases}.$$

a) Utilizando a representação gráfica, determine uma aproximação para as duas (únicas) raízes do sistema.

b) Aplicando o método de Newton-Rapshon, determine uma aproximação para a solução considerando como aproximações iniciais (para cada uma das raízes), $\mathbf{x}^{(0)} = [0.74 \quad 1]^T$ e $\mathbf{x}^{(0)} = [-0.74 \quad -1]^T$.

16. Considere o sistema de equações não linear

$$\begin{cases} x^2 + 2y^2 &= r^2 \\ \frac{1}{10}\cos(x) &= 0 \end{cases}$$

a) Indique um valor de r para o qual o sistema admite pelo menos uma solução.

b) Utilizando a representação gráfica, determine uma aproximação para as raízes do sistema.

c) Aplicando o método de Newton-Rapshon, determine uma aproximação para uma solução considerando como aproximação inicial, $\mathbf{x}^{(0)} = [0.74 \quad -0.3]^T$.

17. Utilize o método de Newton-Raphson para encontrar as soluções aproximadas para o sistema de equações

$$\begin{cases} x^2 - 2x - y &= -\frac{1}{2} \\ x^2 + 4y^2 &= 4 \end{cases}$$

Utilize a representação gráfica para determinar uma aproximação inicial.

18. Determine uma solução aproximada para os sistemas de equações não lineares

a) $\begin{cases} x^2 + y^2 &= 1 \\ x^3 - y &= 0 \end{cases}$

b) $\begin{cases} x^2 + y^2 &= \pi \\ x^3 + y &= 0 \end{cases}$

c) $\begin{cases} x^2 + y &= 4 \\ y - \ln(x) &= 1 \end{cases}$

19. Pretende-se resolver pelo método de Newton o sistema de equações não lineares

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

a) Considerando como aproximação inicial $[x_0\ y_0\ z_0]^T = [0\ 1\ 2]^T$, ao efectuar uma iteração pelo método de Newton, somos conduzidos a um sistema de equações lineares Qual?

b) Resolva o sistema de equações lineares obtido na alínea anterior utilizando o método de Gauss-Seidel considerando como aproximação inicial o vector nulo e efectuando duas iterações.

20. Considere o sistema de equações não-lineares:

$$\begin{cases} 2x - \cos(x+y) &= 2 \\ 3y - \sin(x+y) &= 6 \end{cases}$$

Utilizando o método de Newton Raphson, considerando $[x_0\ \ y_0]^T = [1\ \ 1]$, determine efectuando duas iterações, uma aproximação para a solução.

Capítulo 5

Interpolação polinomial

5.1 Breve introdução histórica

Muitas funções são conhecidas apenas num conjunto finito e discreto de pontos de um intervalo. Neste caso, como não dispomos da sua forma analítica, podemos substituí-la por outra função, que é uma aproximação da função dada e que é deduzida a partir dos dados tabelados. Outras funções têm uma forma analítica muito complexa e portanto, é necessário procurar uma outra função que seja uma aproximação da função dada e cujo manuseio seja bem mais simples. As funções que substituem as funções dadas podem ser de vários tipos: exponencial, logarítmica, trigonométrica e polinomial.

Al-Biruni (973-1050), um grande matemático árabe, já usava interpolação quadrática e é possível que tivesse tido imitadores e discípulos que o fizessem também. Mas foi só no séc. XVII que se efectuaram os primeiros estudos sistemáticos sobre esta matéria, nomeadamente sobre o cálculo das diferenças finitas. Henry Briggs (1556-1630) teve um papel importante nesta matéria ao usar fórmulas de interpolação para tabelar os logaritmos. No entanto, é necessário recuar um pouco no tempo até um colega seu, Thomas Harriot (1560-1621) em Oxford, para encontrar o verdadeiro inventor do cálculo das diferenças finitas. Harriot, tal como Briggs, estava muito interessado em problemas de navegação. Apesar de contribuir de forma notável para esta área da Análise Numérica, os seus trabalhos foram substimados e pouco estudados. Briggs desenvolveu e aplicou ao cálculo logarítmico os trabalhos do seu predecessor tendo sido reconhecido, mais tarde, pelo grande matemático francês Lagrange. No entanto, nesta área como em tantas outras, talvez ninguém tenha feito tanto como o génio inglês Isaac Newton (1643-1727). Ele, aparentemente, desenvolveu os seus estudos desconhecendo os resultados de Harriot e Briggs. O aparecimento das suas ideias surge numa carta datada de 8 de Maio de 1675 mas a publicação definitiva teve que esperar até muito mais tarde. Newton pretendia ajudar um colega seu, John Smith, que estava profundamente interessado no problema de Wallis: determinar a área do quadrante de um círculo dada por

$$\int_0^1 \sqrt{1-x^2}dx$$

Estas preocupações levaram-no ao aprofundamento das suas ideias nesta matéria até produzir o conceito actual de diferença finita. É de notar que o esforço de Newton foi amplamente reconhecido e meritório, facto esse visível na numerosa quantidade de fórmulas

ligadas á teoria da interpolação com o seu nome: Newton; Gregory-Newton; Newton-Gauss; Newton-Cotes; Newton-Bessel, etc. Da generalização de um resultado apresentado por John Wallis que consistia em obter por interpolação, Newton obteve o que é hoje conhecido por Teorema do Binómio sobre o qual Fernando Pessoa disse, pela boca de Álvaro de Campos: "*o Teorema do Binómio é tão belo como a Vénus de Milo; o que há é pouca gente para dar por isso*". Este teorema é considerado como um dos mais brilhantes resultados da Matemática. Note-se, de passagem, que o símbolo "1"para designar infinito foi introduzido por Wallis neste seu trabalho. O matemático suíço Leonhard Euler (1707-1783) também deu um importante contributo no capítulo da interpolação publicando inúmeros resultados e introduzindo uma nova e simples notação que ainda hoje é usada. Joseph Louis Lagrange (1736-1813) também se dedicou a esta área da Análise Numérica, sobretudo inspirado pelas ideias de Briggs. Lagrange trabalhou seriamente estes assuntos publicando numerosos resultados entre os quais poderemos destacar: o estabelecimento da relação entre as derivadas de uma função e as suas diferenças; a apresentação, em 1794/5, da fórmula de interpolação que actualmente possui o seu nome mas que ele atribuia a Newton; a descoberta da interpolação trigonométrica (Alexis Claude Clairaut (1713-1765) também a descobriu independentemente), etc... Outros matemáticos brilhantes que também trabalharam nesta matéria foram: Pierre Simon Laplace (1749-1827) no cálculo das diferenças finitas; Carl Friedrich Gauss (1777-1855) na determinação de fórmulas de quadratura numérica; Augustine Louis Cauchy (1789-1857) em interpolação por fracções racionais; etc... Para terminar é de salientar ainda o nome do matemático francês Charles Hermite (1822- 1901) cujo resultado mais conhecido nesta área é, sem dúvida, a fórmula de interpolação com o seu nome. Hermite foi também um dos primeiros a notar a beleza e importância do teorema dos rezíduos de Cauchy e como este poderia ser usado para obter aproximação de funções.

5.2 Polinómio interpolador

Definição 5.2.1 (Polinómio Interpolador) *Seja $f \in \mathcal{C}([a,b])$ e $x_i \in [a,b]$, $i = 0,1,2,\ldots,n$. Um polinómio $p_n(x)$ de grau menor ou igual que n e que assume os mesmos valores de f nos pontos x_0, x_1,..., x_n, isto é, que satisfaz a condição*

$$f(x_i) = p_n(x_i), \quad i = 0,1,\ldots,n \tag{5.1}$$

é designado por polinómio interpolador *de f nos pontos x_0, x_1,..., x_n.*

Os polinómios interpoladores são gerados através da combinação linear de outras funções polinomiais pertencentes a bases de funções

$$\{\varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)\}.$$

A interpolação pode também ser feita com funções não-polinomiais, contudo iremos apenas considerar as polinomiais.

O Polinómio interpolador p_n é escolhido como combinação linear de uma base de funções

$$p_n(x) = \sum_{j=0}^{n} \varphi_j(x).$$

Impondo a condição $p_n(x)$ interpole os dados (x_i, y_i) significa que

$$p_n(x_i) = \sum_{j=0}^{n} a_j \varphi_j(x_i) = y_i$$

para $0 \leq i \leq n$ correspondendo a um sistema linear $AX = B$, em que X é um vector com $(n+1)$ componentes a_j e as entradas de matriz A, de dimensão $(n+1)^2$, são dadas por $a_{ij} = \varphi_j(x_i)$.

Podemos, por exemplo utilizar a base mónica,

$$\mathcal{B} = \left\{1, x, x^2, \ldots, x^n\right\}, \tag{5.2}$$

que original um polinómio da forma

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n \tag{5.3}$$

onde os a_i são solução do sistema de equações lineares $AX = B$ onde

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}, X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \text{ e } B = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \equiv \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \tag{5.4}$$

Neste capítulo vamos ainda utilizar a base de Lagrange,

$$L_i = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - x_k}{x_i - x_k}, \quad (i = 0, 1, \ldots, n) \tag{5.5}$$

que dá origem ao polinómio da forma

$$\mathbb{P}_n(x) = f(x_0) L_0(x) + f(x_1) L_1(x) + \cdots + f(x_n) L_n(x).$$

e a base de Newton,

$$\pi_k(x) = \prod_{i=0}^{k-1} (x - x_i), \text{ para } k = 0, 1, \ldots, n, \tag{5.6}$$

em que o valor do produto é igual a 1 se os limites da multiplicação forem nulos.

Neste caso, o polinómio interpolador é da forma:

$$p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_1)(x - x_0) + \cdots + a_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}). \tag{5.7}$$

É possível utilizar bases com funções trigonométricas, como por exemplo,

$$\mathcal{B} = \left\{\sin\left(\frac{k\pi x}{L}\right)\right\}_{k=1}^{n} \text{ ou } \mathcal{B} = \left\{\cos\left(\frac{k\pi x}{L}\right)\right\}_{k=0}^{n}, x \in [0, L], L < \infty.$$

A sensibilidade de X, solução do problema, a perturbações nos dados depende do número de condição, $cond(A)$, que por sua vez depende da escolha da base de funções e dos pontos escolhidos.

A matriz A é regular desde que os pontos dados x_i sejam distintos e portanto, o polinómio interpolador existe e é único. Existem muitas técnicas de representar ou calcular um polinómio interpolador, mas em teoria todas devem conduzir ao mesmo resultado.

Vamos de seguida apresentar um exemplo de como determinar um polinómio interpolador para um conjunto de dados. Pretende-se determinar um polinómio de grau menor ou igual que 3 que interpole a função $f(x) = \log_{10}(x)$ no intervalo $(2.3, 2.6)$. Para tal, vamos utilizar quatro pontos: 2.3, 2.4, 2.5 e 2.6. De acordo com a definição temos

$$\begin{aligned} p_3(2.3) &= 0.36173 \\ p_3(2.4) &= 0.38021 \\ p_3(2.5) &= 0.39794 \\ p_3(2.6) &= 0.41497, \end{aligned}$$

Isto é, se $p_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, obtemos o sistema de equações lineares

$$\left\{ \begin{array}{rcl} a_0 + 2.3a_1 + 5.29a_2 + 12.167a_3 &=& 0.36173 \\ a_0 + 2.4a_1 + 5.76a_2 + 13.824a_3 &=& 0.38021 \\ a_0 + 2.5a_1 + 6.25a_2 + 15.625a_3 &=& 0.39794 \\ a_0 + 2.6a_1 + 6.76a_2 + 17.576a_3 &=& 0.41497 \end{array} \right. \tag{5.8}$$

Na forma matricial, obtemos

$$A = \begin{bmatrix} 1 & 2.3 & (2.3)^2 & (2.3)^3 \\ 1 & 2.4 & (2.4)^2 & (2.4)^3 \\ 1 & 2.5 & (2.5)^2 & (2.5)^3 \\ 1 & 2.6 & (2.6)^2 & (2.6)^3 \end{bmatrix}, \quad X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \text{ e } B = \begin{bmatrix} 0.36173 \\ 0.38021 \\ 0.39794 \\ 0.41497 \end{bmatrix}. \tag{5.9}$$

Este sistema admite uma e uma só solução

$$X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -0.3853100000001011 \\ 0.5049666666668093 \\ -0.09750000000005393 \\ 0.008333333333 \end{bmatrix}, \tag{5.10}$$

e portanto o polinómio $p_3(x)$ é único e é dado por

$$\begin{aligned} p_3(x) &= -0.3853100000001011 + 0.5049666666668093x \\ &\quad -0.09750000000005393x^2 + 0.008333333333x^3. \end{aligned}$$

Se, por exemplo, pretendermos calcular uma aproximação para $\log_{10}(2.45)$ temos que

$$\log_{10}(2.45) \approx 0.38916. \tag{5.11}$$

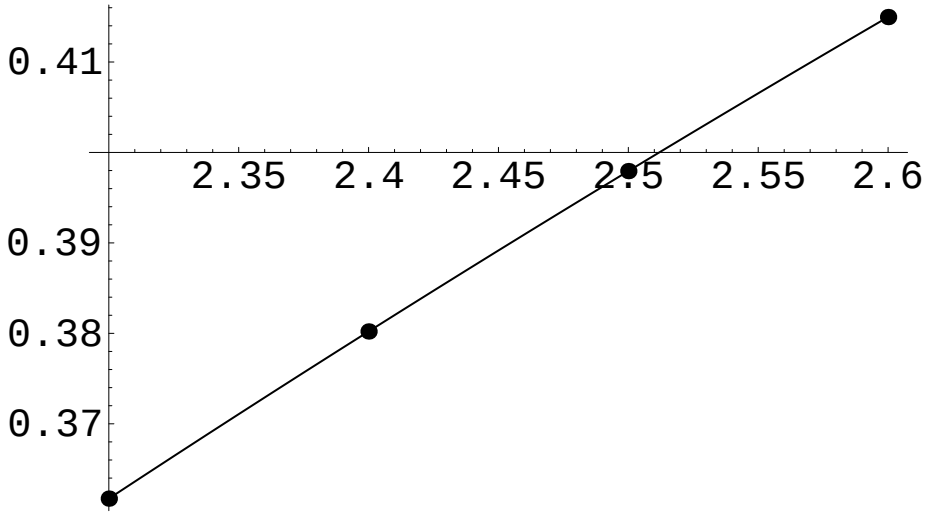

Figura 5.1: Gráficos das funções $f(x) = \log_{10}(x)$ e do polinómio $p(x)$

Comparando com o valor "exacto", $\log_{10}(2.45) = 0.389166608\ldots$. Devemos salientar que o erro cometido na aproximação não excede 0.7×10^{-5}.

No gráfico 5.1 está representada a função $f(x) = \log_{10}(x)$ e os pontos correspondem aos valores de $p_3(x)$. Facilmente se verifica que a função interpoladora e a função $\log_{10}(x)$ tomam os mesmos valores nos pontos considerados.

Nota 5.2.1 *Determine o número de condição da matriz dos coeficientes em (5.9). O que pode concluir? Os cálculos aqui apresentados foram efectuados utilizando o programa Mathematica. Justifique a diferença entre os seus resultados e os resultados aqui apresentados.*

Os polinómios interpoladores são meios muito utilizados na aproximação de funções. As fórmulas desenvolvidas são a base do desenvolvimento de métodos numéricos para o cálculo de raízes de equações (veja-se por exemplo o método da Secante, ver secção 3.4.5), cálculo integral e de derivadas.

Vamos determinar o polinómio interpolador h_3 de grau menor ou igual que 3 de $f(x) = \log_{10}(x)$ e da sua derivada nos pontos 2.4 e 2.5. Então,

$$\begin{aligned} h_3(2.4) &= 0.38021 \\ h_3(2.5) &= 0.39794 \\ h_3'(2.4) &= 0.18096 \\ h_3'(2.5) &= 0.38021 \end{aligned}$$

Temos então

$$\begin{cases} c_0 + 2.4c_1 + 5.76c_2 + 13.824c_3 &= 0.38021 \\ c_0 + 2.5c_1 + 6.25c_2 + 15.625c_3 &= 0.39794 \\ c_1 + 4.8c_2 + 17.28c_3 &= 0.18096 \\ c_1 + 5.0c_2 + 18.75c_3 &= 0.17372 \end{cases} \tag{5.12}$$

Este sistema é possível e determinado e tal polinómio existe e é único, e é dado por

$$h_3(x) = -0.38011 + 0.49872x - 0.09500x^2 + 0.00800x^3. \tag{5.13}$$

Uma pergunta que surge do problema de interpolação polinomial é se, quando aumentamos o grau do polinómio interpolador, este fica mais próximo da função original. Intuitivamente, poderíamos dizer que sim, a exemplo do polinómio de Taylor. Considere a função

$$f(x) = \frac{1}{1+25x^2},$$

no intervalo $[-1, 1]$. A representação gráfica da função nesse intervalo é

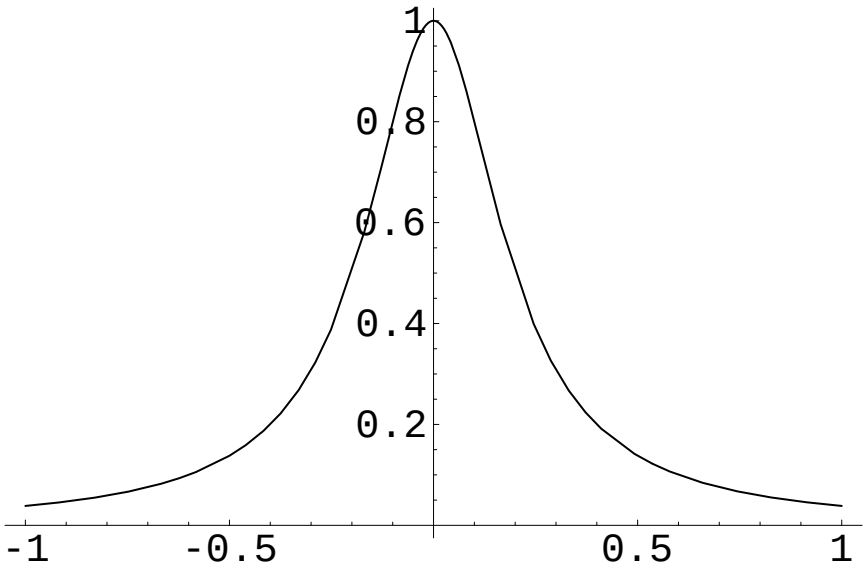

Figura 5.2: Gráfico da função $f(x) = \frac{1}{1+25x^2}$ em $x \in [-1, 1]$.

Utilizando pontos da forma $(x_i, y_i) = \left(-1 + \frac{i}{2}, f\left(-1 + \frac{i}{2}\right)\right)$ para $0 \leq i \leq 4$ obtemos o polinómio interpolador

$$p_4(x) = 1 - \frac{3225}{754}x^2 + \frac{1250}{377}x^4. \tag{5.14}$$

Utilizando pontos da forma $(x_i, y_i) = \left(-1 + \frac{i}{6}, f\left(-1 + \frac{i}{6}\right)\right)$ para $0 \leq i \leq 12$ obtemos o polinómio interpolador

$$\begin{aligned} p_{12}(x) &= 1 - 19.58283572x^2 + 198.7233435x^4 - 955.3733275x^6 + 2201.753993x^8 \\ &\quad -2336.358291x^{10} + 909.8755783x^{12}. \end{aligned} \tag{5.15}$$

A figura 5.3 mostra o gráfico de $f(x)$ assim como os gráficos do polinómios de grau 4(5.14 - a cor verde) e de grau 12 (5.15 - a cor azul) que interpola $f(x)$.

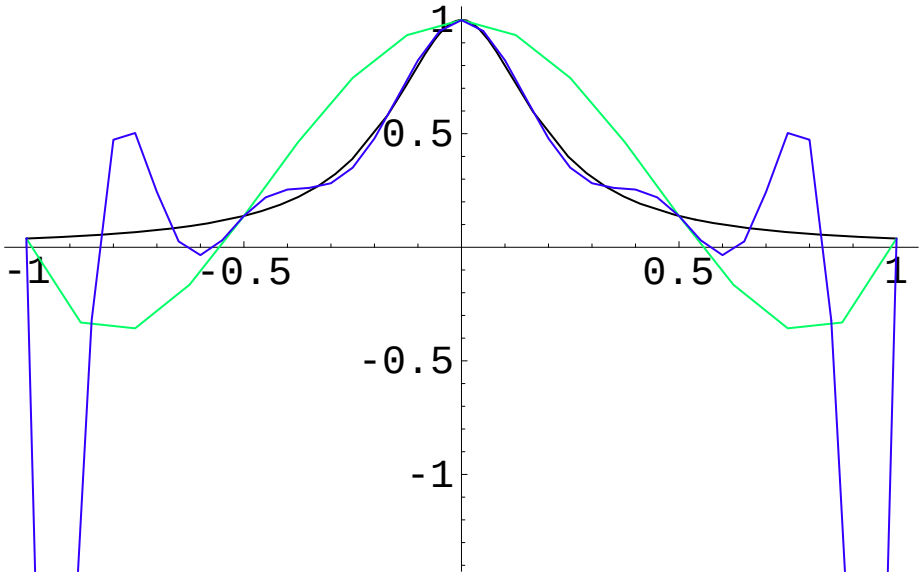

Figura 5.3: Gráfico da função $f(x) = \frac{1}{1+25x^2}$, $p_4(x)$ e $p_{12}(x)$ em $x \in [-1, 1]$.

Vê-se claramente que quanto maior é o grau do polinómio interpolador, maior é diferença entre as funções $f(x)$ e $p_n(x)$, principalmente nas extremidades do intervalo. Podemos então conluir que a sucessão de polinómios interpoladores não converge para $f(x)$. Este fenómeno é designado por *fenónemo de Runge* e a função

$$f(x) = \frac{1}{1+25x^2} \tag{5.16}$$

é designada por *função de Runge*.

Este problema pode ser evitado utilizando os pontos de Chebyshev, definidos por

$$x_i = \cos\left(\frac{i\pi}{n}\right),\ j = 0, \ldots, n, \tag{5.17}$$

onde n é o grau do polinómio.

Utilizando o desenvolvimento em séries de potências, a $f(x)$ em série de potências, numa vizinhança da origem e utilizando 3 termos é dada por

$$s_4(x) = 1 - 25x^2 + 625x^4. \tag{5.18}$$

Neste caso, os gráficos coincidem numa vizinhança da origem. De seguida vamos apresentar dois resultados muito importantes que nos garantem a existência do polinómio interpolador.

Teorema 5.2.1 (Stone-Weierstrass) *Seja $f : [a, b] \to \mathbb{R}$ uma função contínua. Então, para todo o $\varepsilon > 0$, existe um polinómio $p(x)$, de grau menor ou igual a n de forma que*

$$\|f(x) - p(x)\|_\infty \leq \varepsilon,\ \textit{ou seja},\ |f(x) - p(x)| \leq \varepsilon,\ \forall x \in [a, b]. \tag{5.19}$$

Teorema 5.2.2 *Sejam $x_0, x_1, \ldots, x_n$ $(n+1)$ pontos distintos e sejam $y_i = f(x_i)$, para $0 \leq i \leq n$ os valores de uma função contínua $f(x)$ no intervalo $[a,b]$ com $x_0 = a$ e $x_n = b$. Então, existe um único polinómio $p_n(x)$,*

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \tag{5.20}$$

tal que

$$p_n(x_i) = y_i, \quad \forall i = 0, \ldots, n. \tag{5.21}$$

5.2.1 Erro de interpolação

Como vimos anteriormente, o polinómio interpolador coincide com a função num dado conjunto de pontos de suporte. No entanto, aquando da utilização de polinómios de grau muito elevado, habitualmente há grandes diferenças nos valores fora do conjunto de pontos (os polinómios de grau elevado oscilam muito). Consequentemente, é importante determinar o que se passa nos outros pontos do domínio da função, isto é, se para os outros pontos do domínio da função, o polinómio interpolador constitui uma boa ou má aproximação.

Teorema 5.2.3 *Seja $\mathbb{P}_n(x)$ o polinómio de grau menor ou igual que n nos pontos $\{x_0, x_1, \ldots, x_n\}$. Se $f(x) \in \mathcal{C}^n([a,b])$ e se $f^{(n+1)}$ é uma função contínua em (a,b) então, para cada $x^* \in [a,b]$ existe um $\xi = \xi(x^*) \in (a,b)$ tal que*

$$e(x^*) = f(x^*) - \mathbb{P}_n(x^*) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^{n} (x^* - x_j). \tag{5.22}$$

Observação 5.2.1 *Na prática, o erro de interpolação é apresentado na forma*

$$|e(x^*)| = |f(x^*) - \mathbb{P}_n(x^*)| \leq \frac{M_{n+1}}{(n+1)!} \left| \prod_{j=0}^{n} (x^* - x_j) \right|, \tag{5.23}$$

onde

$$M_{n+1} = \max_{x \in [a,b]} \left| f^{(n+1)}(x) \right|.$$

O erro para o polinómio interpolador pode ser obtido por:

i) $|f(x) - p_n(x)| \leq \dfrac{1}{(n+1)!} \left| f^{(n+1)}(\xi)(x - x_0) \cdots (x - x_n) \right|$ se $f(x)$ é conhecida, para algum $\xi \in [a,b]$; ou por

ii) $|f(x) - p_n(x)| \leq M |(x - x_0) \cdots (x - x_n)|$, onde M é o maior módulo das diferenças de ordem $n+1$. Esta majoração é possível pois prova-se que

$$f_{0,\ldots,n} = \frac{f^{(n)}(\xi)}{n!}$$

Exemplo 5.2.1 *Determine uma estimativa para o erro que se cometeu na aproximação obtida em (5.11).*

Resolução 5.2.1 *Atendendo à fórmula (5.23), temos*

$$|e(x^*)| = |\log_{10}(x^*) - \mathbb{P}_3(x^*)| \leq \frac{M_4}{4!}(x^* - 2.3)(x^* - 2.4)(x^* - 2.5)(x^* - 2.6),$$

onde

$$M_4 = \max_{x\in[2.3,2.6]} \left|f^4(x)\right| = \max_{x\in[2.3,2.6]} \frac{6}{x^4 \ln 10} = 0.093116.$$

Fazendo $x^* = 2.45$*, temos*

$$|\log_{10}(2.45) - \mathbb{P}_3(2.45)| \leq \frac{0.093116}{24} |(2.45 - 2.3)(2.45 - 2.4)(2.45 - 2.5)(2.45 - 2.6)|,$$

isto é,

$$|e_3(2.54)| \leq 0.917 \times 10^{-5}$$

■

5.3 Interpolação polinomial linear e quadrática

O caso linear é o caso mais simples da interpolação. Dados dois pontos distintos de uma função $y = f(x)$, $(x_0, f(x_0))$ e $(x_1, f(x_1))$, e $x^* \in (x_0, x_1)$ pretendemos saber, usando a interpolação polinomial, o valor de $y^* = f(x^*)$. Pelo teorema anterior, vamos construir um polinómio de grau um,

$$P_1(x) = a_0 + a_1 x.$$

Mas, $P_1(x)$ tem de ser tal que:

$$\begin{cases} P_1(x_0) &= a_0 + a_1 x_0 = f(x_0) \\ P_1(x_1) &= a_0 + a_1 x_1 = f(x_1) \end{cases}. \tag{5.24}$$

É portanto necessário resolver o sistema de equações lineares (5.24). Este sistema pode ser representado na forma $AX = B$ onde

$$A = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \end{bmatrix}, X = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \text{ e } B = \begin{bmatrix} f(x_0) \\ f(x_1) \end{bmatrix} \equiv \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}.$$

Facilmente se prova que o sistema é possível e determinado uma vez que o determinante da matriz dos coeficientes é

$$\det(A) = |A| = x_1 - x_0.$$

O sistema anterior tem solução única se $\det(A) = x_1 - x_0 \neq 0$, isto é, se se $x_0 \neq x_1$. Ou seja, para pontos distintos o sistema tem solução única.

A solução é dada por $X = A^{-1}B$, isto é,

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \dfrac{x_1}{x_1 - x_0} & \dfrac{x_0}{x_0 - x_1} \\ \dfrac{1}{x_0 - x_1} & \dfrac{1}{x_1 - x_0} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} \dfrac{y_0 x_1 - y_1 x_0}{x_1 - x_0} \\ \dfrac{y_0 - y_1}{x_0 - x_1} \end{bmatrix}$$

Geometricamente, o polinómio $P_1(x)$ é a recta que passa nos pontos $(x_0, f(x_0))$ e $(x_1, f(x_1))$.

Portanto, o polinómio interpolador é da forma

$$\begin{aligned} P_1(x) &= a_0 + a_1 x \\ &= \frac{y_0 x_1 - y_1 x_0}{x_1 - x_0} + \left(\frac{y_0 - y_1}{x_0 - x_1}\right) x. \end{aligned}$$

Queremos agora determinar um polinómio único que passa por três pontos distintos $(x_0, f(x_0))$, $(x_1, f(x_1))$ e $(x_2, f(x_2))$. Procedendo como anteriormente, temos que impor as condições

$$\left\{ \begin{array}{lcl} a + bx_0 + cx_0^2 &=& f(x_0) \\ a + bx_1 + cx_1^2 &=& f(x_1) \\ a + bx_2 + cx_2^2 &=& f(x_2) \end{array} \right. \tag{5.25}$$

O sistema (5.25) pode ser apresentado na forma matricial $AX = B$, onde

$$A = \left[\begin{array}{ccc} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{array} \right], \quad X = \left[\begin{array}{c} a \\ b \\ c \end{array} \right] \text{ e } B = \left[\begin{array}{c} f(x_0) \\ f(x_1) \\ f(x_2) \end{array} \right]. \tag{5.26}$$

O sistema (5.25) tem solução única se e somente se o determinante da matriz dos coeficientes, a matriz A teem determinantes não nulo, isto é,

$$|A| = \left| \begin{array}{ccc} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{array} \right| \neq 0 \Rightarrow (x_2 - x_1)(x_2 - x_0)(x_1 - x_0) \neq 0. \tag{5.27}$$

Isto é, não podem existir dois pontos iguais. A solução do sistema $X = [a \;\; b \;\; c]^T$ é dada por

$$\begin{aligned} c &= \frac{1}{x_2 - x_1}\left(\frac{f(x_2) - f(x_0)}{x_2 - x_0} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}\right) \\ b &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} - c(x_1 + x_0) \\ a &= f(x_0) - bx_0 - cx_0^2 \end{aligned}$$

5.4 Método dos Coeficientes Indeterminados

Seja $f(x)$ uma função contínua num intervalo $[a, b] \subseteq \mathbb{R}$ e sejam $\{x_0, x_1, \ldots, x_n\}$ $(n+1)$ - pontos em $[a, b]$. Habitualmente, tem-se

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b. \tag{5.28}$$

Pretendemos encontrar um polinómio, $p_n(x)$ de grau menor ou igual que n,

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n \tag{5.29}$$

de tal forma que

$$p_n(x_i) = f(x_i), \quad \text{para } 0 \leq i \leq n, \tag{5.30}$$

isto é,

$$\begin{cases} p_n(x_0) &= a_0 + a_1 x_0 + a_2 x_0^2 + \cdots + a_n x_0^n = f(x_0) \\ p_n(x_1) &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_n x_1^n = f(x_1) \\ & \vdots \\ p_n(x_n) &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_n x_n^n = f(x_n) \end{cases} \tag{5.31}$$

que na forma matricial $AX = B$ é,

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^n \end{bmatrix}, \quad X = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \tag{5.32}$$

Nota 5.4.1 *A matriz*

$$\begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix}, \tag{5.33}$$

é designada por matriz de Vandermond. O determinante desta matriz é dado por:

$$\det(A) = \prod_{0 \leq i,j \leq n} (\alpha_j - \alpha_i). \tag{5.34}$$

A matriz de Vandermond aparece no problema de interpolação polinomial. Para um número elevado de pontos, o sistema torna-se bastante grande e além de ser trabalhoso, o problema é mal condicionado o que o torna pouco prático e de rara utilização.

Portanto, para obter os valores de a_0, a_1, $\ldots a_n$ é necessário resolver o sistema de equações lineares (5.32). Por exemplo, sabemos que a solução é da forma

$$X = A^{-1} B. \tag{5.35}$$

Exercício 5.4.1 *Uma função $f(x)$ é dada pela tabela*

x_i	0.1	0.2	0.3	0.4	0.5
$f(x_i)$	-0.135726	-0.0409365	0.0370409	0.100548	0.151633

.

Para determinar o polinómio interpolador, é necessário resolver o sistema de equações lineares $AX = B$, onde

$$A=\begin{bmatrix} 1 & (0.1)^1 & (0.1)^2 & (0.1)^3 & (0.1)^4 \\ 1 & (0.2)^1 & (0.2)^2 & (0.2)^3 & (0.2)^4 \\ 1 & (0.3)^1 & (0.3)^2 & (0.3)^3 & (0.3)^4 \\ 1 & (0.4)^1 & (0.4)^2 & (0.4)^3 & (0.4)^4 \\ 1 & (0.5)^1 & (0.5)^2 & (0.5)^3 & (0.5)^4 \end{bmatrix},\ X=\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}\ e\ B=\begin{bmatrix} -0.135726 \\ -0.0409365 \\ 0.0370409 \\ 0.100548 \\ 0.151633 \end{bmatrix} \tag{5.36}$$

cuja solução é

$$X=\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}=\begin{bmatrix} -0.249963 \\ 1.249126166 \\ -1.117601666 \\ 0.5126333333 \\ -0.1223333333 \end{bmatrix}. \tag{5.37}$$

Portanto, o polinómio interpolador de $f(x)$ é

$$p_4(x)=-0.249963+1.249126166x-1.117601666x^2+0.5126333333x^3-0.1223333333x^4. \tag{5.38}$$

5.5 Interpolação de Lagrange

Teorema 5.5.1 (Lagrange) *Seja f definida em $[a,b]\subseteq\mathbb{R}$ e $\{x_0,x_1,\ldots,x_n\}$ um conjunto de $n+1$ pontos distintos em $[a,b]$. O polinómio p_n de grau $\leq n$ interpolador de f nos pontos $x_0,x_1,\ldots,x_n$, existe e é único, sendo definido por*

$$\mathbb{P}_n(x)=a_0L_0(x)+a_1L_1(x)+\cdots+a_nL_n(x), \tag{5.39}$$

onde

$$a_i=f(x_i),\ \ L_i=\prod_{\substack{k=0\\k\neq i}}^{n}\frac{x-x_k}{x_i-x_k},\ \ (i=0,1,\ldots,n) \tag{5.40}$$

Devemos salientar que cada um dos polinómios $L_i(x)$ goza das seguintes propriedades:

- $L_i(x_j)=0$, para $i\neq j$;
- $L_i(x_i)=1$.

Teorema 5.5.2 (Erro de Interpolação) *Seja P_n o polinómio de grau menor ou igual a n, interpolador de $f(x)$ nos pontos $\{x_0,x_1,\ldots,x_n\}\in[a,b]$. Se $f\in\mathcal{C}^n([a,b])$ e se $f^{(n+1)}$ é uma função contínua em (a,b), então para cada $x\in[a,b]$ existe $\xi=\xi(\bar{x})\in(a,b)$ tal que*

$$e(b\bar{x})=|f(\bar{x})-p_n(\bar{x})|=\left|\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{j=0}^{n}(\bar{x}-x_j)\right|. \tag{5.41}$$

Nota 5.5.1 *Na prática, o erro de interpolação é calculado na forma*

$$e\left(b\bar{x}\right)=\left|f\left(\bar{x}\right)-p_{n}\left(\bar{x}\right)\right|\leq\frac{\max_{x\in[a,b]}\left|f^{(n+1)(x)}\right|\prod_{j=0}^{n}\left(\bar{x}-x_{j}\right)}{(n+1)!} \tag{5.42}$$

Exercício 5.5.1 *Uma função é dada pela tabela*

x	-2	1	2	4
$y\left(x\right)$	25	-8	-15	-23

a) Encontre o polinómio interpolador de Lagrange de $f\left(x\right)$.

b) Indique uma aproximação $f\left(0\right)$ e $f\left(3\right)$.

Resolução 5.5.1 *a) Neste caso, $0\leq n\leq3$ e temos os pontos $x_{0}=-2$, $x_{1}=1$, $x_{2}=2$, $x_{3}=4$, $y\left(x_{0}\right)=25$, $y\left(x_{1}\right)=-8$, $y\left(x_{2}\right)=-15$ e $y\left(x_{3}\right)=-23$. Portanto, (5.39) fica*

$$\mathbb{P}_{3}\left(x\right)=y\left(x_{0}\right)L_{0}\left(x\right)+y\left(x_{1}\right)L_{1}\left(x\right)+y\left(x_{2}\right)L_{2}\left(x\right)+y\left(x_{3}\right)L_{3}\left(x\right), \tag{5.43}$$

onde

$$\begin{aligned}
L_{0} &= \prod_{\substack{k=0\\k\neq0}}^{3}\frac{x-x_{k}}{x_{0}-x_{k}} & L_{1} &= \prod_{\substack{k=0\\k\neq1}}^{3}\frac{x-x_{k}}{x_{1}-x_{k}}\\
&= \frac{x-x_{1}}{x_{0}-x_{1}}\frac{x-x_{2}}{x_{0}-x_{2}}\frac{x-x_{3}}{x_{0}-x_{3}} & &= \frac{x-x_{0}}{x_{1}-x_{0}}\frac{x-x_{2}}{x_{1}-x_{2}}\frac{x-x_{3}}{x_{1}-x_{3}}\\
&= \frac{x-1}{-2-1}\frac{x-2}{-2-2}\frac{x-4}{-2-4} & &= \frac{x-(-2)}{1-(-2)}\frac{x-2}{1-2}\frac{x-4}{1-4}\\
&= -\frac{1}{72}\left(x-1\right)\left(x-2\right)\left(x-4\right), & &= \frac{1}{9}\left(x+2\right)\left(x-2\right)\left(x-4\right),
\end{aligned}$$

$$\begin{aligned}
L_{2} &= \prod_{\substack{k=0\\k\neq2}}^{3}\frac{x-x_{k}}{x_{2}-x_{k}} & L_{3} &= \prod_{\substack{k=0\\k\neq3}}^{3}\frac{x-x_{k}}{x_{3}-x_{k}}\\
&= \frac{x-x_{0}}{x_{2}-x_{0}}\frac{x-x_{1}}{x_{2}-x_{1}}\frac{x-x_{3}}{x_{2}-x_{3}} & &= \frac{x-x_{0}}{x_{3}-x_{0}}\frac{x-x_{1}}{x_{3}-x_{1}}\frac{x-x_{2}}{x_{3}-x_{2}}\\
&= \frac{x-1}{2-(-2)}\frac{x-2}{2-1}\frac{x-4}{2-4} & &= \frac{x+2}{4-(-2)}\frac{x-1}{4-1}\frac{x-2}{4-2}\\
&= -\frac{1}{8}\left(x+2\right)\left(x-1\right)\left(x-4\right), & &= \frac{1}{36}\left(x+2\right)\left(x-1\right)\left(x-2\right),
\end{aligned}$$

Portanto, (5.43) fica

$$\begin{aligned}\mathbb{P}_3(x) &= y(x_0)L_0(x) + y(x_1)L_1(x) + y(x_2)L_2(x) + y(x_3)L_3(x) \\ &= -\frac{25}{72}(x-1)(x-2)(x-4) - \frac{8}{9}(x+2)(x-2)(x-4) \\ &\quad +\frac{15}{8}(x+2)(x-1)(x-4) - \frac{23}{36}(x+2)(x-1)(x-2) \\ &= x^2 - 10x + 1 \equiv p_2(x) \end{aligned} \tag{5.44}$$

A representação gráfica de $p_2(x)$ é:

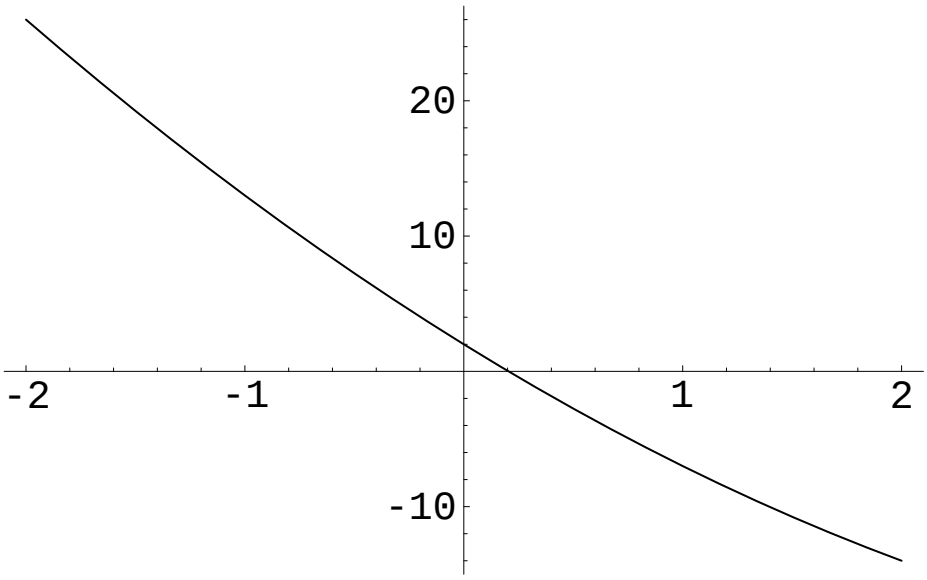

Figura 5.4: Gráfico da função $p(x) = x^2 - 10x + 2$.

Verifique que $\mathbb{P}_3(x_i) = f(x_i) = y_i$.

b) Tendo em conta (5.44),

$$f(0) \approx \mathbb{P}_3(0) = 1$$

e que

$$f(3) \approx \mathbb{P}_3(x) = -20.$$

■

Nota 5.5.2 *A ordem dos pontos não é importante. Uma ordem ascendente ou descendente torna o conjunto de pontos $\{x_0, x_1, \ldots, x_n\}$ mais legível.*

Nota 5.5.3 *Este problema também pode ser resolvido de outro modo: Pretende-se determinar o polinómio de 3 grau que passa pelos pontos $P_1(-2, 25)$, $P_2(1, -8)$, $P_3(2, -15)$ e $P_1(4, -23)$. Vamos supor que o polinómio é da forma*

$$p(x) = ax^3 + bx^2 + cx + d.$$

Então,

$$\left\{\begin{array}{rcr} p(-2) & = & 25 \\ p(1) & = & -8 \\ p(2) & = & -15 \\ p(4) & = & -23 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcr} -8a+4b-2c+d & = & 25 \\ a+b+c+d & = & -8 \\ 8a+4b+2c+d & = & -15 \\ 64a+16b+4c+d & = & -23 \end{array}\right.$$

cuja solução é

$$(a,b,c,d) = (0,1,-10,1) \tag{5.45}$$

Também obtemos o polinómio de segundo grau,

$$p(x) = x^2 - 10x + 1.$$

É fácil verificar que $p(-2) = 25$, $p(1) = -8$, $p(2) = -15$ *e que* $p(4) = -23$. *Com este polinómio obtemos os seguintes valores aproximados para* $f(0) \approx 1$ *e* $f(3) \approx -20$.

Exercício 5.5.2 *Determine aproximações de* $\cos\left(\frac{\pi}{8}\right)$ *usando os polinómios interpoladores de Lagrange de grau 2 e 4 no intervalo* $[0,\pi]$. *Compare os resultados obtidos e indique um majorante para o erro.*

Resolução 5.5.2 *Para aproximarmos a função* $f(x) = \cos(x)$ *no ponto* $x = \frac{\pi}{8}$, *minimizando o erro, temos de escolher pelo menos um ponto antes e um ponto depois. Assim podemos escolher os pontos* $x_0 = 0, x_1 = \frac{\pi}{2}$ *e* $x_2 = \pi$. *O polinómio é da forma*

$$\begin{aligned} p_2(x) =& f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x) \\ =& f(0)L_0(x) + f\left(\frac{\pi}{2}\right)L_1(x) + f(\pi)L_2(x) \end{aligned}$$

onde

$$\begin{aligned} L_0(x) =& \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} = \frac{\left(x-\frac{\pi}{2}\right)(x-\pi)}{\left(0-\frac{\pi}{2}\right)(0-\pi)} = \frac{2x^2-3\pi x+\pi^2}{\pi^2} \\ L_1(x) =& \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} = \frac{(x-0)(x-\pi)}{\left(\frac{\pi}{2}-0\right)\left(\frac{\pi}{2}-\pi\right)} = \frac{-4x^2+4\pi x}{\pi^2} \\ L_2(x) =& \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} = \frac{(x-0)\left(x-\frac{\pi}{2}\right)}{(\pi-0)\left(\pi-\frac{\pi}{2}\right)} = \frac{2x^2-\pi x}{\pi^2}. \end{aligned}$$

Substituindo no polinómio fica

$$p_2(x) = 1\times\frac{2x^2-3\pi x+\pi^2}{\pi^2} + 0\times\frac{-4x^2+4\pi x}{\pi^2} - 1\times\frac{2x^2-\pi x}{\pi^2} = -\frac{2}{\pi}x + 1$$

Logo

$$\cos\left(\frac{\pi}{8}\right) \approx p_2\left(\frac{\pi}{8}\right) = -\frac{2}{\pi}\frac{\pi}{8} + 1 = 0.75.$$

Por seu turno, $p_4\left(\frac{\pi}{8}\right) = 0.93121843353823$.

O valor exacto é $\cos\left(\frac{\pi}{8}\right) = 0.92387953251129$. *Depreende-se que a segunda aproximação é melhor que a primeira. Vamos de seguida, efectuar o cálculo do majorante:*

$$\begin{aligned}\left|f\left(\frac{\pi}{8}\right) - p_2\left(\frac{\pi}{8}\right)\right| &\leqslant \left|\frac{\cos^{(3)}(\xi)}{3!}\left(\frac{\pi}{8}-0\right)\left(\frac{\pi}{8}-\frac{\pi}{2}\right)\left(\frac{\pi}{8}-\pi\right)\right| \\ &= \left|\frac{\sin(\xi)}{6}\times\frac{21\pi^3}{512}\right| \leqslant \frac{1}{6}\times\frac{21\pi^3}{512} = 0.21195696949424\end{aligned}$$

■

Exemplo 5.5.1 *Considere a função* $f(x) = \cos(x)$, $x \in [0,\pi]$. *Determine o número de pontos a considerar em* $[0,\pi]$ *de tal forma a que o erro máximo da aproximação de* $f(x)$ *por um polinómio interpolador seja inferior a* $\frac{1}{2}$.

Resolução 5.5.3 *Para* $x \in [0,\pi]$,

$$|f(x) - \mathbb{P}_n(x)| \leq \frac{\max_{x\in[0,\pi]}\left|f^{(n+1)(x)}\right|}{(n+1)!}|w(x^*)| = \frac{|w(x)|}{(n+1)!} \leq \frac{\pi^{n+1}}{(n+1)!}.$$

Portanto, para determinar o valor de n que satisfaz a condição é necessário procurar n tal que

$$\frac{\pi^{n+1}}{(n+1)!} \leq \frac{1}{2}.$$

Verificamos que com $n = 7$, $\frac{\pi^8}{8!} \approx 0.23 < \frac{1}{2}$.

■

Apesar de o método de Lagrange ser eficiente para calcular o polinómio interpolador, ele possui uma desvantagem. Suponha que o polinómio $p_{n-1}(x)$ é calculado para os pontos $x_0,\ldots,x_{n-1}$, e que, por algum motivo, também seja necessário o cálculo com mais um ponto x_n. Para se calcular o polinómio interpolador $p_n(x)$ sobre os pontos $x_0,\ldots,x_{n-1},x_n$ pelo método de Lagrange, é necessário refazer todos os cálculos, uma vez que o grau dos $L_j(x)$ agora é diferente.

5.5.1 Interpolação de Lagrange segmentada

Em geral, o aumento do grau do polinómio implica um aumento da precisão. No entanto, existem funções para as quais não podemos afirmar que um aumento do grau do polinómio leva a um aumento da precisão, isto é, corresponda a um aumento da "proximidade"entre o polinómio interpolador e a função a interpolar, isto é, o valor de

$$|f(x) - p_n(x)|$$

para $x \in [a,b]$ é menor. Como já mencionámos anteriormente, o uso de um número elevado de pontos dá origem a polinómios de grau muito elevado que sofrem muitas oscilações.

Consideremos um intervalo $[a,b]$ que vamos dividir em k subintervalos de comprimento h_j, definidos por

$$I_j = [x_j, x_{j+1}]\,,\ j = 0, 1, \ldots, k-1.$$

Defina-se

$$h = \max_{0 \le j \le k-1} h_j.$$

Temos que

$$[a, b] = \cup_{j=0}^{k-1} I_j.$$

Podemos definir um polinómio interpolador de Lagrange em cada um dos intervalos I_j, de grau n usando $n+1$ pontos. No caso em que os pontos são igualmente espaçados, temos que

$$h = h_j,\ \ j = 0, 1, \ldots, k-1$$

Consideremos que os pontos são definidos por

$$\left\{x_j^{(i)} :\ 0 \le i \le n\right\}.$$

Para qualquer função f contínua no intervalo $[a, b]$ a interpolação segmentada coincide um cada I_j com o polinómio interpolador de $f|_{I_j}$ nos $n+1$ pontos

$$\left\{x_j^{(i)} :\ 0 \le i \le n\right\}.$$

O erro da interpolação segmentada é dado pelo

Teorema 5.5.3 *Sejam $f \in \mathcal{C}^{m+1}[a, b]$ e suponhamos que $[a, b] = \cup_{j=0}^{k-1} I_j$. Se p_n^j são os polinómios interpoladores de $f|_{I_j}$ nos $n+1$ pontos $\left\{x_j^{(i)} :\ 0 \le i \le n\right\}$, então, o erro da interpolação segmentada verifica*

$$\max_{0 \le j \le k-1} \left\|e_n^j\right\|_\infty \le \frac{1}{(n+1)!} \left\|f^{(n+1)}\right\|_\infty \max_{0 \le j \le k-1} \left\|\omega_{n+1}^j\right\|_\infty \tag{5.46}$$

onde $\omega_{n+1}^j = \prod_{i=0}^{n} \left(x - x_j^{(i)}\right)$.

5.6 O método de Newton

Vamos agora apresentar um exemplo utilizando o método de Newton, isto é, utilizando a base (5.6). A base

$$\pi_k(x) = \prod_{i=0}^{k-1} (x - x_i)\,,\ \text{para } k = 0, 1, \ldots, n$$

goza da seguinte propriedade: para $i < j$,

$$\pi_j(x_i) = 0,$$

pelo que a matriz dos coeficientes no sistema $AX = B$ é uma matriz triangular inferior.

Exemplo 5.6.1 *Utilizando o método de Newton, determine um polinómio $p_2(x)$ que interpole uma função $f(x)$ desconhecida da qual sabemos que*

x_i	-2	0	1
$f(x_i)$	-27	-1	10

Resolução 5.6.1 *Para este problema, o sistema linear $AX = B$ onde*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & x_1 - x_0 & 0 \\ 1 & x_2 - x_0 & (x_2 - x_0)(x_2 - x_1) \end{bmatrix}, \; X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \; e \; B = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \end{bmatrix},$$

o que neste caso corresponde às martizes

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 3 \end{bmatrix}, X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \; e \; B = \begin{bmatrix} -27 \\ -1 \\ 0 \end{bmatrix},$$

cuja solução é

$$X = \begin{bmatrix} -27 \\ 13 \\ -4 \end{bmatrix}.$$

Portanto, o polinómio é

$$\begin{aligned} p_2(x) &= -27 + 13(x+2) - 4(x+2)(x-0) \\ &= -1 + 5x - 4x\ 2. \end{aligned}$$

polinewton

■

5.7 Diferenças divididas

Vamos de seguida apresentar um outro modo de calcular o polinómio interpolador que exige um menor esforço computacional.

Definição 5.7.1 (Diferenças divididas) *Seja f definida em $[a,b] \subseteq \mathbb{R}$ e sejam $x_0, x_1, \ldots, x_n$ pontos distintos desse intervalo. O quociente,*

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}, \tag{5.47}$$

é designado por diferença dividida de primeira ordem *de f relativamente aos argumentos x_{i+1} e x_i. As diferenças divididas de ordem superior definem-se de forma recursiva. Deste modo, define-se* diferença dividida de ordem *k relativamente aos argumentos x_i, x_{i+1}, ..., x_{i+k}, com $i + k < n$, por*

$$f\left[x_i, x_{i+1}, \ldots, x_{i+k}\right] = \frac{f\left[x_i, x_{i+1}, \ldots, x_{i+k}\right] - f\left[x_i, x_{i+1}, \ldots, x_{i+k-1}\right]}{x_{i+k} - x_i}, . \tag{5.48}$$

Vamos representar a diferença $f\left[x_i, x_{i+1}, \ldots, x_{i+j}\right]$ *por* $f_{i,i+j}$.

Habitualmente, as diferenças divididas são obtidas pela tabela das diferenças divididas:

x_i	$f(x_i)$	$f(x_{i+1})$	$f(x_{i+2})$	$f(x_{i+3})$	$\cdots$
x_0	f_0				
		$f_{0,1}$			
x_1	f_1		$f_{0,2}$		
		$f_{1,2}$		$f_{0,3}$	
x_2	f_2		$f_{1,3}$		$\cdots$
		$f_{2,3}$		$\cdots$	
x_3	f_3		$\cdots$		$\cdots$
		$\cdots$		$f_{n-3,n}$	
$\cdots$	$\cdots$		$f_{n-2,n}$		
		$f_{n-1,n}$			
x_n	f_n				

As diferencças divididas também estão definidas para pontos não distintos, de acordo com a seguinte formula:

$$f\left[x_i, x_i, \ldots, x_i\right] = \frac{f^{(r)}(x_i)}{r!}, \tag{5.49}$$

onde $x_i, x_i, \ldots, x_i$ correspondem a $r+1$ pontos.

Note que

$$\begin{aligned} f\left[x_i, x_i\right] &= \lim_{x \to x_i} f\left[x, x_i\right] \\ &= \lim_{x \to x_i} \frac{f(x) - f(x_i)}{x - x_i} = f'(x_i) \end{aligned}$$

Se trocarmos a ordem dos argumentos das diferenças divididas, por exemplo

$$f\left[x_{i+1}, x_i\right] = \frac{f(x_i) - f(x_{i+1})}{x_i - x_{i+1}} = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = f\left[x_i, x_{i+1}\right]$$

obtemos o mesmo valor.

Prova-se que o valor das diferenças divididas não se altera se permutarmos os argumentos entre si. Um resultado muito importante é o

Teorema 5.7.1 *As diferenças divididas são invariantes para qualquer permutação dos índices de suporte.*

Como $p_n(x)$ é um polinómio interpolador, o *erro* é dado por:

$$|f(x) - p_n(x)| = \left| \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x - x_0)\dots(x - x_n) \right|$$

se $f(x)$ é conhecida.

Caso $f(x)$ não seja conhecida ou $f^{(n+1)}(x)$ seja muito difícil de obter, é possível obter um majorante para o erro pois:

$$f_{[0,\dots,n]} = \frac{f^{(n)}(\xi)}{n!} \quad \text{para} \quad \xi \in [a,b].$$

Assim, se $f(x)$ for desconhecida podemos considerar

$$|f(x) - p_n(x)| \leqslant |M(x - x_0)\dots(x - x_n)|$$

onde M é o maior módulo das diferenças divididas de ordem $n+1$, i.e, $M = |f[x_0, x_1, \dots, x_{n+1}]|$.

5.8 Fórmula de Newton para diferenças divididas

Tendo por objectivo diminuir o esforço computacional aquando da determinação do polinómio interpolador, vamos definir o polinómio interpolador de $f(x)$ nos pontos x_0, x_1, x_2 ,..., x_n de grau menor ou igual que n por

$$\mathbb{P}_n = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n \prod_{k=0}^{n-1}(x - x_k), \tag{5.50}$$

onde os a_i, $i = 1, \dots, n$ são constantes apropriadas a determinar. Coloca-se então o problema de determinar cada um dos a_i.

Para determinar a_0, devemos ter em conta que, se $\mathbb{P}_n(x)$ pode ser escrito na forma (5.50) então,

$$a_0 = \mathbb{P}_n(x_0) = f(x_0).$$

Do mesmo modo, obtemos que

$$f(x_0) + a_1(x - x_1) = \mathbb{P}_n(x_1) = f(x_1) \Rightarrow a_1 = f[x_0, x_1].$$

De forma análoga, provamos que

$$a_i = f[x_0, x_1, \dots, x_i], \quad i = 1, 2, \dots, n. \tag{5.51}$$

Então, o polinómio interpolador de Lagrange de f nos pontos x_0, x_1, x_2 ,..., x_n pode ser escrito na forma

$$\begin{aligned}\mathbb{P}_n(x) &= f(x_0) + f[x_0, x_1](x - x_0) + f(x_0, x_1, x_2)(x - x_0)(x - x_1) \\ &\quad + \cdots + f[x_0, x_1, x_2, \ldots, x_n](x - x_0)(x - x_1)\cdots(x - x_{n-1}) \\ &= f(x_0) + \sum_{i=1}^{n} f[x_0, x_1, \ldots, x_i] \prod_{j=0}^{i-1}(x - x_j) \qquad (5.52)\end{aligned}$$

A fórmula (5.52) é designada por fórmula interpoladora de Newton das diferenças divididas.

Observação 5.8.1 *a) os coeficientes da fórmula de Newton estão ao longo da diagonal das tabela das diferenças divididas;*

b) uma vez determinado $\mathbb{P}_n(x)$*, para determinar* $\mathbb{P}_{n+1}(x)$ *basta fazer:*

$$\mathbb{P}_{n+1}(x) = \mathbb{P}_n(x) + f[x_0, x_1, \ldots, x_{n+1}] \prod_{j=0}^{n}(x - x_j). \qquad (5.53)$$

c)

O erro para o polinómio interpolador pode ser obtido por:

$$|f(x) - p_n(x)| \leq M\,|(x - x_0)\cdots(x - x_n)|,$$

onde M é o maior módulo das diferenças de ordem $n+1$. Esta majoração é possível pois prova-se que

$$f_{0,\ldots,n} = \frac{f^{(n)}(\xi)}{n!}$$

Teorema 5.8.1 *Seja* f *definida em* $[a, b] \subseteq \mathbb{R}$ *e sejam* $x_0, x_1, \ldots, x_n$ *pontos distintos desse intervalo. Então,*

$$f[x_0, x_1, \ldots, x_n] = \sum_{i=0}^{n} \frac{f(x_i)}{w'(x_i)}, \qquad (5.54)$$

onde $w(x) = \prod_{j=0}^{n}(x - x_j)$.

Teorema 5.8.2 (Valor intermédio de Lagrange generalizado) *: Seja* $f \in \mathcal{C}^2([a, b])$ *uma função conhecida nos pontos (distintos)* $x_0, x_1, x_2, \ldots, x_n \in [a, b]$*. Então, existe um* $\xi \in (a, b)$ *tal que*

$$f[x_0, x_1, x_2, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}. \qquad (5.55)$$

O teorema anterior permite-nos concluir que, na ausência de informação sobre $f^{(n+1)}$, uma boa estimativa para o erro de interpolação pode ser dada por

$$f[x_0, x_1, x_2, \ldots, x_n] \times \prod_{j=0}^{n} (x - x_j) \tag{5.56}$$

desde que as diferenças de ordem $(n+1)$ não variem muito.

Exemplo 5.8.1 *Considere a tabela*

x_i	1	-1	-2
$f(x_i)$	0	-3	-4

Determine uma aproximação para $f(0)$ utilizando interpolação quadrática.

Resolução 5.8.1 *A tabela de diferenças divididas é*

x_i	$f(x_i)$	$f_{i,i+1}$	$f_{i,i+2}$
-2	-4		
		$\frac{-3-(-4)}{-1-(-2)} = \frac{-3+4}{1} = 1$	
-1	-3		$\frac{\frac{3}{2}-(1)}{1-(-2)} = \frac{1}{6}$
		$\frac{0-(-3)}{1-(-1)} = \frac{3}{2}$	
1	0		

Portanto,

$$\mathbb{P}_2(x) = -4 + 1(x-(-2)) + \frac{1}{6}(x+1)(x+2) = -2 + x + \frac{1}{6}(x+2)(x+1).$$

Consequentemente, o polinómio interpolador de f é dado por

$$p_2(x) = -2 + x + \frac{1}{6}(x+2)(x+1). \tag{5.57}$$

$$f(0) \approx \mathbb{P}_2(0) = -\frac{5}{3}.$$

Nota 5.8.1 *Devemos salientar que apresentando os elementos da tabela $(x_i, f(x_i))$ por ordem decrescente, obtemos o mesmo polinómio.*

x_i	$f(x_i)$	$f_{i,i+1}$	$f_{i,i+2}$
1	0		
		$\frac{-3-0}{-1-(1)} = \frac{-3}{-2} = \frac{3}{2}$	
-1	-3		$\frac{\frac{3}{2}-(1)}{1-(-2)} = \frac{1}{6}$
		$\frac{-4-(-3)}{-2-(-1)} = \frac{-1}{-1} = 1$	
-2	-4		

Portanto, o polinómio interpolador de $f(x)$ nos pontos indicados é

$$\begin{aligned} P_2(x) &= 0 + \frac{3}{2}(x-1) + \frac{1}{6}(x-1)(x+1) \\ &= (x-1)\left(\frac{3}{2} + \frac{1}{6}(x+1)\right). \end{aligned}$$

■

Exemplo 5.8.2 *Considere a seguinte tabela de valores de uma função real de variável real, $f(x)$:*

x_i	0	1	2	4
$y(x_i)$	2	2	3	6

Determine o polinómio interpolador da função $f(x)$, de grau 3, usando a tabela das diferenças divididas e determine um majorante para o erro cometido.

Resolução 5.8.2 *A tabela das diferenças finitas fica, então:*

$$\begin{aligned} f[x_0,x_1] &= \frac{f(x_1)-f(x_0)}{x_1-x_0} = \frac{2-2}{1-0} = 0 \\ f[x_1,x_2] &= \frac{f(x_2)-f(x_1)}{x_2-x_1} = \frac{3-2}{2-1} = 1 \\ f[x_2,x_3] &= \frac{f(x_3)-f(x_2)}{x_3-x_2} = \frac{6-3}{4-2} = \frac{3}{2} \\ f[x_0,x_1,x_2] &= \frac{f[x_1,x_2]-f[x_0,x_1]}{x_2-x_0} = \frac{1-0}{2-0} = \frac{1}{2} \\ f[x_1,x_2,x_3] &= \frac{f[x_2,x_3]-f[x_1,x_2]}{x_3-x_1} = \frac{\frac{3}{2}-1}{4-1} = \frac{1}{6} \\ f[x_0,\ldots,x_3] &= \frac{f[x_1,x_2,x_3]-f[x_0,x_1,x_2]}{x_3-x_0} = \frac{\frac{1}{6}-\frac{1}{2}}{4-0} = -\frac{1}{12} \end{aligned}$$

Isto é,

x_i	$f(x_i)$	$f[x_i,x_{i+1}]$	$f[x_i,\ldots,x_{i+2}]$	$f[x_i,\ldots,x_{i+3}]$
0	2			
		0		
1	2		$\frac{1}{2}$	
		1		$-\frac{1}{12}$
2	3		$\frac{1}{6}$	
		$\frac{3}{2}$		
4	6			

$$\begin{aligned}\mathbb{P}_3(x) &= 2 + 0(x-0) + \frac{x(x-1)}{2} - \frac{x(x-1)(x-2)}{12} \\ &= 2 + \frac{x(x-1)}{2} - \frac{x(x-1)(x-2)}{12}.\end{aligned}$$

■

Exemplo 5.8.1 *Os seguintes dados são relativos à função real de variável real* $f(x) = xe^{-1}$.

x_i	0	0.1	0.2	0.3
$f(x_i)$	0	0.094837	0.163746	0.22245

a) Determine o polinómio interpolador de Newton para a função $f(x)$.

b) Determine uma aproximação para $f(0.15)$.

c) Determine, no intervalo $[0,1]$ *a solução da equação* $xe^{-x} = 0.15$

Resolução 5.8.3

x_i	$f(x_i)$	$f[x_{i+1}, x_i]$	$f[x_{i+2}, x_{i+1}, x_i]$	$f[x_{i+3}, \ldots, x_i]$
0.0	0			
		$\frac{0.094837-0}{0.1-0.0} = 094837$		
0.1	0.0904837		$\frac{0.732623-094837}{0.2-0.0} = -0.861075$	
		$\frac{0.163746-0.094837}{0.2-0.1} = 0.732623$		$\frac{-0.738165-(-0.861075)}{0.3-0.0} = 0.4097$
0.2	0.163746		$\frac{0.58499-0.732623}{0.3-0.1} = -0.738165$	
		$\frac{0.22245-0.163746}{0.3-0.2} = 0.58499$		
0.3	0.22245			

Logo, o polinómio interpolador de $f(x)$ *nos pontos indicados é:*

$$\begin{aligned} p_3(x) &= 0 + 094837(x-0.0) - 0.861075x(x-0.1) + 0.4097 + 0.4097x(x-0.1)(x-0.2) \\ &= 0.9991385x - 0.9839850000000001x^2 + 0.4097x^3 \end{aligned}$$

e $f(0.15) \approx p_3(0.15) = 0.129114$. *O valor exacto do erro é*

$$|f(0.15) - p_3(0.15)| = |0.129106 - 0.129114| = 7.653536241325476^{-6}.$$

■

5.8.1 Interpolação polinomial segmentada

Nesta secção vamos utilizar polinómios segmentados contínuos para interpolar uma função num conjunto de pontos.

Um polinómio segmentado contínuo é uma função polinomial contínua formada pela "ligação"de "segmentos"de vários polinómios diferentes. As abcissas dos pontos de ligação são designados por "nós".

Seja f uma função real de variável real definida em $[a,b] \subset \mathbb{R}$, tabulada para os valores do argumento $x_0,\, x_1,\, \ldots,\, x_n \in [a,b]$.

O exemplo mais simples de interpolação polinomial segmentada é quando a função interpoladora è constituída por polinómios com grau menor ou igual que um, em cada subintervalo $[x_i, x_{i+1}]$. Tal função que denotaremos por $q(x)$ pode ser definida por

$$q_i(x) = f(x_i) + (x - x_i) f[x_i, x_{i+1}]. \tag{5.58}$$

5.9 Interpolação de Hermite

O objectivo da interpolação de Hermite é representar uma função por um polinómio interpolando a função e as suas derivadas até certa ordem nalguns pontos do seu domínio.

Procuramos o polinómio h interpolador de uma função f e da sua primeira derivada f', em $n+1$ pontos distintos $x_0, x_1, \ldots, x_n$. Isto é, h deve satisfazer as seguintes $2n+2$ condições:

$$h(x_i) = f(x_i), \;\; h'(x_i) = f'(x_i), \text{ para } \{x_0, x_1, \ldots, x_n\} \tag{5.59}$$

Teorema 5.9.1 (Hermite) *Seja $f \in \mathcal{C}^{2n+2}([a,b])$ e $\{x_0, x_1, \ldots, x_n\}$ um conjunto de $n+1$ pontos distintos em $[a,b]$. O polinómio h_{2n+1} de grau menor ou igual que $2n+1$ interpolador de f e f' nos pontos $x_0, x_1, \ldots, x_n$, existe e é único, sendo definido por*

$$\begin{aligned} h_{2n+1}(x) = a_0 H_0(x) + a_1 H_1(x) + \cdots + a_n H_n(x) + \bar{a}_0 H_0(x) \\ + \bar{a}_1 \overline{H}_1(x) + \cdots + \bar{a}_n \overline{H}_n(x), \end{aligned} \tag{5.60}$$

onde

$$a_i = f(x_i), \; \bar{a}_i = f'(x_i), \; H_i = \left(1 - 2L_i'(x_i)(x - x_i)\right) [L_i(x_i)]^2, \tag{5.61}$$

$$\overline{H}_i(x) = (x - x_i)[L_i(x)]^2, \tag{5.62}$$

com

$$L_i(x) = \frac{\prod_{\substack{j=0 \\ j \neq i}}^{n} (x - x_j)}{\prod_{\substack{j=0 \\ j \neq i}}^{n} (x_i - x_j)}, \;\; (i = 0, 1, \ldots, n) \tag{5.63}$$

Teorema 5.9.2 (Erro na interpolação de Hermite) *Seja p_{2n+1} o polinómio de grau menor ou igual a $2n+1$, interpolador de Hermite da função f nos pontos distintos $x_0, x_1, \ldots, x_n \in [a,b]$. Se $f \in \mathcal{C}[a,b]$ então para cada $x \in [a,b]$, existe $\xi \in [a,b]$ tal que*

$$e(x) = f(x) - p_{2n+1}(x) = \frac{f^{(2n+2)}(\xi)}{(2n+2)!}\omega^2(x) \tag{5.64}$$

onde $\omega = \prod_{j=0}^{n}(x - x_j)$.

Exemplo 5.9.1 *Considere a seguinte tabela relativa a uma função* $f(x)$*:*

i	x_i	$f(x_i)$	$f'(x_i)$
0	1.3	0.6200860	−0.5220232
1	1.6	0.4554022	−0.5698959
2	1.9	0.2818186	−0.5811571

Utilizando os polinómios de Hermite, determine uma aproximação para $f(1.5)$.

Resolução 5.9.1 *Em primeiro lugar, vamos determinar os polinómios de Lagrange e as suas derivadas:*

$$\begin{aligned}
L_{2,0}(x) &= \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} = \frac{50}{9}x^2 - \frac{175}{9}x + \frac{152}{9},\\
L'_{2,0}(x) &= \frac{100}{9}x - \frac{175}{9},\\
L_{2,1}(x) &= \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_2-x_2)} = -\frac{100}{9}x^2 + \frac{320}{9}x - \frac{247}{9}\\
L'_{2,1}(x) &= -\frac{200}{9}x + \frac{320}{9},\\
L_{2,2}(x) &= \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} = \frac{50}{9}x^2 - \frac{145}{9}x + \frac{104}{9},\\
L'_{2,2}(x) &= \frac{100}{9}x - \frac{145}{9}.
\end{aligned}$$

Os polinómios de Hermite, $H_{2,j}(x)$ *e* $\overline{H}_{2,j}(x)$

$$\begin{aligned}
H_{2,0}(x) &= [1 - 2(x-1.3)(-5)]\left(\frac{50}{9}x^2 - \frac{175}{9}x + \frac{152}{9}\right)^2\\
&= (10x - 12)\left(\frac{50}{9}x^2 - \frac{175}{9}x + \frac{152}{9}\right)^2,\\
H_{2,1}(x) &= 1 \cdot \left(-\frac{100}{9}x^2 + \frac{320}{9}x - \frac{247}{9}\right)^2,\\
H_{2,2}(x) &= 10(2-x)\left(\frac{50}{9}x^2 - \frac{145}{9}x + \frac{104}{9}\right)^2.
\end{aligned}$$

Os polinómios $\overline{H}_{2,j}(x)$ *são dados por*

$$\begin{aligned}
\overline{H}_{2,0}(x) &= (x-1.3)\left(\frac{50}{9}x^2-\frac{175}{9}x+\frac{152}{9}\right)^2,\\
\overline{H}_{2,1}(x) &= (x-1.6)\left(-\frac{100}{9}x^2+\frac{320}{9}x-\frac{247}{9}\right)^2,\\
\overline{H}_{2,2}(x) &= (x-1.9)\left(\frac{50}{9}x^2-\frac{145}{9}x+\frac{104}{9}\right)^2.
\end{aligned}$$

Portanto,

$$\begin{aligned}
H_5(x) &= 0.620086H_{2,0}(x)+0.4554022H_{2,1}(x)+0.2818186H_{2,2}(x)\\
&\quad -0.5220232\overline{H}_{2,0}(x)-0.5698959\overline{H}_{2,1}(x)-0.5811571\overline{H}_{2,2}(x). \quad (5.65)
\end{aligned}$$

e portanto,

$$\begin{aligned}
H_5(1.5) &= 0.620086\left(\frac{4}{27}\right)+0.4554022\left(\frac{64}{81}\right)+0.2818186\left(\frac{5}{81}\right)\\
&\quad -0.5220232\left(\frac{4}{405}\right)-0.5698959\left(\frac{-32}{405}\right)-0.5811571\left(\frac{-2}{405}\right)\\
&= 0.5118277. \quad (5.66)
\end{aligned}$$

■

A obtenção do polinómio interpolador de Hermite pode ser feita de várias maneiras. Também pode ser obtido numa forma que generaliza o polinómio interpolador de Newton das diferenças divididas. Consideremos os $2n+2$ pontos $x_0, x_0, x_1, x_1, \ldots, x_n, x_n$. Podemos verificar que o polinómio de grau $2n+1$ é dado por

$$H_{2n+1}(x) = f[z_0] + \sum_{k=1}^{2n+1} f[z_0,\ldots,z_k](x-z_0)(x-z_1)\cdots(x-z_{k-1}), \quad (5.67)$$

isto é,

$$\begin{aligned}
p(x) &= f(x_0)+f[x_0,x_0](x-x_0)+f[x_0,x_0,x_1](x-x_0)^2\\
&\quad +f[x_0,x_0,x_1,x_1](x-x_0)^2(x-x_1)\\
&\quad +\cdots+f[x_0,x_0,\ldots,x_n,x_n](x-x_0)^2(x-x_1)^2\cdots(x-x_{n-1})(x-x_n) \quad (5.68)
\end{aligned}$$

As diferenças divididas representadas estão generalizadas para pontos não distintos de acordo com o seguinte resultado:

$$f[x_i,x_i,\ldots,x_i] = \frac{f^{(r)}}{r!}, \quad (5.69)$$

onde $x_i, x_i, \ldots, x_i$ corresponde a $r+1$ pontos.

Note que

$$\begin{aligned} f[x_i, x_i] &= \lim_{x \to x_i} f[x, x_i] \\ &= \lim_{x \to x_i} \frac{f(x) - f(x_i)}{x - x_i} = f'(x_i) \end{aligned}$$

As diferenças são obtidas de acordo com a seguinte tabela:

z	$f(z)$	Primeira ordem	Segunda ordem
$z_0 = x_0$	$f[z_0] = f(x_0)$		
		$f[z_0, z_1] = f'(x_0)$	
$z_1 = x_0$	$f[z_1] = f(x_0)$		$f[z_0, z_1, z_2] = \frac{f[z_1, z_2] - f[z_0, z_1]}{z_2 - z_0}$
		$f[z_1, z_2] = \frac{f[z_2] - f[z_1]}{z_2 - z_1}$	
$z_2 = x_1$	$f[z_2] = f(x_1)$		$f[z_1, z_2, z_3] = \frac{f[z_2, z_3] - f[z_1, z_2]}{z_3 - z_1}$
		$f[z_2, z_3] = f'(x_1)$	
$z_3 = x_1$	$f[z_3] = f(x_1)$		$f[z_2, z_3, z_4] = \frac{f[z_3, z_4] - f[z_2, z_3]}{z_4 - z_2}$
		$f[z_3, z_4] = \frac{f[z_4] - f[z_3]}{z_4 - z_3}$	
$z_4 = x_2$	$f[z_4] = f(x_2)$		$f[z_3, z_4, z_5] = \frac{f[z_4, z_5] - f[z_3 - z_4]}{z_5 - z_3}$
		$f[z_4, z_5] = f'(x_2)$	
$z_5 = x_2$	$f[z_5] = f(x_2)$		

Exemplo 5.9.2 (O exemplo 5.9.1) *Utilizando a fórmula (5.68), determine o polinómio interpolador de Hermite para os dados do problema 5.9.1.*

Resolução 5.9.2 *A tabela de diferenças divididas para este problema é:*

x_i	$f(x_i)$	$1ordem$	$2ordem$	$3ordem$	$4ordem$	$5ordem$
1.3	0.6200860					
		-0.5220232				
1.3	0.6200860		-0.0897427			
		-0.5489460		0.0663657		
1.6	0.4554022		-0.0698330		0.0026663	
		-0.59698959		0.0679655		-0.0027738
1.6	0.4554022		-0.0290537		0.0010020	
		-0.5786120		0.0685667		
1.9	0.2818186		-0.0084837			
		-0.5811571				
1.9	0.2818186					

Portanto, o polinómio de Hermite que interpola a função $f(x)$ é dado por

$$\begin{aligned} H_5(x) &= 0.6200860 - 0.5220232(x - 1.3) - 0.0897427(x - 1.3)^2 \\ &\quad +0.0663657(x - 1.3)^2(x - 1.6) + 0.0026663(x - 1.3)^2(x - 1.6)^2 \\ &\quad -0.0027738(x - 1.6)^2(x - 1.6)^2(x - 1.9) \end{aligned}$$

e portanto, $H_5(1.5)$ *é dado por*

$$\begin{aligned} H_5(1.5) &= 0.6200860 - 0.5220232(1.5-1.3) - 0.0897427(1.5-1.3)^2 \\ &\quad +0.0663657(1.5-1.3)^2(1.5-1.6) + 0.0026663(1.5-1.3)^2(1.5-1.6)^2 \\ &\quad -0.0027738(1.5-1.6)^2(1.5-1.6)^2(1.5-1.9) \\ &= 0.5118277 \end{aligned}$$

■

Exemplo 5.9.3 *Determine o polinómio interpolador de Hermite de grau* 3 *para a função* $f(x) = \sin(x)$*, para* $x \in \left[0, \frac{\pi}{2}\right]$*. Indique uma aproximação para* $\sin\left(\frac{\pi}{4}\right)$ *e indique o erro absoluto (exacto) e um majorante para o erro absoluto.*

Resolução 5.9.3 *Como* $H_n(x)$ *tem grau* $2n+1$ *e pretendemos um polinómio de terçeiro grau que interpole* $f(x)$ *temos que* $2n+1=3$ *o que implica que* $n=1$*. Isto é, necessitamos de considerar* 2 *pontos* x_0 *e* x_1*. Vamos considerar os extremos do intervalo, isto é,* $x_0 = 0$ *e* $x_1 = \frac{\pi}{2}$*. Se* $f(x) = \sin(x)$ *então* $f'(x) = \cos(x)$*. Neste caso temos que* $f(0) = 0$ *e* $f'(0) = 1$ *e* $f\left(\frac{\pi}{2}\right) = 1$ *e* $f'\left(\frac{\pi}{2}\right) = 0$*.*

A tabela das diferenças finitas generalizadas é dada por

x_i	$f(x_i)$			
0	0			
		1		
0	0		$\frac{4-2\pi}{\pi^2}$	
		$\frac{2}{\pi}$		$\frac{-16+4\pi}{\pi^3}$
$\frac{\pi}{2}$	1		$-\frac{4}{\pi^2}$	
		0		
$\frac{\pi}{2}$	1			

Logo,

$$\begin{aligned} p_3(x) &= x + \frac{4-2\pi}{\pi^2}x^2 + \frac{-16+4\pi}{\pi^3}x^2\left(x - \frac{\pi}{2}\right) \\ &= x\left(1 + x\left(-0.231 - 0.921\left(x - \frac{\pi}{2}\right)\right)\right) \end{aligned}$$

Facilmente se verifica que:

$$\begin{aligned} p_3(0) &= 0 = \sin(0) \\ p_3\left(\frac{\pi}{2}\right) &= 1 = \sin\left(\frac{\pi}{2}\right). \end{aligned}$$

A derivada é

$$p_3'(x) = 1 + \frac{8-4\pi}{\pi^2}x + \frac{-16+4\pi}{\pi^3}\left(3x^2 - \pi x\right).$$

Facilmente se verifica que:

$$\begin{aligned} p_3'(0) &= 1 = \cos(0) \\ p_3'\left(\frac{\pi}{2}\right) &= 0 = \cos\left(\frac{\pi}{2}\right). \end{aligned}$$

Vamos agora determinar uma aproximação para $\sin\left(\frac{\pi}{4}\right)$. *Temos então*

$$\sin\left(\frac{\pi}{4}\right) \approx p_3\left(\frac{\pi}{4}\right) = 0.69635. \tag{5.70}$$

Portanto, o erro absoluto é:

$$\left|\sin\left(\frac{\pi}{4}\right) - p_3\left(\frac{\pi}{4}\right)\right| = 0.0107572. \tag{5.71}$$

A majoração teórica para o erro é dada por

$$e_a\left(\frac{\pi}{4}\right) = \left|\sin\left(\frac{\pi}{4}\right) - p_3\left(\frac{\pi}{4}\right)\right| \leq \max_{x\in\left[0,\frac{\pi}{2}\right]} \left|\frac{\frac{d^4}{dx^4}\sin(x)}{4!}\left(\frac{\pi}{4}-0\right)\left(\frac{\pi}{4}-\frac{\pi}{2}\right)\right| \approx 0.0257020947945. \tag{5.72}$$

■

Exemplo 5.9.4 *Construa o polinómio de Hermite de* 3 *grau para a função* $g(x)$ *definida pela tabela:*

x	1	2	3	4
$g(x)$	0	$\frac{15}{2}$	$\frac{80}{3}$	$\frac{255}{4}$
$g'(x)$	4	$\frac{49}{4}$	$\frac{244}{9}$	$\frac{769}{16}$

e obtenha uma aproximação para o valor de $g(1.5)$.

Resolução 5.9.4 *Dado um conjunto com* $n+1$ *pontos, o polinómio de Hermite que lhe corresponde tem grau igual a* $2n+1$. *Consequentemente, se pretendemos obter o polinómio de Hermite de grau* 3, *temos que* $2n+1=3$, *donde,* $n=1$. *Isto é, necessitamos dos pontos* x_0 *e* x_1. *Como pretendemos determinar uma aproximação para* $g(1.5)$, *vamos escolher* $x_0=1$ *e* $x_1=2$ *como pontos de interpolação. A tabela de interpolação é dada por*

z_i	$g(z_i)$	$g(z_i, z_{i+1})$	$g[z_i,\ldots,z_{i+2}]$	$g[z_i,\ldots,z_{i+3}]$
1	0			
		$f[x_0,x_0] = f'(x_0) = 4$		
1	0		$\frac{7}{2}$	
		$f[x_1,x_0] = \frac{15}{2}$		$\frac{5}{4}$
2	$\frac{15}{2}$		$\frac{19}{4}$	
		$f[x_1,x_1] = f'(x_1) = \frac{49}{4}$		
2	$\frac{15}{2}$			

Então, o polinómio pedido é dado por

$$H_3(x) = 0 + 4(x-1) + \frac{7}{2}(x-1)^2(x-2) = \frac{5}{4}x^3 - \frac{3}{2}x^2 + \frac{13}{4}x - 3 \tag{5.73}$$

e uma aproximação para $g(1.5)$ *é dada por*

$$g(1.5) \approx H_3(1.5) = \frac{87}{32} = 2.71875. \tag{5.74}$$

■

5.10 Splines Cúbicos

A spline linear apresenta a desvantagem de ter derivada primeira descontínua nos nós.

Se a função $f(x)$ está tabelada em $(n+1)$ pontos e a aproximarmos de grau n que a interpola sobre os pontos tabelados, o resultado dessa aproximação pode ser desastroso. Uma alternativa é interpolar $f(x)$ em grupos de poucos pontos, obtendo-se polinômio de grau menor, e impor condições para que a função de aproximação seja contínua e tenha derivadas contínuas até uma certa ordem.

O nome *spline* vem de uma régua elástica, usada em desenhos de engenharia, que pode ser curvada de forma a passar por um dado conjunto de pontos (x_i, y_i), que tem o nome de *spline*. Sob certas hipóteses (de acordo com a teoria da elasticidade) a curva definida pela régua pode ser descrita aproximadamente como sendo uma função por partes, cada qual um polinômio cúbico, de tal forma que ela e suas duas primeiras derivadas são contínuas. A terceira derivada, entretanto, pode ter descontinuidades nos pontos x_i. Tal função é uma spline cúbica interpoladora com nós nos pontos x_i.

Se usarmos splines quadráticas, teremos que $S_2(x)$ tem derivadas contínuas até à ordem 1 apenas e, portanto, a curvatura de $S_2(x)$ pode trocar nos nós. Por esta razão, as splines cúbicas são mais usadas. Uma spline cúbica, $S_3(x)$, é uma função polinomial por partes, contínua, onde cada parte, $s_k(x)$, é um polinômio de grau 3 no intervalo $[x_{k+1}, x_k]$, $k = 1, 2, \ldots, n$. $S_3(x)$ tem a primeira e segunda derivadas contínuas, o que faz com que a curva $S_3(x)$ não tenha picos e nem troque abruptamente de curvatura nos nós.

O Spline cúbico (polinómio interpolador de terceiro grau) é o spline mais utilizado uma vez que fornece uma excelente aproximação aos pontos tabelados e o seu cálculo não é complicado.

Seja $f(x)$ uma função contínua no intervalo $[a, b]$ e consideremos os pontos $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$, e consideremos a seguinte divisão de $[a, b]$:

$$[a, b] = \bigcup_{k=0}^{n} S_k = \bigcup_{k=0}^{n} [x_k, x_{k+1}]. \tag{5.75}$$

Em cada subintervalo S_k vamos definir um polinómio cúbico diferente, que será denotado por $S_k(x)$, isto é,

$$S(x) = \begin{cases} S_0(x) & x \in [x_0, x_1] \\ S_1(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_{n-1}(x) & x \in [x_{k-1}, x_k] \end{cases} \tag{5.76}$$

Cada um dos polinómios satisfaz as seguintes condições de interpolação:

i) $S(x_k) = y_k, \quad {\scriptstyle (k=0,\ldots,n)}$

ii) $S_k(x_{k+1}) = S_{k+1}(x_{k+1}), \quad {\scriptstyle (k=0,\ldots,n-2)}$ continuidade de S_k

iii) $S'_k(x_{k+1}) = S'_{k+1}(x_{k+1}), \quad {\scriptstyle (k=0,\ldots,n-2)}$ continuidade de S'_k

iv) $S''_k(x_{k+1}) = S''_{k+1}(x_{k+1}), \quad {\scriptstyle (k=0,\ldots,n-2)}$ continuidade de S''_k

Para cada k $_{(k=0,\ldots,n-1)}$,

$$S_k(x) = a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3 \tag{5.77}$$

em que a_k, b_k, c_k, d_k são constantes a determinar.

As condições de interpolação (i)

$$S(x_k) = y_k, \quad (k = 0, \ldots, n)$$

determinam $a_k = y_k$, $_{(k=0,\ldots,n-1)}$.

Temos agora $3n$ coeficientes a determinar. Precisamos, pois, de $3n$ condições.

Da continuidade de S, obtemos as condições

$$\begin{aligned} &S_k(x_{k+1}) = S_{k+1}(x_{k+1}) = y_{k+1}, \qquad {\scriptstyle k=0,\ldots,n-2} \\ &S_{n-1}(x_n) = y_n \end{aligned}$$

o que leva a

$$y_{k+1} = y_k + b_k h_k + c_k h_k^2 + d_k h_k^3, \qquad {\scriptstyle k=0,\ldots,n-1} \tag{5.78}$$

onde $h_k = x_{k+1} - x_k$, $k = 0, \ldots, n-1$.

A expressão (5.78) origina n equações.

Da continuidade de S' obtemos as condições

$$S'_k(x) = b_k + 2c_k(x - x_k) + 3d_k(x - x_k)^2$$

$$S'_k(x_{k+1}) = S'_{k+1}(x_{k+1}), \qquad {\scriptstyle k=0,\ldots,n-2}$$

obtém-se

$$b_{k+1} = b_k + 2c_k h_k + 3d_k h_k^2, \qquad {\scriptstyle k=0,\ldots,n-2} \tag{5.79}$$

que origina mais $n-1$ equações.

Da continuidade de S'', $(S''_k(x) = 2c_k + 6d_k(x - x_k))$

$$S''_k(x_{k+1}) = S''_{k+1}(x_{k+1}), \qquad {\scriptstyle k=0,\ldots,n-2}$$

obtém-se

$$c_{k+1} = c_k + 3d_k h_k, \qquad {\scriptstyle k=0,\ldots,n-2} \tag{5.80}$$

o que origina mais $n-1$ equações.

Então, de (5.78), (5.79) e (5.80) temos $3n-2$ equações para determinar $3n$ incógnitas.

São necessárias mais duas condições adicionai. É habitual considerar as condições

$$S''_0(x_0) = S''_{n-1}(x_n) = 0 \tag{5.81}$$

originando o chamado spline cúbico natural.

No entanto, também pode recorrer-se a outras condições, como por exemplo,

$$S_0'(x_0) = y_0' = d_0$$

e

$$S_{n-1}'(x_n) = y_n' = d_n \tag{5.82}$$

obtendo o designado *spline cúbico completo.*

5.10.1 Spline cúbico natural

O spline cúbico natural que interpola f nos pontos $x_0, \ldots, x_n$ é da forma

$$S_k(x) = a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3, \quad x \in [x_k, x_{k+1}], \quad {\scriptstyle k=0,\ldots,n-1}$$

sendo as constantes dadas pelas equações (5.78), (5.79), (5.80) e (5.81), ou seja,

$$\begin{aligned} &h_k = x_{k+1} - x_k, \quad a_k = f(x_k), \\ &b_k = f_{k,k+1} - \tfrac{h_k}{3}(2c_k + c_{k+1}), \quad d_k = \tfrac{c_{k+1}-c_k}{3h_k} \end{aligned} \qquad {\scriptstyle (k=0,1,\ldots,n-1)}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ h_0 & 2(h_0+h_1) & h_1 & 0 & \ldots & 0 & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 3f_{1,2}-3f_{0,1} \\ 3f_{2,3}-3f_{1,2} \\ \vdots \\ 3f_{n-1,n}-3f_{n-2,n-1} \\ 0 \end{bmatrix}.$$

A coluna dos termos independentes é da forma

$$\begin{bmatrix} 0 \\ 3f_{1,2}-3f_{0,1} \\ 3f_{2,3}-3f_{1,2} \\ \vdots \\ 3f_{n-1,n}-3f_{n-2,n-1} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{3(f(x_2)-f(x_1))}{h_1} - \dfrac{3(f(x_1)-f(x_0))}{h_0} \\ \dfrac{3(f(x_3)-f(x_2))}{h_2} - \dfrac{3(f(x_2)-f(x_1))}{h_1} \\ \vdots \\ \vdots \\ \dfrac{3(f(x_n)-f(x_{n-1}))}{h_{n-1}} - \dfrac{3(f(x_{n-1})-f(x_{n-2}))}{h_{n-2}} \\ 0 \end{bmatrix}$$

5.10.2 Processo do cálculo do Spline Natural

a) Cálculo dos a_j:

$$a_j = f(x_j), \quad j = 0, \ldots, n-1.$$

b) cálculo dos c_j:

$$h_{j-1}c_{j-1} + 2(h_{j-1} + h_j)c_j + h_j c_{j-1} = \frac{3}{h_j}(a_{j+1} - a_j) - \frac{3}{h_{j-1}}(a_j - a_{j-1}), j = 0, 1, \ldots, n-1$$

que é um sistema da forma:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \\ h_0 & 2(h_0+h_1) & h_1 & 0\cdots & 0 & 0 & \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0\cdots & 0 & 1 & \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \cdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 3f_{1,2} - 3f_{0,1} \\ 3f_{2,3} - 3f_{1,2} \\ \vdots \\ 3f_{n-1,n} - 3f_{n-2,n-1} \\ 0 \end{bmatrix}$$

onde

$$\begin{bmatrix} 0 \\ 3f_{1,2} - 3f_{0,1} \\ 3f_{2,3} - 3f_{1,2} \\ \vdots \\ 3f_{n-1,n} - 3f_{n-2,n-1} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{3(f(x_2) - f(x_1))}{h_1} - \dfrac{3(f(x_1) - f(x_0))}{h_0} \\ \dfrac{3(f(x_3) - f(x_2))}{h_2} - \dfrac{3(f(x_2) - f(x_1))}{h_1} \\ \vdots \\ \vdots \\ \dfrac{3(f(x_n) - f(x_{n-1}))}{h_{n-1}} - \dfrac{3(f(x_{n-1}) - f(x_{n-2}))}{h_{n-2}} \\ \\ 0 \end{bmatrix}$$

c) cálculo dos b_j

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}), \quad j = 0, 1, \ldots n-1$$

d) cálculo dos d_j

$$d_j = \frac{c_{j+1} - c_j}{3h_j}, \quad j = 0, 1, \ldots n-1$$

O spline fica então

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3, \; x \in [x_j, x_{j+1}], \; j = 0, 1, 2, \ldots, n-1.$$

5.10.3 Processo do cálculo do Spline Completo

a) Cálculo dos a_j:

$$a_j = f(x_j), \quad j = 0, \ldots, n-1.$$

b) cálculo dos c_j:

$$h_{j-1}c_{j-1} + 2(h_{j-1} + h_j)c_j + h_j c_{j-1} = \frac{3}{h_j}(a_{j+1} - a_j) - \frac{3}{h_{j-1}}(a_j - a_{j-1}), j = 1, \ldots, n-1$$

que é um sistema da forma:

$$\begin{bmatrix} 2h_0 & h_0 & 0 & 0 & \ldots & 0 & 0 \\ h_0 & 2(h_0+h_1) & h_1 & 0 & \ldots & 0 & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \ldots & h_{n-1} & 2h_{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \cdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} 3f_{0,1} - 3f'(x_0) \\ 3f_{1,2} - 3f_{0,1} \\ 3f_{2,3} - 3f_{1,2} \\ \vdots \\ 3f_{n-1,n} - 3f_{n-2,n-1} \\ 3f'(x_n) - 3f_{n-1,n} \end{bmatrix}$$

onde

$$\begin{bmatrix} 3f_{0,1} - 3f'(x_0) \\ 3f_{1,2} - 3f_{0,1} \\ 3f_{2,3} - 3f_{1,2} \\ \vdots \\ 3f_{n-1,n} - 3f_{n-2,n-1} \\ 3f'(x_n) - 3f_{n-1,n} \end{bmatrix} = \begin{bmatrix} \frac{3(f(x_1) - f(x_0))}{h_0} - 3f'(x_0) \\ \frac{3(f(x_2) - f(x_1))}{h_1} - \frac{3(f(x_1) - f(x_0))}{h_0} \\ \frac{3(f(x_3) - f(x_2))}{h_2} - \frac{3(f(x_2) - f(x_1))}{h_1} \\ \vdots \\ \vdots \\ \frac{3(f(x_n) - f(x_{n-1}))}{h_{n-1}} - \frac{3(f(x_{n-1}) - f(x_{n-2}))}{h_{n-2}} \\ 3f'(x_n) - \frac{3(f(x_n - f(x_{n-1})))}{h_{n-1}} \end{bmatrix}$$

c) cálculo dos b_j

$$b_j = \frac{1}{h_j}(a_{j+1} - a_j) - \frac{h_j}{3}(2c_j + c_{j+1}), \quad j = 0, 1, \ldots n-1$$

d) cálculo dos d_j

$$d_j = \frac{c_{j+1} - c_j}{3h_j}, \quad j = 0, 1, \ldots n-1$$

O spline fica então

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3,\ x \in [x_j, x_{j+1}],\ j = 0, 1, 2, \ldots, n-1.$$

5.10.4 Spline Cúbico Completo

O cálculo do Spline Cúbico Completo apenas difere do anterior na obtenção dos c_k que agora resulta do sistema

$$\begin{bmatrix} 2h_0 & h_0 & 0 & 0 & \ldots & 0 & 0 \\ h_0 & 2(h_0+h_1) & h_1 & 0 & \ldots & 0 & 0 \\ 0 & h_1 & 2(h_1+h_2) & h_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2(h_{n-2}+h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & 0 & \ldots & h_{n-1} & 2h_{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} f_{0,1} - 3f'(x_0) \\ 3f_{1,2} - 3f_{0,1} \\ 3f_{2,3} - 3f_{1,2} \\ \vdots \\ 3f_{n-1,n} - 3f_{n-2,n-1} \\ 3f'(x_n) - f_{n,n-1} \end{bmatrix}$$

5.10.5 Erro de Interpolação

O erro de interpolação cometido utilizando o spline cúbico é dado pelo seguinte resultado:

Teorema 5.10.1 *Seja $f \in \mathcal{C}^4([x_0, x_n])$ e seja $S(x)$ o sline cúbico (natural ou completo) que interpolador de f em $[x_0, x_n]$. Então,*

$$|f(x) - S(x)| \leq \frac{5}{384} M h^4 \tag{5.83}$$

onde

$$M = \max_{x \in [x_9, x_n]} \left| f^{(4)}(x) \right| \ \text{ e } h = \max_{0 \leq i \leq n} h_i \tag{5.84}$$

sendo $h_i = x_{i+1} - x_i$ para $0 \leq i \leq n$.

Exemplo 5.10.1 *Determine o Spline natural que interpole a função $y = f(x)$ onde*

x_i	1	2	4	5
$y_i = f(x_i)$	2	1	4	3

Resolução 5.10.1 *Como vimos anteriormente, os coeficientes dos polinómios podem ser calculados da seguinte forma:*

a) Distâncias horizontais (h_j):

$$\begin{aligned} h_1 &= x_1 - x_0 = 2 - 1 = 1 \\ h_2 &= x_2 - x_1 = 4 - 2 = 2 \\ h_3 &= x_3 - x_2 = 5 - 4 = 1 \end{aligned}$$

b) Distâncias verticais:

$$\begin{aligned}\Delta_1 &= f(x_1) - f(x_0) = 1 - 2 = -1\\ \Delta_2 &= f(x_2) - f(x_1) = 4 - 1 = 3\\ \Delta_3 &= f(x_3) - f(x_2) = 3 - 4 = -1\end{aligned}$$

c) Cálculo dos c_j:

$$\begin{aligned}c_{22} &= 2(h_0 + h_1) = 2(1 + 2) = 6\\ c_{33} &= 2(h_2 + h_1) = 2(2 + 1) = 6\end{aligned}$$

Os termos independentes do sistema são

$$b_{21} = 3\left(\tfrac{3}{2} - \tfrac{-1}{1}\right) = \tfrac{15}{2}$$

$$b_{31} = 3\left(\tfrac{-1}{1} - \tfrac{3}{2}\right) = -\tfrac{15}{2}$$

d) Resolução do sistema tridiagonal:

$$\begin{bmatrix} 1 & 0 & 0 & 0\\ h_0 & 2(h_0+h_1) & h_1 & 0\\ 0 & h_1 & 2(h_1+h_2) & h_2\\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_0\\ c_1\\ c_2\\ c_3 \end{bmatrix} = \begin{bmatrix} 0\\ c_{21}\\ c_{31}\\ 0 \end{bmatrix}$$

Cuja solução é $c_1 = 0$, $c_2 = \frac{15}{8}$, $c_3 = -\frac{15}{8}$ e $c_4 = 0$.

e) Cálculo dos b_j

$$b_0 = \tfrac{1}{h_0}(a_1 - a_0) - \tfrac{h_0}{3}(2c_0 + c_1) = -1 - \tfrac{1}{3}\left(2 \times 0 + \tfrac{15}{8}\right) = -\tfrac{13}{8}$$

$$b_1 = \tfrac{1}{h_1}(a_2 - a_1) - \tfrac{h_1}{3}(2c_1 + c_2) = \tfrac{3}{2} - \tfrac{2}{3}\left(2 \times \tfrac{15}{8} + \tfrac{-15}{8}\right) = \tfrac{1}{4}$$

$$b_2 = \tfrac{1}{h_2}(a_3 - a_2) - \tfrac{h_2}{3}(2c_2 + c_3) = -1 - \tfrac{1}{3}\left(2 \times \left(\tfrac{-15}{8}\right) + 0\right) = \tfrac{1}{4}$$

f) Cálculo dos d_j:

$$d_0 = \tfrac{c_1 - c_0}{3h_0} = \tfrac{\frac{15}{8} - 0}{3 \times 1} = \tfrac{5}{8}$$

$$d_1 = \tfrac{c_2 - c_1}{3h_1} = \tfrac{-\frac{15}{8} - \frac{15}{8}}{3 \times 2} = -\tfrac{5}{8}$$

$$d_2 = \tfrac{c_3 - c_2}{3h_2} = \tfrac{0 - \left(-\frac{15}{8}\right)}{3 \times 1} = \tfrac{5}{8}$$

g) O spline cúbico é então:

$$S(x) = \begin{cases} S_0(x) &= 2 - \frac{13}{8}(x-1) + 0(x-1)^2 + \frac{5}{8}(x-1)^3, \\ &= \frac{1}{8}\left(5x^3 - 15x^2 + 2x + 24\right), & \text{para } 1 \leq x \leq 2; \\ S_1(x) &= 1 + \frac{1}{4}(x-2) + \frac{15}{8}(x-2)^2 - \frac{5}{8}(x-2)^3, \\ &= -\frac{5}{8}x^3 + \frac{45}{8}x^2 - \frac{59}{4}x + 13 & \text{para } 2 \leq x \leq 4; \\ S_2(x) &= 4 + \frac{1}{4}(x-4) - \frac{15}{8}(x-4)^2 + \frac{5}{8}(x-4)^3, \\ &= \frac{5}{8}x^3 - \frac{75}{8}x^2 + \frac{181}{4}x - 67 & \text{para } 4 \leq x \leq 5. \end{cases} \tag{5.85}$$

De seguida, apresentamos a representação gráfica das funções $S_i(x)$, para $i = 0, 1, 2$ e de $S(x)$:

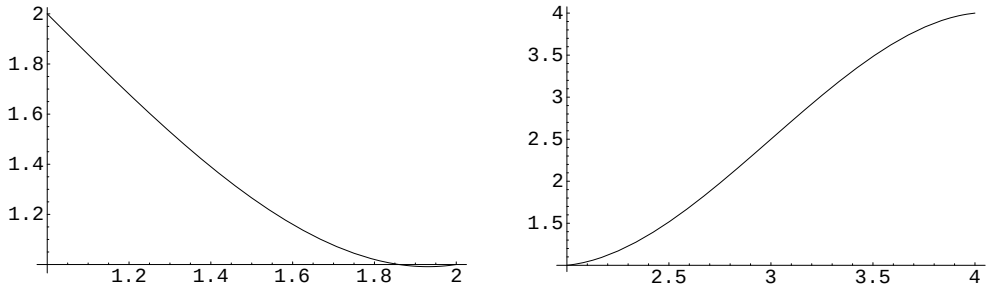

Figura 5.5: Gráficos de $S_0(x)$ e de $S_1(x)$

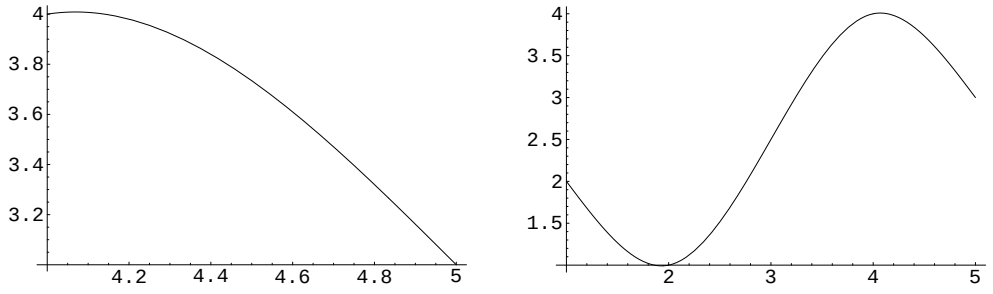

Figura 5.6: Gráficos de $S_2(x)$ e de $S(x)$

Verifique que:

1. *$S(x)$ é uma função contínua em $[1, 5]$;*

2. $S'(x)$ *é uma função contínua em* $[1,5]$*;*

3. $S''(x)$ *é uma função contínua em* $[1,5]$*;*

4. $S''(1) = S''(5) = 0$.

■

5.11 Exercícios

1. Determine o polinómio de terceiro grau da forma $y = ax^3 + bx^2 + cx + d$ que "passa" pelos pontos $(1,10)$, $(2,26)$, $(-1,2)$ e $(0,4)$, utilizando o método de eliminação de Gauss para o resolver.

2. A seguinte tabela corresponde à função $f(x) = \frac{1}{x}$

x	3.35	3.4	3.5	3.6
$f(x)$	0.298507	0.294118	0.285714	0.277778

   Encontre valores aproximados para $f(3.44)$ utilizando a interpolação linear, quadrática e cúbica. Calcule o valor do erro para cada um dos casos.

3. Uma função g é conhecida exclusivamente através da tabela

x	-2	-1	1	2	3
$g(x)$	-16	0	2	0	4

   a) Calcule uma estimativa de $g(1.65)$, usando o polinómio interpolador de Lagrange de grau 2.

   b) Determine a melhor estimativa de $g(1.65)$ que os dados permitem.

4. Determine o polinómio de Lagrange, $p(x)$, que passa pelos pontos $(-3,1)$, $(-2,2)$, $(1,-1)$ e $(3,10)$. Calcule $p(0)$.

5. Determine aproximações de $\cos(\frac{\pi}{8})$ usando os polinómios interpoladores de Lagrange de grau 2 e 4 no intervalo $[0,\pi]$. Compare os resultados obtidos e indique um majorante para o erro.

6. A seguinte tabela lista a população, em milhares de pessoas, de 1930 a 1980 num certo país.

Ano	1930	1940	1950	1960	1970	1980
população	123.203	131.669	150.697	179.323	203.212	226.505

   Utilize o polinómio de Lagrange para estimar a população no ano 1965.

7. O tempo t que um automóvel leva a passar de uma velocidade inicial, de 30 Km/h, para uma velocidade v, está descrito na seguinte tabela

v (Km/h)	30	45	50	75
t (s)	0.0	1.8	4.3	9.4

Estime o tempo necessário para atingir 48 Km/h.

8. Determine o polinómio de Lagrange, de grau 2 em x e de grau 3 em y, interpolador da função $f(x, y) = -2y^3 + x^2 + 4y^2 - 3x - 1$, no conjunto $(x, y) \in [0, 1] \times [0, 1]$.

9. Pretende-se calcular o valor de $\sqrt{1.005}$. Suponha que é dada a seguinte tabela relativa à função $f(x) = \sqrt{x}$

x_i	1.00	1.01	1.02	1.03	1.04	1.05
$f(x_i)$	1.0000	1.0050	1.0100	1.0149	1.0189	1.0247

a) Determine uma aproximação para $\sqrt{1.005}$, utilizando o polinómio interpolador de Newton de grau 3. Indique a tabela da diferenças divididas.

b) Calcule um majorante do erro da estimativa obtida na alínea anterior.

c) Calcule o erro.

10. Seja $f(x)$ dada pela seguinte tabela.

x_i	-2	0	2	4	6
$f(x_i)$	1	2	-1	2	3

Determine uma aproximação para o valor de $f(-1.5)$ usando a fórmula interpoladora de Newton.

11. Dada a tabela de valores de uma determinada função real

x	0	1	2	4
y	2	2	3	6

a) Determine o polinómio interpolador da função, de grau 2, usando a tabela das diferenças divididas, e calcule um majorante para o erro cometido.

b) A partir do polinómio da alínea anterior, encontre o polinómio de grau 3 que interpola a função nos quatro pontos tabelados.

c) Indique a melhor estimativa de $y(1.85)$, que os dados permitem.

12. Num teste para determinar a elongação dum material em função da temperatura obtiveram-se os seguintes valores

Temperatura (^{0}C)	70	78	83	90	95
Elongação (%)	3	5	9	11	17

Preveja a elongação a obter se a temperatura for 80 ^{0}C.

13. A diferença de voltagem V que atravessa uma resistência para vários valores de corrente I foi medida e registada na tabela seguinte

I	0.25	0.75	1.25	1.5	2.0
V	-0.45	-0.60	0.70	1.88	6.0

Utilize a interpolação polinomial para estimar a voltagem para $I = 1.1$.

14. Foi feito um teste para relacionar a tensão e a deformação numa barra de alumínio, tendo-se obtido os seguintes valores

Tensão	1	2	3	4	5	6	7	8	9	10	11	12
Deformação	2	4	6	6	6	7	8	7.5	7	7.5	8	7.5

Usando a interpolação determine o valor da deformação correspondente a uma tensão de 7.4. Indique uma estimativa do erro cometido.

15. Considere a tabela

x	-1	0	1
$f(x)$	0	1	0
$f'(x)$	0	0	0

a) Defina a estimativa de $f(0.25)$ recorrendo ao polinómio interpolador de Hermite.

b) Determine outra estimativa do mesmo valor, definindo o polinómio de Hermite unicamente no segmento que contém o 0.25

c) Use a fórmula interpoladora de Newton para calcular uma terceira aproximação de $f(0.25)$.

16. Construa o polinómio de Hermite de grau 3 para função $g(x)$, definida a seguir.

x	1	2	3	4
$g(x)$	0	$\frac{14}{2}$	$\frac{84}{3}$	$\frac{245}{4}$
$g'(x)$	4	$\frac{44}{5}$	$\frac{254}{8}$	$\frac{769}{17}$

Obtenha uma aproximação de $g'(2.5)$.

17. Construa o polinómio de Hermite de grau 3 para função $f(x)$, definida a seguir.

x	0.1	0.2	0.3	0.4
$f(x)$	0.2	1.42	0.84	.245
$f'(x)$	0.4	0.44	0.254	0.77

Obtenha uma aproximação para:

a) $f(0.26)$;

b) $f'(0.26)$;

c) $f(0.35)$;

d) $f'(0.35)$.

18. Pretende-se interpolar a função $sin(\pi x)$ no intervalo $[0,1]$, por um spline cúbico natural, numa malha uniforme.
Construa o spline cúbico natural que interpola a função nos nós $x_0 = 0$, $x_1 = \frac{1}{4}$, $x_2 = \frac{1}{2}$, $x_3 = \frac{3}{4}$ e $x_4 = 1$.

19. Determine o spline cúbico natural que interpola a função $f(x) = x(1 + x^2)$, nos pontos $x_0 = -1$, $x_1 = 0$ e $x_2 = 1$.

20. De uma função real conhecem-se apenas os valores

x	0.0	0.5	0.7	1.0
$f(x)$	0.0	0.6	1.2	2.2

    a) Determine o spline cúbico natural que interpola a função nos pontos dados.

    b) Calcule uma estimativa de $f(0.3)$.

21. As funções de Bessel aparecem muitas vezes em engenharia e no estudo de campos eléctricos. Estas funções são bastante complexas de avaliar directamente, por isso são muitas vezes compilados em tabelas, por exemplo

x	1.8	2.0	2.2	2.4	2.6
$J_0(x)$	0.3400	0.2239	0.1104	0.0025	0.0968

    Estime $J_0(2.1)$

    a) Utilizando interpolação polinomial.

    b) Utilizando splines cúbicos.

    c) Calcule os erros das aproximações anteriores sabendo que o valor exacto é 0.1666.

22. Determine o Spline natural que interpole a função $y = f(x)$ onde

x_i	-1	-2	4	5
$y_i = f(x_i)$	3	2	-1	-2

23. Determine o Spline cúbico completo que interpole a função $y = f(x)$ onde

x_i	-1	-2	0	1	2
$y_i = f(x_i)$	3	0	-1	7	3

Capítulo 6

O método dos mínimos quadrados

Nesta secção, vamos estudar a aproximação de funções numa perspectiva diferente da interpolação. Dada uma função $f(x)$ definida num intervalo $[a,b]$, pretende-se encontrar uma função $p(x)$ de tal forma que o erro entre a função $f(x)$ e $p(x)$ seja mínimo (no caso discreto e contínuo), isto é, dada uma norma $\|\cdot\|$ interessa determinar

$$\min \|f(x) - p(x)\|. \tag{6.1}$$

Neste caso, exigimos apenas que essa função "aproximadora"tome valores (nesses pontos) de forma a minimizar a distância aos valores dados e a função aproximadora . . . falamos em minimizar, no sentido dos *mínimos quadrados*, isto é, pretendemos minimizar, em alguma norma, o erro entre a função conhecida, quer num conjunto finito de pontos ou num intervalo (casos discreto ou contínuo).

Sejam u e v duas funções contínuas em $[a,b]$. Então, o produto interno entre as funções u e v está definido por

$$(u,v) = \int_a^b u(x)\, v(x)\, dx, \tag{6.2}$$

e a norma de $u(x)$ é dada por

$$\|u\| = \|u\|_2 = \left(\int_a^b u^2(x)\, dx \right)^{\frac{1}{2}}. \tag{6.3}$$

Portanto

$$\|u\|^2 = \|u\|_2^2 = \int_a^b u^2(x)\, dx. \tag{6.4}$$

No caso discreto,

$$(u,v) = \sum_{i=0}^{n} u(x_i)\, v(x_i)\, dx, \tag{6.5}$$

e a norma de $u(x)$ é dada por

$$\|u\| = \|u\|_2 = \left(\sum_{i=0}^{n} u^2 (x_i) \right)^{\frac{1}{2}}. \tag{6.6}$$

Portanto

$$\|u\|^2 = \|u\|_2^2 = \sum_{i=0}^{n} u^2 (x_i). \tag{6.7}$$

O anteriormente descrito corresponde a minimizar:

- $\|f(x_i) - p_n(x_i)\|_2 = \left(\sum_{i=0}^{n} [f(x_i) - p_n(x_i)]^2 dx \right)^{\frac{1}{2}}$;
- $\|f(x) - p_n(x)\|_2 = \left(\int_a^b [f(x) - p_n(x)]^2 dx \right)^{\frac{1}{2}}$.

Isto é importante em termos de aplicações, já que podemos ter valores obtidos experimentalmente com uma certa incerteza. Ao tentar modelelar essa experiência, com uma certa classe de funções, seria inadequado exigir que a função aproximadora interpolasse esses pontos.

Um caso simples, em que se aplica esta teoria é o caso da regressão linear, em que tentamos adaptar a um conjunto de pontos e valores dados, a "melhor recta", que (neste caso) será a recta que minimiza a soma quadrática das diferenças entre os valores dados ao valores da recta, nesses pontos, isto é,

$$\min_i \|f(x_i) - (ax_i + b)\|_2 = \min_i \left(\sum_{i=0}^{n} [f(x_i) - (ax_i + b)]^2 \right)^{\frac{1}{2}}, \tag{6.8}$$

esta é uma perspectiva discreta, em que o conjunto de valores dados é finito.

Neste caso, interessa determinar os valores de a e de b que minimizam (6.8). Da análise, sabemos que é necessário calcular as derivadas parciais de (6.8) em ordem a a e a b e igualá-las a zero. Mas, o problema de minimizar a norma é equivalente a minimizar o seu quadrado, portanto, é necessário resolver

$$\begin{cases} \dfrac{\partial Q}{\partial a} = -2 \displaystyle\sum_{i=0}^{n} [f(x_i) - (ax_i + b)] x_i = 0 \\ \dfrac{\partial Q}{\partial b} = -2 \displaystyle\sum_{i=0}^{n} [f(x_i) - (ax_i + b)] = 0 \end{cases} \tag{6.9}$$

Nota 6.0.1 *Calcule as derivadas parciais, resolva o sistema anterior e determine a fórmula para a recta dos mínimos quadrados.*

Podemos também pensar num caso contínuo, em que apesar de conhecermos a função, não apenas em certos pontos, mas em todo um intervalo, estamos interessados em aproximar essa função (... no sentido dos mínimos quadrados) por funções de uma outra classe, mais adequadas ao problema que pretendemos resolver. Neste caso, é necessário minimizar o integral

$$\left\| f(x) - \left(\sum_i a_i \varphi_i \right) \right\|_2 = \left(\int_a^b \left[f(x) - \left(\sum_i a_i \varphi_i \right) \right]^2 dx \right)^{\frac{1}{2}}. \tag{6.10}$$

Por exemplo, podemos estar interessados em determinar qual a "melhor recta"que aproxima a função $\sin(x)$ no intervalo $[0,1]$. Neste caso, é necessário minimizar o integral

$$Q(a,b) = \|\sin(x) - (ax+b)\|_2 = \left(\int_0^1 [\sin(x) - (ax+b)]^2 dx \right)^{\frac{1}{2}}. \tag{6.11}$$

Em resumo, dada uma função $f(x)$ vamos aproximá-la por uma função da forma

$$p(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \ldots + a_n\varphi_n(x) \tag{6.12}$$

que é a combinação linear das funções $\{\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x)\}$. A função $p(x)$, habitualmente, é designada *polinómio generalizado*.

Para as determinar é necessário determinar o valro das constantes a_i, $0 \leq i \leq n$ que minimizam o integral

$$\|f(x) - p_n(x)\|_2^2 = \int_a^b [f(x) - (a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots a_n\varphi_n(x))]^2 dx.$$

Para tal é necessário calcular as derivadas parciais em ordem a cada um dos $a_i's$, para $0 \leq i \leq n$ e iguar a 0, isto é,

$$\frac{\partial Q}{\partial a_i} = \int_a^b (f(x) - (a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots a_n\varphi_n(x)))\,\varphi_i(x)\,dx = 0,\ 0 \leq i \leq n. \tag{6.13}$$

Isto é, obtemos o sistema de equações lineares nas incógnitas a_0, ..., a_n,

$$\int_a^b f(x)\,\varphi_i(x)\,dx = a_0 \int_a^b \varphi_0(x)\,\varphi_i(x)\,dx + \cdots + a_n \int_a^b \varphi_n(x)\,\varphi_i(x)\,dx \tag{6.14}$$

para $0 \leq i \leq n$.

O problema é então calcular o valor das constantes $a_0, \ldots, a_n$. Devemos salientar que, para garantir que $p(x)$ exista e seja único exigimos que $\{\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x)\}$ sejam linearmente independentes, num contexto apropriado.

6.1 Caso discreto

Consideremos, de novo, um conjunto de pontos $\{x_0, \ldots, x_n\}$ a que estão associados, respectivamente, os valores $f(x_0)$, ..., $f(x_n)$.

Temos que considerar agora uma classe de funções, entre as quais vamos tentar encontrar a que "melhor aproxima"aquele conjunto de valores, nos pontos dados.

Vamo-nos concentrar em funções da forma:

$$g(x) = a_0\phi_0(x) + \ldots + a_n\phi_n(x) \tag{6.15}$$

em que $\phi_0, \ldots, \phi_n$ são funções base (linearmente independentes), e são conhecidas.

Habitualmente, considere-se, por simplicidade, a base canónica

$$\mathcal{B} = \{1, x, x^2, \ldots, x^n\},$$

que gera um subespaço vectorial de dimensão $n + 1$.

Assim

$$p_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n. \tag{6.16}$$

Neste caso, apenas teremos que determinar os parâmetros a_0, ...,a_n , de forma a que a soma quadrática das diferenças entre os $f(x_i)$ e os $p_n(x_i)$ seja mínima.

Faz pois sentido introduzir a distância $\|f(x) - p_n(x)\|$ em que

$$\|u\|^2 = \sum_{i=0}^{n} u^2(x_i), \tag{6.17}$$

que está associada ao produto interno (no caso discreto)

$$(u, v) = \sum_{i=0}^{n} u(x_i)\, v(x_i). \tag{6.18}$$

A norma e o produto interno estão bem definidos para funções que assumem quaisquer valores nos pontos $x_0, \ldots, x_n$. Convém-nos trabalhar com estas noções, já que aquilo que iremos ver, de seguida, será exactamente igual no caso contínuo, apenas a norma e o produto interno serão diferentes (substituiremos o somatório por um integral ...).

Pretende-se pois encontrar os parâmetros $a_0, \ldots, a_n$ que minimizem a distância entre f e g, ou, o que é equivalente, minimizem :

$$Q = \|f(x) - p_n(x)\|^2 = (f - p_n(x), f - p_n(x)). \tag{6.19}$$

Para obtermos esse mínimo, começamos por procurar os valores $a_0, \ldots, a_m$ tais que todas as derivadas parciais de Q sejam nulas, isto é:

$$\frac{\partial Q}{\partial S}(a_0, \ldots, a_n) = 0, (\text{ para } j = 0, \ldots, n). \tag{6.20}$$

Calculamos a derivada parcial, usando as propriedades da derivação do produto interno :

$$\begin{aligned}\frac{\partial Q}{\partial a_i} &= \frac{\partial}{\partial a_i}(f-p_n(x), f-p_n(x)) \\ &= \left(\frac{\partial}{\partial a_i}(f-p_n(x)), (f-p_n(x))\right) + \left((f-p_n(x)), \frac{\partial}{\partial a_i}(f-p_n(x))\right) \\ &= 2\left((f-p_n(x)), \frac{\partial}{\partial a_i}(f-p_n(x))\right) = -2\left((f-p_n(x)), \frac{\partial}{\partial a_i} g\right) \end{aligned} \tag{6.21}$$

Por outro lado

$$\frac{\partial Q}{\partial a_i} = \frac{\partial}{\partial a_i}(a_0\varphi_0 + \ldots + a_i\varphi_i + \cdots + a_n\varphi_n) = \varphi_i \tag{6.22}$$

e assim obtemos, para cada $i = 0, \ldots, n$:

$$(f-g, \varphi_i) = 0. \tag{6.23}$$

Podemos ainda substituir a expressão de $p_n(x)$ e obtemos um sistema de equações lineares:

$$\sum_{j=0}^{n} a_i(\phi_i, \phi_j) = (f, \varphi_i), \quad \text{para } i = 0, \ldots, n \tag{6.24}$$

designado por *sistema normal*, que escrevemos matricialmente:

$$\begin{bmatrix} (\varphi_0,\varphi_0) & \ldots & (\varphi_0,\varphi_n) \\ \vdots & \ddots & \vdots \\ (\varphi_n,\varphi_0) & \ldots & (\varphi_n,\varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (f,\varphi_0) \\ \vdots \\ (f,\varphi_n) \end{bmatrix} \tag{6.25}$$

Exemplo 6.1.1 *No caso de considerarmos a aproximação através de funções polinomiais, temos como funções base, $\varphi_0 = 1, \ldots, \varphi_n = x^n$, e assim obtemos:*

$$\begin{bmatrix} \sum_{i=0}^m 1 & \sum_{i=0}^m x_i & \sum_{i=0}^m x_i^2 & \ldots & \sum_{i=0}^m x_i^n \\ \sum_{i=0}^m x_i & \sum_{i=0}^m x_i^2 & \sum_{i=0}^m x_i^3 & \ldots & \sum_{i=0}^m x_i^{n+1} \\ \sum_{i=0}^m x_i^2 & \sum_{i=0}^m x_i^3 & \sum_{i=0}^m x_i^4 & \ldots & \sum_{i=0}^m x_i^{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^m x_i^n & \sum_{i=0}^m x_i^{n+1} & \sum_{i=0}^m x_i^{n+2} & \ldots & \sum_{i=0}^m x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^m f(x_i) \\ \sum_{i=0}^m x_i f(x_i) \\ \sum_{i=0}^m x_i^2 f(x) \\ \vdots \\ \sum_{i=0}^m x_i^m f(x_i) \end{bmatrix}$$

No método dos mínimos quadrados podemos definir o erro (mais conhecido por erro padrão) pela expressão

$$\frac{1}{m+1}\left(\sum_{i=0}^{m}(f(x_i)-p_n(x_i))^2\right)^{\frac{1}{2}}$$

Exemplo 6.1.2 *Os valores da função g são apresentados na tabela seguinte*

x	0.50	0.55	0.60	0.65	0.70	0.75	0.80
$g(x)$	1.2	1.0	0.7	0.4	0.1	-0.2	-0.6

a) Aproxime a função g por uma parábola, recorrendo ao método dos mínimos quadrados, e defina uma estimativa de $g(0.65)$.

b) Calcule o erro padrão da resposta da alínea anterior.

Resolução 6.1.1 *Como queremos um parábola, a função pretendida é*

$$p_2(x) = a_2x^2 + a_1x + a_0$$

Seleccionamos então a base $\{1, x, x^2\}$. *Logo para encontrarmos os coeficientes* a_0, a_1 *e* a_2 *temos de resolver o sistema*

$$\begin{bmatrix} \sum_{i=0}^{6} 1 & \sum_{i=1}^{m} x_i & \sum_{i=0}^{6} x_i^2 \\ \sum_{i=0}^{6} x_i & \sum_{i=0}^{6} x_i^2 & \sum_{i=0}^{6} x_i^3 \\ \sum_{i=0}^{6} x_i^2 & \sum_{i=0}^{6} x_i^3 & \sum_{i=0}^{6} x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{6} g(x_i) \\ \sum_{i=0}^{6} g(x_i)x_i \\ \sum_{i=0}^{6} g(x_i)x_i^2 \end{bmatrix}$$

Antes de resolvermos o sistema temos de calcular os somatórios, para isso é melhor colocarmos tudo numa tabela que facilita os cálculos e a visualização dos resultados.

	x_i	x_i^2	x_i^3	x_i^4	$g(x_i)$	$x_ig(x_i)$	$x_i^2g(x_i)$
	0.50	0.2500	0.125000	0.06250000	1.2	0.60	0.3000
	0.55	0.3025	0.166375	0.09150625	1.0	0.55	0.3025
	0.60	0.3600	0.216000	0.12960000	0.7	0.42	0.2520
	0.65	0.4225	0.274625	0.17850625	0.4	0.26	0.1690
	0.70	0.4900	0.343000	0.24010000	0.1	0.07	0.0490
	0.75	0.5625	0.421875	0.31640625	-0.2	-0.15	-0.1125
	0.80	0.6400	0.512000	0.40960000	-0.6	-0.48	-0.3840
$\sum$	4.55	3.0275	2.058875	1.42821875	2.6	1.27	0.576

Logo o sistema fica

$$\begin{bmatrix} 7 & 4.55 & 3.0275 \\ 4.55 & 3.0275 & 2.058875 \\ 3.0275 & 2.058875 & 1.42821875 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.6 \\ 1.27 \\ 0.576 \end{bmatrix} \Leftrightarrow \begin{cases} a_0 = 2.307142857 \\ a_1 = 0.1904761905 \\ a_2 = -4.761904762 \end{cases}$$

Então a parábola que melhor se aproxima dos pontos dados é

$$p_2(x) = 2.307142857 + 0.1904761905x - 4.761904762x^2$$

Logo podemos concluir que $g(0.65) \approx p_2(0.65) = 0.419047619$

Para o cálculo do erro temos de construir uma tabela da forma

x_i	$g(x_i)$	$p_2(x_i)$	$g(x_i) - p_2(x_i)$	$(g(x_i) - p_2(x_i))^2$
0.50	1.2	1.21190476175	-0.01190476175	$0.14172335232427 \times 10^{-3}$
0.55	1.0	0.97142857127	0.02857142873	$0.81632653967347 \times 10^{-3}$
0.60	0.7	0.70714285698	-0.00714285698	$0.05102040583673 \times 10^{-3}$
0.65	0.4	0.41904761888	-0.01904761888	$0.36281178499772 \times 10^{-3}$
0.70	0.1	0.10714285697	-0.00714285697	$0.05102040569388 \times 10^{-3}$
0.75	-0.2	-0.22857142875	0.02857142875	$0.81632654081633 \times 10^{-3}$
0.80	-0.6	-0.58809523828	-0.01190476172	$0.14172335160996 \times 10^{-3}$
				$\sum = 0.00238095238095$

então o erro é

$$E_p = \frac{1}{7}\sqrt{0.00238095238095} = 0.00697071480678$$

■

6.2 O Caso Contínuo

Vamos considerar agora que conhecemos a função f não apenas em alguns pontos, mas sim num determinado intervalo $[a, b]$. Mais uma vez estamos interessados em aproximar f por funções da forma

$$p_n(x) = a_0\varphi_0(x) + \ldots + a_n\varphi_n(x) \tag{6.26}$$

ou seja, com dependência linear dos parâmetros.

A única diferença existente, face ao caso discreto, está na norma e no produto interno :

$$\|u\|^2 = \int_a^b u\,(x)^2\,dx \tag{6.27}$$

que está associada o produto interno

$$(u, v) = \int_a^b u\,(x)\,v\,(x)\,dx. \tag{6.28}$$

Tudo se deduz de forma semelhante, e obtemos também um sistema normal, cuja única diferença está no significado dos produtos internos.

No caso em que consideramos como funções base, os polinómios,

$$\varphi_0(x) = 1, \ldots, \varphi_n(x) = x^n, \tag{6.29}$$

obtemos agora o sistema normal

$$\begin{bmatrix} \int_a^b 1\,dx & \int_a^b x\,dx & \int_a^b x^2\,dx & \ldots & \int_a^b x^n\,dx \\ \int_a^b x\,dx & \int_a^b x^2\,dx & \int_a^b x^3\,dx & \ldots & \int_a^b x^{n+1}\,dx \\ \int_a^b x^2\,dx & \int_a^b x^3\,dx & \int_a^b x^4\,dx & \ldots & \int_a^b x^{n+2}\,dx \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \int_a^b x^n\,dx & \int_a^b x^{n+1}\,dx & \int_a^b x^{n+2}\,dx & \ldots & \int_a^b x^{2n}\,dx \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \int_a^b f(x)\,dx \\ \int_a^b xf(x)\,dx \\ \int_a^b x\,2f(x)\,dx \\ \vdots \\ \int_a^b x^n f(x)\,dx \end{bmatrix} \tag{6.30}$$

Podemos obter este sistema a partir da relação (6.14) considerando a base (6.29):

$$\mathcal{B} = \left\{1, x, x^2, \ldots, x^n\right\}.$$

Após alguns cálculos, concluímos que o polinómio pretendido se obtém ao resolver o sistema linear de $n + 1$ equações e $n + 1$ incógnitas dado por (6.30).

A matriz dos coeficientes em (6.30) é designada por *Matriz de Hilbert*, e é extremamente mal condicionada. Com efeito, já para $n = 3$ obtemos $\text{Cond}\,(A) = 28375$, e para $n = 4$ já atinge 943656, continuando a crescer fortemente. Temos, assim, problemas de condicionamento e consequentemente de instabilidade numérica, para este tipo de matrizes.

No caso contínuo, o erro padrão é

$$\frac{1}{b-a}\left(\int_a^b (f(x)-p_n(x))^2\,dx\right)^{\frac{1}{2}}. \tag{6.31}$$

Exemplo 6.2.1 *Encontre a aproximação dos mínimos quadrados de grau 2 da função* $f(x) = \cos(\pi x)$, $x \in [0,1]$.

Determine o erro padrão.

Resolução 6.2.1 *Para aproximarmos a função dada por um polinómio de grau 2, vamos utilizar a base* $\{1, x, x^2\}$.
Nesta base o polinómio é

$$p_2(x) = a_0 + a_1 x + a_2 x^2$$

As constantes a_0, a_1, a_2 *são solução do sistema*

$$\begin{bmatrix} \int_0^1 1\,dx & \int_0^1 x\,dx & \int_0^1 x^2\,dx \\ \int_0^1 x\,dx & \int_0^1 x^2\,dx & \int_0^1 x^3\,dx \\ \int_0^1 x^2\,dx & \int_0^1 x^3\,dx & \int_0^1 x^4\,dx \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \int_0^1 \cos(\pi x)\,dx \\ \int_0^1 x\cos(\pi x)\,dx \\ \int_0^1 x^2\cos(\pi x)\,dx \end{bmatrix}$$

ou seja

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{2}{\pi^2} \\ -\frac{2}{\pi^2} \end{bmatrix}$$

Resolvendo o sistema obtemos

$$\begin{cases} a_0 = \dfrac{12}{\pi^2} \\ a_1 = -\dfrac{24}{\pi^2} \\ a_2 = 0 \end{cases}$$

Logo a aproximação pedida é

$$p_2(x) = \frac{12}{\pi^2} - \frac{24}{\pi^2}x. \tag{6.32}$$

O erro padrão é dado por

$$\begin{aligned} E_p &= \frac{1}{b-a}\sqrt{\int_a^b \left(f(x)-p_2(x)\right)^2 dx} \\ &= \frac{1}{1-0}\sqrt{\int_a^b \left(\cos(\pi x)-\left(\frac{12}{\pi^2}-\frac{24}{\pi^2}x\right)\right)^2 dx} \\ &= \sqrt{\frac{1}{2}-\frac{48}{\pi^4}} \approx 0.0850462 \end{aligned}$$

■

Observação 6.2.1 (Dependência não linear) *: Quando não há dependência linear dos coeficientes, há duas possibilidades a considerar:*

1. *Método exacto: Efectuamos ainda a derivação $\frac{\partial Q}{\partial a_j}$ mas isso irá levar à resolução de um sistema não linear.*
2. *Método aproximado: Quando possível, por transformação de variável, reduzimos a forma da função a aproximar ao caso linear, e aí usamos o método linear descrito acima, regressando às variáveis anteriores por transformação inversa.*

Um exemplo habitual, é considerar $g(x) = ae^{bx}$. Assim, como queremos que $f(x) \approx g(x)$, usamos $\log(f(x)) \approx \log(g(x)) = \log(a) + bx$. Definindo $F(x) = \log(f(x))$, $A = \log(a)$, $B = b$, procedemos à aproximação habitual de F usando os mínimos quadrados (neste caso regressão linear) e tendo encontrado os valores A e B, usamos transformação inversa para obter $a = e^A$, $b = B$.

6.3 Polinómios Ortogonais

Para polinómios de graus elevados necessitamos de calcular

$$\frac{(n+1)(n+4)}{2}$$

integrais, ou somatórios, e resolver um sistema linear de ordem $(n+1)$. Como apenas sabemos que a matriz dos coeficientes é simétrica, a sua resolução é difícil porque envolve muitos cálculos. Mais, o problema é mal condicionado.

Com o objectivo de ultrapassarmos este problema, é necessário considerar bases para além da base

$$\mathcal{B} = \left\{1, x, x^2, x^3, \ldots, x^n\right\}.$$

Vamos considerar bases em que os respectivos polinómios são ortogonais em $[a, b]$, isto é, satisfazem a condição

$$\int_a^b \varphi_i(x)\varphi_j(x)\, dx = 0, \quad i \neq j$$

Recordemos a expressão (6.14):

$$\int_a^b f(x)\,\varphi_i(x)\,dx = a_0\int_a^b \varphi_0(x)\,\varphi_i(x)\,dx + \cdots + a_n\int_a^b \varphi_n(x)\,\varphi_i(x)\,dx,$$

para $0 \leq i \leq n$.

Uma vez que as funções $\{\varphi_i(x)\}_{i=0}^n$ são ortogonais, isto é,

$$\int_a^b \varphi_i(x)\varphi_j(x)\;dx = 0, \quad i \neq j$$

o que implica que o sistema normal (6.25)

$$\begin{bmatrix} (\varphi_0,\varphi_0) & \cdots & (\varphi_0,\varphi_n) \\ \vdots & \ddots & 0 \\ (\varphi_n,\varphi_0) & \cdots & (\varphi_n,\varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (f,\phi_0) \\ \vdots \\ (f,\phi_n) \end{bmatrix}$$

fica

$$\begin{bmatrix} (\varphi_0,\varphi_0) & 0 & \cdots & 0 \\ 0 & (\varphi_1,\varphi_1) & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (\varphi_n,\varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (f,\varphi_0) \\ \vdots \\ (f,\varphi_n) \end{bmatrix}. \tag{6.33}$$

Deste modo, a matriz do sistema reduz-se a uma matriz diagonal e a solução desse sistema é então

$$a_k = \frac{\displaystyle\int_a^b f(x)\,\varphi_k(x)dx}{\displaystyle\int_a^b (\varphi_k(x))^2 dx}, \quad {\scriptstyle k=0,\ldots,n.} \tag{6.34}$$

Como exemplo de polinómios ortogonais, vamos considerar os polinómios de *Legendre*, que podem ser obtidos pela seguinte fórmula de recorrência

$$\begin{cases} L_0(x) = 1, \;\; L_1(x) = x \\ \\ L_{n+1}(x) = \dfrac{2n+1}{n+1}xL_n(x) - \dfrac{n}{n+1}L_{n-1}(x), \;\; n \geqslant 1 \end{cases}, \;\; x \in [-1,1]. \tag{6.35}$$

Uma desvantagem destes polinómios é que eles só são ortogonais no intervalo $[-1,1]$. Este problema resolve-se com uma mudança de variável nos polinómios da base, isto é, é necessário utilizar a transformação:

$$\begin{array}{rcl} [a,b] & \mapsto & [-1,1] \\ z & \to & x = \dfrac{2z}{b-a} - \dfrac{b+a}{b-a} \end{array} \tag{6.36}$$

Exemplo 6.3.1 *Pretende-se aproximar a função $f(x) = e^{-x}$ no intervalo $[-1,1]$ por um polinómio. Defina a parábola que melhor aproxima f no intervalo dado, recorrendo ao método dos mínimos quadrados e usando polinómios de Legendre. Determine o erro comentido.*

Resolução 6.3.1 *Como o intervalo é $[-1,1]$, para definirmos uma parábola precisamos de utilizar os polinómios*

$$P_0(x) = 1, \;\; P_1(x) = x, \;\; P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$$

isto é, a base utilizada é $\left\{1, x, \frac{3}{2}x^2 - \frac{1}{2}\right\}$. Então o polinómio procurado é

$$p_2(x) = a_0 \times 1 + a_1 \times x + a_2 \times \left(\frac{3}{2}x^2 - \frac{1}{2}\right),$$

onde as constantes são determinadas pelo sistema

$$\begin{bmatrix} \int_{-1}^{1} P_0(x)P_0(x)dx & \int_{-1}^{1} P_0(x)P_1(x)dx & \int_{-1}^{1} P_0(x)P_2(x)dx \\ \int_{-1}^{1} P_1(x)P_0(x)dx & \int_{-1}^{1} P_1(x)P_1(x)dx & \int_{-1}^{1} P_1(x)P_2(x)dx \\ \int_{-1}^{1} P_2(x)P_0(x)dx & \int_{-1}^{1} P_2(x)P_1(x)dx & \int_{-1}^{1} P_2(x)P_2(x)dx \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} P_0(x)f(x)dx \\ \int_{-1}^{1} P_1(x)f(x)dx \\ \int_{-1}^{1} P_2(x)f(x)dx \end{bmatrix}$$

Como os polinómios são ortogonais, só é necessário calcular os elementos da diagonal principal, um a vez que

$$\int_{-1}^{1} P_i(x)\, P_j(x)\, dx = 0, \;\; \text{para } i \neq j.$$

Logo o sistema fica

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{5} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} e - e^{-1} \\ -2e^{-1} \\ e - 7e^{-1} \end{bmatrix}$$

Resolvendo o sistema obtemos

$$\begin{cases} a_0 = \dfrac{e - e^{-1}}{2} \\ a_1 = -3e^{-1} \\ a_2 = \dfrac{5e - 35e^{-1}}{2} \end{cases}$$

Substituindo no polinómio obtemos

$$\begin{aligned} p_2(x) &= \frac{e - e^{-1}}{2} \times 1 - 3e^{-1} \times x + \frac{5e - 35e^{-1}}{2} \times \left(\frac{3}{2}x^2 - \frac{1}{2}\right) \\ &= \frac{e - e^{-1}}{2} - 3e^{-1} \times x + \frac{5e - 35e^{-1}}{2} \times \left(\frac{3}{2}x^2 - \frac{1}{2}\right) \end{aligned}$$

O erro padrão é dado por

$$\begin{aligned} E_p =& \frac{1}{1+1}\left(\int_{-1}^{1}(f(x)-p_2(x))^2dx\right)^{\frac{1}{2}} \\ =& \frac{1}{2}\sqrt{\int_{-1}^{1}(e^{-x}-0.536721526x^2+1.103638324x-0.9962940184)^2\ dx} \\ =& 0.0189774441 \end{aligned}$$

■

Exemplo 6.3.2 *Pretende-se aproximar a função $g(x) = \sinh(x)$ por um polinómio de terceiro grau no intervalo $[0,1]$. Quais deverão ser os elementos da base do espaço a que pertence a aproximação de $g(x)$ que minimiza o trabalho relativo à resolução do sistema de equações que permite definir os parâmetros do modelo matemático?*

Resolução 6.3.2 *Para minimizar o trabalho de resolução do sistema temos de utilizar polinómios ortogonais.*

$$P_0(x)=1,\ \ P_1(x)=x,\ \ P_2(x)=\frac{3}{2}x^2-\frac{1}{2},\ \ P_3(x)=\frac{5}{2}x^3-\frac{3}{2}x$$

Os polinómios de Legendre só são ortogonais em $[-1,1]$, por isso temos de fazer uma mudança de variável dos polinómios para o intervalo $[0,1]$.
Essa mudança é

$$\begin{aligned} [0,1] &\mapsto [-1,1] \\ z &\to x=\frac{2z}{1-0}-\frac{1+0}{1-0}=2z-1 \end{aligned}$$

então obtemos os polinómios

$$\begin{aligned} P_0(z) &= 1, \\ P_1(z) &= 2z-1, \\ P_2(z) &= \frac{3}{2}(2z-1)^2-\frac{1}{2}=6z^2-6z+1, \\ P_3(z) &= \left(\frac{5}{2}(2z-1)^3-\frac{3}{2}(2z-1)\right)=20z^3-30z^2+12z-1 \end{aligned}$$

então a base é.

$$\{1, 2x-1, 6x^2-6x+1, 20x^3-30x^2+12x-1\}.$$

Verifique que

$$\int_0^1 (2x-1)\left(6x^2-6x+1\right)dx=0,\ \ \int_0^1 1\times(2x-1)\,dx=0.$$

■

6.4 Método dos mínimos quadrados não linear

O caso mais geral do método dos mínimos quadrados é quando se pretende obter uma função que não é polinomial.

Neste caso, procuramos primeiro uma função polinomial equivalente, ou seja, por exemplo:

- $y = ae^{bx} \longrightarrow \ln(y) = \ln(a) + bx$
- $y = ab^{x} \longrightarrow \ln(y) = \ln(a) + x\ln(b)$
- $y = ax^{b} \longrightarrow \ln(y) = \ln(a) + b\ln(x)$
- $y = ae^{a+bx} \longrightarrow \ln(y) = a + bx$
- $y = \dfrac{1}{a+bx} \longrightarrow \dfrac{1}{y} = a + bx$

Aplicamos depois o método ao polinómio resultante e, concluímos, definindo a função original.

6.5 Exercícios

1. Determine a recta dos mínimos quadrados relativa aos dados $(x_i, f(x_i))$, $0 \leq i \leq n$, onde

x_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$f(x_i)$	0.201	0.399	0.75	1.04	1.25	1.399	1.45	1.58

2. Determine a parábola dos mínimos quadrados relativa aos dados $(x_i, f(x_i))$, $0 \leq i \leq n$, onde

x_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$f(x_i)$	-0.201	0.399	0.75	1.64	1.25	1.399	-1.45	-1.58

3. Determine a parábola dos mínimos quadrados relativa aos dados $(x_i, f(x_i))$, $0 \leq i \leq n$, onde

x_i	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$
$f(x_i)$	-1.3	1.55	1.2	-1.45

   Determine o erro padrão cometido.

4. Pretende-se interpolar a função $f(x) = sin(\pi x)$ no intervalo $[0, 1]$, um polinómio cúbico numa malha uniforme com 5 pontos, isto é, $x_0, x_1, \ldots, x_4$. Determine o polinómio dos mínimos quadrados que melhor aproxima $f(x)$.

   Considere agora o caso contínuo:

a) Determine a parábola dos mínimos quadrados que melhor aproxima $f(x)$ em $[0,1]$. Considere a base $\mathcal{B} = \{1, x, x^2\}$;

b) Determine o polinómio cúbico dos mínimos quadrados que melhor aproxima $f(x)$ em $[0,1]$. Considere a base $\mathcal{B} = \{1, x, x^2, x^3\}$;

c) Determine a parábola dos mínimos quadrados que melhor aproxima $f(x)$ em $[-1,1]$. Utilize polinómios de Legendre;

d) Determine o polinómio cúbico dos mínimos quadrados que melhor aproxima $f(x)$ em $[-1,1]$. Utilize polinómios de Legendre;

e) Determine a parábola dos mínimos quadrados que melhor aproxima $f(x)$ em $[0,1]$. Utilize polinómios de Legendre;

f) Determine o polinómio cúbico dos mínimos quadrados que melhor aproxima $f(x)$ em $[1,3]$. Utilize polinómios de Legendre.

Em cada um dos casos, determine o erro padrão.

5. Determine os valores de a e b de tal forma que a função $y = a\sin(x) + b\cos(x)$ aproxime no sentido dos mínimos quadrado os seguintes dados:

x_i	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$
$f(x_i)$	-1.3	1.55	1.2	-1.45

Determine o erro padrão cometido.

6. Determine os valores de a e b de tal forma que a função $y = a\sin(x) + b\cos(x)$ aproxime no sentido dos mínimos quadrado da parábola $y = 1 - x^2$ para $x \in [-1,1]$. Determine o erro padrão.

7. Determine a aproximação polinomial cúbica dos mínimos quadrados que a função $f(x) = \dfrac{\cos(\pi x)}{1+x^2}$ no intervalo $\mathcal{I} = [0,1]$. Determine o erro padrão cometido.

8. Determine a aproximação polinomial cúbica dos mínimos quadrados que a função $f(x) = e^{-x^2}$ no intervalo $\mathcal{I} = [0,2]$. Determine o erro padrão cometido.

9. Considere a função real de variável real $f(x) = \dfrac{\sin(x)}{x}$ no intervalo $\mathcal{I} = [1,3]$.

a) Determine a aproximação polinomial cúbica dos mínimos quadrados que aproxima a função $f(x)$. Determine o erro padrão cometido.

b) Utilizando os polinómios de Legendre, determine a aproximação polinomial cúbica dos mínimos quadrados que aproxima $f(x)$ no intervalo $\mathcal{I} = [1,3]$.

10. Considere a base $\mathcal{B} = \{\sin(i\pi x)\}_{i=1}^{5}$ para $x \in [0,1]$. Determine uma função $\varphi(x)$ que pertençe ao espaço gerado por $\mathcal{B}$ que melhor aproxima a função $f(x) = e^{-x}\sin(x)$ no sentido dos mínimos quadrados.

11. Pretende-se aproximar a função $f(x) = e^{-2x}(1+x)$ no intervalo $[-1,1]$ por um polinómio. Defina a parábola que melhor aproxima f no intervalo dado, recorrendo ao método dos mínimos quadrados e usando polinómios de Legendre. Determine o erro comentido.

12. Pretende-se aproximar a função $f(x) = \frac{x}{1+x^2}$ no intervalo $[-1,1]$ por um polinómio. Defina a parábola que melhor aproxima f no intervalo dado, recorrendo ao método dos mínimos quadrados e usando polinómios de Legendre. Determine o erro comentido.

13. Resolva o problema 12 mas considerando o intervalo:

    a) $\mathcal{I} = [0,1]$;
    b) $\mathcal{I} = [1,2]$;
    c) $\mathcal{I} = [-1,2]$;
    d) $\mathcal{I} = [-2,4]$;

Capítulo 7

Quadratura numérica

Frequentemente, quando pretendemos calcular um integral definido, é impossível fazer os cálculos, quer seja por se tratar duma expressão muito complicada, seja por estarmos perante uma função não primitivável simbolicamente como por exemplo as funções

$$f(x) = \frac{e^x}{x}, \quad f(x) = \frac{\sin(x)}{x}.$$

Os métodos de integração numérica permitem calcular o valor aproximado de um integral definido sem conhecer uma expressão analítica para a sua primitiva. O método básico envolvido nesta aproximação é chamado de quadratura numérica e consiste em:

$$\int_a^b f(x)\,dx \approx \sum_{i=0}^{n} \alpha_i f(x_i)$$

Nesta secção, vamos estudar como calcular

$$\int_a^b f(x)\,dx \tag{7.1}$$

baseados na aproximação de

$$\int_a^b p(x)\,dx \tag{7.2}$$

onde $p(x)$ é uma função que aproxima f em (a,b).

Suponhamos que $f(x) \approx f_n(x)$, $f_n(x)$ aproximação para f,

$$\int_a^b f(x)\,dx \approx \int_a^b f_n(x)\,dx,$$

então o erro da aproximação é dado pela fórmula

$$|E(I_n)| \leq (b-a) \max_{a\leq x\leq b} |f(x) - f_n(x)|. \tag{7.3}$$

7.1 Regra do trapézio

1. Regra dos trapézios (simples): Seja $f \in \mathcal{C}^2([a,b])$. Este método baseia-se numa aproximação de $f(x)$ pela recta que une os pontos $(a, f(a))$ e $(b, f(b))$. Usando a função interpoladora de Lagrange para obter a recta que passa pelos dois pontos, obtermos

$$f(x) \approx p_1(x) = f(a)\frac{x-b}{a-b} + f(b)\frac{x-a}{b-a}.$$

Logo,

$$\int_a^b f(x)\,dx \approx \int_a^b p_1(x)\,dx = \frac{b-a}{2}\left[f(a)+f(b)\right],$$

isto é,

$$\int_a^b f(x)\,dx \approx \frac{b-a}{2}\left[f(a)+f(b)\right] \tag{7.4}$$

e o erro é dado por

$$E_n = -\frac{(b-a)^3}{12} f''(\eta), \quad \eta \in (a,b). \tag{7.5}$$

Devemos salientar que o número η não está especificado o que implica que não é possível apresentar um valor correcto para o erro cometido. No entanto, é possível encontrar um intervalo no qual ele se encontra, isto é,

$$\frac{(b-a)^3}{12} \min_{a\le x\le b} \left|f''(x)\right| \le |E_n| \le \frac{(b-a)^3}{12} \max_{a\le x\le b} \left|f''(x)\right|. \tag{7.6}$$

2. Regra dos trapézios (**composta**: Se $b-a$ não for suficientemente pequeno o erro de integração para a regra dos trapézios, E_n pode não satisfazer a precisão desejada. Assim, é conveniente dividir o intervalo de integração em subintervalos mais pequenos, $[x_i, x_{i+1}]$, de amplitude constante e dada por $\Delta x = \frac{b-a}{n}$ e aplicamos a formula (7.4) a cada um dos subintervalos, isto é,

$$\int_{x_i}^{x_{i-1}} f(x)\,dx \approx \frac{x_i - x_{i-1}}{2}\left[f(x_i)+f(x_{i-1})\right].$$

Se n é o número de subintervalos a ser considerados então:

$$h = \frac{b-a}{n} \,(\equiv \Delta x) \tag{7.7}$$

$$x_i = x_0 + ih, \quad i = \overline{1,n}. \tag{7.8}$$

Portanto,

$$\int_a^b f(x)\,dx = \frac{h}{2}\left[f(x_i) + f(x_{i-1}) - \frac{h^3}{12}f''(\eta_i)\right],\ \ x_{i-1} < \eta_i < x_i. \tag{7.9}$$

Assim,

$$\int_a^b f(x)\,dx \approx h\left(\frac{1}{2}f(x_0) + f(x_1) + \cdots + f(x_{n-1}) + \frac{1}{2}f(x_n)\right) \tag{7.10}$$

$$|E_n(f)| = \frac{b-a}{12}h^2 M_2,\ \ M_2 = \max_{a\le x\le b}\left|f''(x)\right| \tag{7.11}$$

Podemos ainda afirmar que

$$\frac{(b-a)^3}{12n^2}\min_{x\in[a,b]}\left|f''(x)\right| \le |E_n| \le \frac{(b-a)^3}{12n^2}\max_{x\in[a,b]}\left|f''(x)\right|. \tag{7.12}$$

Consequentemente, se $f''(x) \neq 0$, $\forall x \in [a,b]$ o erro é sempre maior que 0. O que é que se passa quando $f''(x) = \text{const}\ \forall x \in [a,b]$?

Exemplo 7.1.1 *Utilizando a regra do trapézio com $n = 3$, determine uma aproximação para o valor do integral*

$$\int_{-1}^{1}\left(1-x^2\right)dx.$$

Resolução 7.1.1 *Neste caso temos $a = -1$, $b = 1$ e*

$$f(x) = 1 - x^2,$$

e $h = \dfrac{1-(-1)}{3} = \dfrac{2}{3}$. Consequentemente,

$$x_i = -1 + i\frac{2}{3},\ \ i = 0,1,2,3.$$

Temos então,

$$\begin{aligned}\int_{-1}^{1}\left(1-x^2\right)dx &\approx \frac{1}{2}\left(\frac{2}{3}\right)\left(f(-1) + 2f\left(-\frac{1}{3}\right) + 2f\left(\frac{1}{3}\right) + f(1)\right)\\ &= \frac{1}{3}\left(0 + \frac{16}{9} + \frac{16}{3} + 0\right) = \frac{32}{27} = 1.185185185.\end{aligned}$$

Devemos salientar que o valor exacto é $\dfrac{4}{3} = 1.33333333\,(3)$.
Obviamente,

$$\int_{-1}^{1}\left(1-x^2\right)dx = x - \frac{x^3}{3}\Big|_{x=-1}^{x=1} = \left(1-\frac{1}{3}\right) - \left(-1+\frac{1}{3}\right) = 2 - \frac{2}{3} = \frac{4}{3}.$$

■

Exemplo 7.1.2 *Utilizando a regra do trapézio com $n = 6$, determine uma aproximação para o valor do integral*

$$\int_1^2 \frac{1}{x} dx.$$

Resolução 7.1.2 *Neste caso temos $a = 1$, $b = 2$ e*

$$f(x) = \frac{1}{x},$$

e $h = \dfrac{2-1}{6} = \dfrac{1}{6}$. Consequentemente,

$$x_i = 1 + i\frac{1}{6} = \frac{6+i}{6}, \quad i = 0, 1, 2, 3, 4, 5, 6.$$

Temos então,

$$\begin{aligned} \int_1^2 \frac{1}{x} dx &\approx \frac{1}{12}\left(1 + 2\left(\frac{6}{7}\right) + 2\left(\frac{3}{4}\right) + 2\left(\frac{2}{3}\right) + 2\left(\frac{3}{5}\right) + 2\left(\frac{6}{11}\right) + \frac{1}{2}\right) \\ &= \frac{9631}{13860} = 0.694877345. \end{aligned}$$

O valor exacto é,

$$\int_1^2 \frac{1}{x} dx = \ln(x)\Big|_{x=1}^{x=2} = \ln(2) - \ln(1) = \ln(2) \approx 0.693147181.$$

Como $f(x) = \frac{1}{x}$ então $f''(x) = \frac{2}{x^3}$ e então, tendo em conta (7.12) obtemos a seguinte estimativa para o erro

$$\frac{1}{432} \min_{x\in[1,2]} \left|\frac{2}{x^3}\right| \leq |E_n| \leq \frac{1}{432} \max_{x\in[1,2]} \left|\frac{2}{x^3}\right| [1].$$

Como $f''(x)$ é estritamente decrescente em $[1, 2]$ temos que

$$0.0005787 < \frac{1}{1738} \leq |E_n| \leq \frac{1}{216} < 0.0046297.$$

■

Exemplo 7.1.3 *Utilizando a regra do trapézio com $n = 9$, determine uma aproximação para o valor do integral*

$$\int_1^e e^x \ln(x)\, dx.$$

[1] Neste caso, como $f''(x) > 0$, $\forall x \in [1, 2]$ podemos omitir "| |".

Resolução 7.1.3 *Neste caso a função* $f(x)$ *é* $f(x) = e^x \ln(x)$ *e temos que* $\Delta x = h = \frac{e-1}{9}$ *com*

$$x_i = 1 + i\frac{e-1}{9}, i = \overline{0,9}.$$

Com um arredondamento com 9 *casas decimais obtemos os valores*

$$\begin{array}{rclrcl} x_0 &=& 1.0, & f(x_0) &=& 0.0, \\ x_1 &=& 1.190920203, & f(x_1) &=& 0.575868251, \\ x_2 &=& 1.381840406, & f(x_2) &=& 1.287915841, \\ x_3 &=& 1.572760609, & f(x_3) &=& 2.182623220, \\ x_4 &=& 1.763680813, & f(x_4) &=& 3.310156060, \\ x_5 &=& 1.954601016, & f(x_5) &=& 4.732251791, \\ x_6 &=& 2.145521219, & f(x_6) &=& 6.524244600, \\ x_7 &=& 2.336441422, & f(x_7) &=& 8.778523818, \\ x_8 &=& 2.527361625, & f(x_8) &=& 11.608640254, \\ x_9 &=& 2.718281828, & f(x_9) &=& 15.154262241. \end{array}$$

Aplicando a regra do Trapézio temos a seguinte aproximação para o valor do integral

$$\begin{aligned} \int_1^e e^x \ln(x)\, dx &\approx 0.095460102\,(0.0 + 2 \times (0.575868251)\, 2 + \times (1.287915841) + 2 \times (2.182623220) \\ &\quad +2 \times (3.310156060) + 2 \times (4.732251791) + 2 \times (6.524244600) \\ &\quad +2 \times (8.778523818) + 2 \times (11.608640254) + 15.154262241) \\ &= 8.892367151. \end{aligned}$$

A função a integrar é $f(x) = e^x \ln(x)$ *e a sua derivada de segunda ordem é*

$$f''(x) = e^x \left(\frac{2}{x} - \frac{1}{x^2} + \ln(x) \right).$$

Devemos salientar que

$$f'''(x) = e^x \left(\frac{3}{x} - \frac{3}{x^2} + \frac{3}{x^3} + \ln(x) \right),$$

que é positiva para $x \geq 1$. *Consequentemente,* $f''(x)$ *é crescente em* $[1, e]$ *e portanto,*

$$0.0141876 < \frac{(e-1)^3}{972} f''(1) \leq |E_n| \leq \frac{(e-1)^3}{972} f''(e) < 0.1265863.$$

■

Exemplo 7.1.4 *Utilizando a regra do trapézio com* $n = 6$ *determine uma aproximação para o integral*

$$\mathcal{I} = \int_0^1 e^{-x^2} dx.$$

Indique o erro cometido.

Resolução 7.1.4 *Como* $n = 6$ *vem que* $h = \dfrac{1-0}{6} = \dfrac{1}{6}$ *e consequentemente*

$$x_i = 0 + \frac{i}{6} = \frac{i}{6}, i = \overline{0,6},$$

isto é,

$$x_0 = 0,\ x_1 = \frac{1}{6},\ x_2 = \frac{2}{6},\ x_3 = \frac{3}{6},\ x_4 = \frac{4}{6},\ x_5 = \frac{5}{6},\ x_1 = 1.$$

Consquentemente,

$$\int_0^1 e^{-x^2} dx \approx \frac{1}{12}\left(f(x_0) + 2f(x_1) + 2f(x_2) + 2f(x_3) + 2f(x_4) + 2f(x_5) + f(x_6)\right)$$

O gráfico da função é

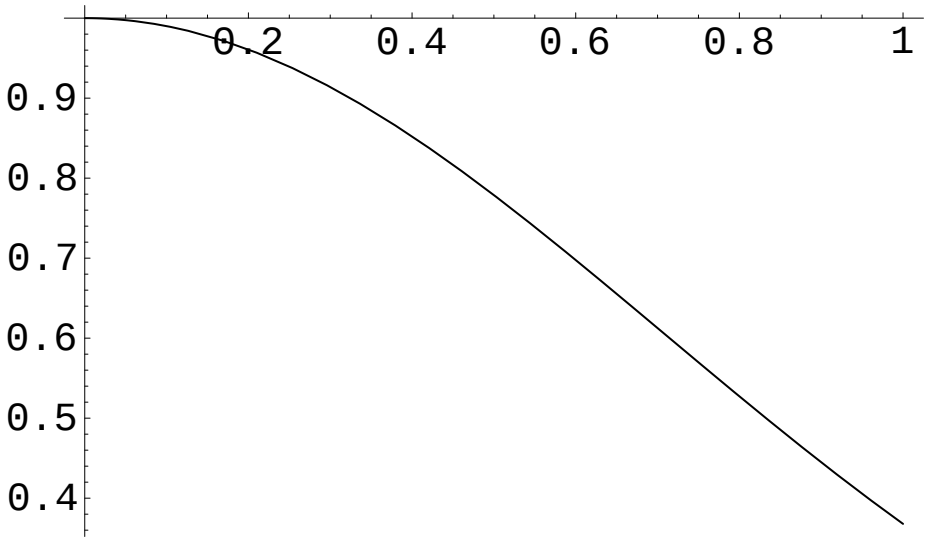

Figura 7.1: Gráfico da função $f(x) = e^{-x^2}$ para $x \in [0,1]$

e o gráfico do módulo da derivada de quarta ordem,

$$f^{(4)}(x) = e^{-x^2}\left(16x^4 - 48x^2 + 12\right)$$

é

Devemos salientar que existe um valor $x^* \in [0,1]$ *tal que* $f^{(4)}(x^*) = 0$*, isto é, o mínimo da função em* $[0,1]$ *é zero.*

■

7.2 O método de Simpson

O outro método que vamos apresentar é o designado Regra de Simpson. Neste método aproximamos a função a integrar $f(x)$ por parábolas e integramos em cada uma delas. Consideremos o intervalo $[a,b]$ e uma sua partição $\{x_0, x_1, \ldots, x_n\}$, isto é,

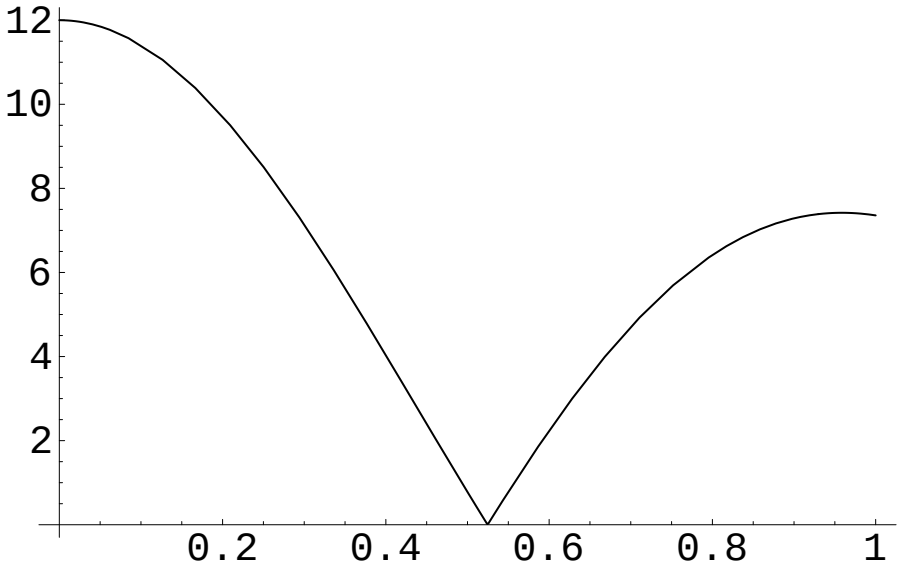

Figura 7.2: Gráfico da função $f(x) = \left| e^{-x^2} \left(16x^4 - 48x^2 + 12\right) \right|$ para $x \in [0,1]$

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n, \tag{7.13}$$

onde tal como anteriormente

$$x_i = a + i\frac{b-a}{n}, \quad i = 0, 1, 2, \ldots, n.$$

No caso da regra de Simpson, exigimos que n seja um *número par*. A (única !) função quadrática que interpola $f(x)$ no intervalo $[x_{i-2}, x_i]$, isto é, que interpola $f(x)$ nos pontos $(x_{i-2}, f(x_{i-2}))$, $(x_{i-1}, f(x_{i-1}))$ e $(x_i, f(x_i))$ é

$$\begin{aligned} y &= \frac{(x-x_{i-2})(x-x_{i-1})}{(x_i-x_{i-2})(x_i-x_{i-1})} f(x_i) + \frac{(x-x_{i-1})(x-x_i)}{(x_{i-2}-x_{i-1})(x_{i-2}-x_i)} f(x_{i-2}) \\ &\quad + \frac{(x-x_i)(x-x_{i-2})}{(x_{i-1}-x_i)(x_{i-1}-x_{i-2})} f(x_{i-1}). \end{aligned} \tag{7.14}$$

Aplicando sucessivamente a fórmula (7.14) obtemos a seguinte aproximação

$$\begin{aligned} \int_a^b f(x)\,dx &\approx \frac{h}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-1}) + f(x_n)\right) \\ &\approx \frac{h}{3}\left(f(x_0) + 4\sum_{i=1}^{\frac{n}{2}} f(x_{2i-1}) + 2\sum_{i=1}^{\frac{n}{2}-1} f(x_{2i}) + f(x_n)\right). \end{aligned} \tag{7.15}$$

Uma aproximação para o erro cometido é dada pela fórmula

$$E_n = \left| \frac{(b-a)^5}{180n^4} f^{(4)}(\eta) \right|,$$

para algum $\eta \in [a,b]$. Devemos salientar que, mais uma vez, o número η não está especificado o que implica que não é possível apresentar um valor correcto para o erro cometido. No entanto, tal como para a regra do trapézio, é possível encontrar um intervalo no qual ele se encontra, isto é,

$$\frac{(b-a)^5}{180n^4} \min_{x\in[a,b]} \left|f^{(4)}(x)\right| \leq |E_n| \leq \frac{(b-a)^5}{180n^4} \max_{x\in[a,b]} \left|f^{(4)}(x)\right| \tag{7.16}$$

Exemplo 7.2.1 *Utilize a regra do trapézio com $n = 6$ para aproximar o valor de*

$$\int_1^2 \frac{1}{x} dx.$$

Determine uma aproximação para o erro cometido.

Resolução 7.2.1 *Tal como anteriormente temos que $h = \frac{1}{6}$ e que*

$$x_i = \frac{6+i}{6}, \ \ i = \overline{0,6}.$$

Aplicando (7.15) obtemos

$$\begin{aligned} \int_1^2 \frac{1}{x} dx &\approx \frac{1}{18}\left(1 + 4\left(\frac{6}{7}\right) + 2\left(\frac{3}{4}\right) + 4\left(\frac{2}{3}\right) + 2\left(\frac{3}{5}\right) + 4\left(\frac{6}{11}\right) + \frac{1}{2}\right) \\ &= \frac{14411}{20790} = 0.693147181. \end{aligned}$$

Como $f(x) = \frac{1}{x}$ vem que $f^{(x)}(x) = \frac{24}{x^5}$ e portanto,

$$\frac{1}{233280} \min_{x\in[1,2]} \left|\frac{24}{x^5}\right| \leq |E_n| \leq \frac{1}{233280} \max_{x\in[1,2]} \left|\frac{24}{x^5}\right|.$$

Como $f^{(4)}(x)$ é estritamente decrescente para qualquer $x > 0$, em particular em $[1,2]$, obtemos a estimativa

$$0.0000003215 < \frac{1}{311040} \leq |E_n| \leq \frac{1}{9720} 0.000102881$$

■

Exemplo 7.2.2 *Utilizando a regra do trapézio com $n = 8$, determine uma aproximação para o valor do integral*

$$\int_1^e e^x \ln(x)\, dx.$$

Resolução 7.2.2 *Tal como anteriormente, temos $f(x) = e^x \ln(x)$ mas, $h = \dfrac{e-1}{8}$ e portanto,*

$$x_i = 1 + i\frac{e-1}{8} = \frac{8 + i(e-1)}{8}, \ \ i = \overline{0,8}$$

$$\begin{array}{rcll} x_0 & = & 1, & f(x_0) = 0, \\ x_1 & = & 1.214785228, & f(x_1) = 0.655608177, \\ x_2 & = & 1.429570457, & f(x_2) = 1.492717202, \\ x_3 & = & 1.644355685, & f(x_3) = 2.575108450, \\ x_4 & = & 1.859140914, & f(x_4) = 3.980031703, \\ x_5 & = & 2.073926142, & f(x_5) = 5.803451205, \\ x_6 & = & 2.288711371, & f(x_6) = 8.165809696, \\ x_7 & = & 2.503496599, & f(x_7) = 11.21889280, \\ x_8 & = & 2.718281828, & f(x_8) = 15.15426224. \end{array}$$

Aplicando a regra de Simpson, temos

$$\begin{aligned} \int_1^e e^x \ln(x)\, dx &\approx 0.071595076\,(0.0 + 4\,(0.655608177) + 2\,(1.492717202) \\ &\quad +4\,(2.575108450) + 2\,(3.980031703) + 4\,(5.803451205) \\ &\quad +2\,(8.165809696) + 4\,(11.21889280) + 15.15426224) \\ &= 8.837955521 \end{aligned}$$

A função a integrar é $f(x) = e^x \ln(x)$ *e a sua derivada de quarta ordem é*

$$f^{(4)}(x) = e^x \left(\frac{4}{x} - \frac{6}{x^2} + \frac{8}{x^3} + \frac{6}{x^4} + \ln(x) \right).$$

Prova-se que $f^{(5)}(x) \geq 0$ *para* $x \geq 1$. *Consequentemente,* $f^{(4)}(x)$ *é crescente em* $[1, e]$ *e portanto podemos apresentar a estimativa*

$$0 = \frac{(e-1)^5}{737280} f^{(4)}(1) \leq |E_n| \leq \frac{(e-1)^5}{737280} f^{(4)}(e) < 0.0006.$$

■

Exemplo 7.2.3 *Determine uma aproximação para o integral*

$$\mathcal{I} = \int_0^3 e^{1-x^2} dx,$$

utilizando as regras do Trapézio e de Simpson com $n = 2, 4, 6, 8$. *Em cada um dos casos, determine um intervalo que contenha o erro.*

Resolução 7.2.3 *Vamos resolver o problema com* $n = 6$. *Os outros casos ficam ao cuidado do leitor.*

O gráfico da função $f(x) = e^{1-x^2}$ *no intervalo* $\mathcal{I} = [0, 3]$ *é*

■

Exemplo 7.2.4 *Determine uma aproximação para o integral*

$$\mathcal{I} = \int_1^2 x e^{1-x^2} dx,$$

utilizando as regras do Trapézio e de Simpson com $n = 2, 4, 6, 8$. *Em cada um dos casos, determine um intervalo que contenha o erro.*

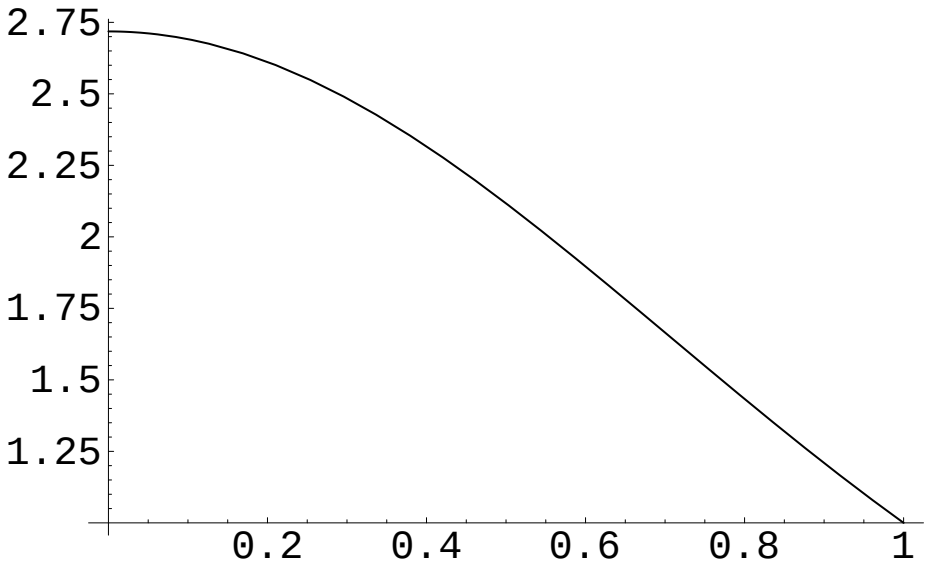

Figura 7.3: Gráfico da função $f(x) = e^{1-x^2}$ para $x \in [0, 1]$

7.3 Método de Romberg

O Método de Romberg utiliza a Regra do Trapézio "repetida"para obter aproximações preliminares e em seguida aplica um processo de extrapolação de Richardson para melhor a aproximação.

O método de Romberg é um método iterativo que se baseia no uso da regra dos trapézios composta juntamente com a estimativa do erro da fórmula de Euler-Maclaurin.

Consideremos uma sucessão de aplicações da regra dos trapézios composta em que o número de intervalos é sucessivamente aumentado para o dobro em cada aplicação. Concretamente, sejam

$$h = h_0 = b - a,\ h_k = \frac{h_{k-1}}{2} = \frac{b-a}{2^k},\ \ k = 1, 2, \ldots$$

e $R_{k,1}$ o valor obtido pela aplicação da regra do trapézios composta com 2^{k-1} subintervalos de comprimento h_k.

De acordo com

$$E_h(f) = I(f) - I_h(f) = c_2 h^2 + c_4 h^4 + \ldots,$$

O erro tem a seguinte expressão

$$E_k = I - R_{k,1} = c_2 h_k^2 + c_4 h_k^4 + \ldots \tag{7.17}$$

onde as constantes c_2, c_4, ... dependem das derivadas da função mas não do parâmetro h_k. Notemos que em (7.17) apenas figuram potências pares de h_k.

Em face do que dissemos, podemos escrever

$$E_{k+1} = I - R_{k+1,1} = c_2 \left(\frac{h_k}{2}\right)^2 + c_4 \left(\frac{h_k}{2}\right)^4 + \dots \tag{7.18}$$

Multiplicando (7.18) por 4 e subtraindo-lhe (7.17), vem

$$I = \frac{4R_{k+1,1} - R_{k,1}}{4-1} + \mathcal{O}(h_k^4).$$

Esta relação mostra que o valor

$$R_{k+1,2} = \frac{4R_{k+1,1} - R_{k,1}}{4-1}$$

difere do valor exacto do integral a menos de um termo de ordem h_k^4, donde se conclui que $R_{k+1,2}$ constituirá uma melhor aproximação.

Obtemos deste modo uma nova sucessão $\{R_{k+1,2}\}$ que converge para o valor exacto $I(f)$ com maior rapidez. Se repetirmos este processo com os novos valores $R_{k+1,2}$ obteremos uma nova sucessão $\{R_{k+1,3}\}$ cujo erro será $\mathcal{O}(h_k^6)$. Não é difícil concluir que é válida a seguinte expressão geral

$$R_{k+1,m+1} = \frac{4^m R_{k+1,m} - R_{k,m}}{4^m - 1}, \quad m = 1, 2, \dots, k$$

ou, ainda,

$$R_{k+1,m+1} = R_{k+1,m} + \frac{R_{k+1,m} - R_{k,m}}{4^m - 1}, \quad m = 1, 2, \dots, k$$

Por cada incremento do valor de m ganhamos uma potência de h_k^2 na expressão do erro, ou seja,

$$I = R_{k+1,m+1} + \mathcal{O}(h_k^{2m+2}).$$

A precisão do algoritmo de Romberg é limitada apenas pela regularidade da função e pelos erros de arredondamento.

Resumindo, os cálculos necessários à aplicação do método são apresentados na seguinte tabela:

$$\begin{array}{cccccccc}
\mathcal{O}(h^2) & & \mathcal{O}(h^4) & & \dots & & \mathcal{O}(h^{2n}) \\
R_{1,1} & & & & & & \\
 & \searrow & & & & & \\
R_{2,1} & \rightarrow & R_{2,2} & & & & \\
 & \searrow & & \searrow & & & \\
\vdots & \rightarrow & \vdots & \ddots & & & \\
 & \searrow & & \searrow & & \searrow & \\
R_{n,1} & \rightarrow & R_{n,2} & \rightarrow & \cdots & \rightarrow & R_{n,n}
\end{array}$$

As entradas da tabela anterior podem obter-se a partir da expressão

$$R_{k+1,m+1} = \frac{4^m R_{k+1,m} - R_{k,m}}{4^m - 1}, \quad m = 1, 2, \dots, k \tag{7.19}$$

quer por colunas quer por linhas.

A análise que fizemos mostra que as colunas e as diagonais, deste quadro, convergem para o valor exacto do integral.

Uma estimativa do erro pode ser obtida através da expressão

$$E_A \approx |R_{n,n} - R_{n-1,n-1}| \tag{7.20}$$

Exemplo 7.3.1 *Utilize o método de Romberg três vezes para resolver o seguinte integral:*

$$\int_1^{1.6} f(x)\ dx = \int_1^{1.6} \frac{2x}{x^2-4}\ dx$$

Resolução 7.3.1 $m = 0$; $2^0 = 1$, *logo vamos aplicar a regra dos trapézio com* 1 *subintervalo,* $h = 1.6 - 1 = 0.6$.

$$R_{1,1} = \frac{0.6}{2}(f(1) + f(1.6)) = 0.3(-0.666666667 - 2.222222222) = -0.866666667$$

$m = 1$

$2^1 = 2$, *regra dos trapézios com 2 subintervalos,* $h = \frac{1.6-1}{2} = 0.3$.

$$\begin{aligned} R_{2,1} = \frac{0.3}{2}(f(1) + 2f(1.3) + f(1.6)) = \\ = 0.15(-0.666666667 + 2 \times (-1.125541126) - 2.222222222) = \\ = -0.770995671 \end{aligned}$$

$$\begin{aligned} R_{2,2} = \frac{4R_{2,1} - R_{1,1}}{4-1} = \frac{4 \times (-0.770995671) - (-0.866666667)}{3} = \\ = -0.73910534 \end{aligned}$$

$m = 2$

$2^2 = 4$, *regra dos trapézios com 4 subintervalos,* $h = \frac{1.6-1}{4} = 0.15$.

$$\begin{aligned} R_{3,1} =& \frac{0.15}{2}(f(1) + 2f(1.15) + 2f(1.3) + 2f(1.45) + f(1.6)) = \\ =& 0.075(-0.666666667 + 2 \times (-0.8590102708) + 2 \times (-1.125541126) \\ & + 2 \times (-1.528326746) - 2.222222222) = \\ =& -0.743598388 \\ R_{3,2} =& \frac{4R_{3,1} - R_{2,1}}{4-1} = \frac{4 \times (-0.743598388) - (-0.770995671)}{3} = \\ =& -0.73446596 \\ R_{3,3} =& \frac{4^2 R_{3,2} - R_{2,2}}{4^2-1} = \frac{16 \times (-0.73446596) - (-0.73910534)}{15} = \\ =& -0.73415667 \end{aligned}$$

$m = 3$

$2^3 = 8$, *regra dos trapézios com 8 subintervalos,* $h = \frac{1.6-1}{8} = 0.075$.

$$\begin{aligned}
R_{4,1} =& \frac{0.075}{2}(f(1) + 2f(1.075) + 2f(1.15) + 2f(1.225) + 2f(1.3) + 2f(1.375) + \\
& + 2f(1.45) + 2f(1.525) + f(1.6)) = \\
=& -0.73640433 \\
R_{4,2} =& \frac{4R_{4,1} - R_{3,1}}{4-1} = \frac{4 \times (-0.73640433) - (-0.743598388)}{3} = \\
=& -0.73400631 \\
R_{4,3} =& \frac{4^2 R_{4,2} - R_{3,2}}{4^2 - 1} = \frac{16 \times (-0.73400631) - (-0.73446596)}{15} = \\
=& -0.73397567 \\
R_{4,4} =& \frac{4^3 R_{4,3} - R_{3,3}}{4^3 - 1} = \frac{64 \times (-0.73397567) - (-0.73415667)}{63} = \\
=& -0.73397279
\end{aligned}$$

Pondo tudo numa tabela fica

$$\begin{array}{cccccccc}
\mathcal{O}(h^2) & & \mathcal{O}(h^4) & & \mathcal{O}(h^6) & & \mathcal{O}(h^8) \\
-0.866666667 & & & & & & \\
 & \searrow & & & & & \\
-0.770995671 & \rightarrow & -0.73910534 & & & & \\
 & \searrow & & \searrow & & & \\
-0.73640433 & \rightarrow & -0.73446596 & \rightarrow & -0.73415667 & & \\
 & \searrow & & \searrow & & \searrow & \\
-0.73640433 & \rightarrow & -0.73400631 & \rightarrow & -0.73397567 & \rightarrow & -0.73397279
\end{array}$$

Portanto, o valor aproximado para o integral é

$$\int_1^{1.6} \frac{2x}{x^2 - 4}\, dx \approx -0.73397279. \tag{7.21}$$

Finalmente uma estimativa do erro pode ser

$$E_A \approx |R_{4,4} - R_{3,3}| = |-0.73397279 - (-0.73415667)| = 0.00018388$$

■

7.4 Quadratura de Gauss

Todas as fórmulas anteriores basearam-se na determinação dos pesos a_i de modo a se ter

$$\int_a^b p_{n-1}(x)\, dx = \sum_{i=1}^{n} a_i p_{n-1}(x_i)$$

para um polinómio $p_{n-1}(x)$ de grau não superior a $n-1$, com $x_1, \ldots, x_n$ n pontos igualmente espaçados e previamente fixados.

Na quadratura de Gauss os pontos de integração x_i e os pesos a_i são determinados em simultâneo de forma a que a igualdade

$$\int_a^b p_m(x)\ dx = \sum_{i=1}^{n} a_i p_m(x_i) \tag{7.22}$$

seja verificada para o maior m possível.

Iremos apenas considerar o intervalo de integração $[-1, 1]$, uma vez que, por meio de uma mudança de variável, podemos sempre escrever

$$\int_a^b g(t)\ dt = \left(\frac{b-a}{2}\right) \int_{-1}^{1} g\left(\frac{a+b+(b-a)x}{2}\right) dx = \int_{-1}^{1} f(x)\ dx$$

para a, $b \in \mathbb{R}$, $a < b$.

Para que a fórmula (7.22) seja de grau m, temos de ter:

$$\int_{-1}^{1} x^k\ dx = \sum_{i=1}^{n} a_i x_i^k$$

para $k = 0, \ldots, m$.

Obtendo-se deste modo o seguinte sistema não linear:

$$\sum_{i=1}^{n} a_i x_i^k = \int_{-1}^{1} x^k\ dx = \frac{1-(-1)^{k+1}}{k+1}, \quad k = 0, 1, \ldots, m \tag{7.23}$$

com $2n$ incógnitas $(x_1, \ldots, x_n; a_1, \ldots, a_n)$
Pelo que o sistema deverá ter $2n$ equações e portanto $m = 2n-1$.

Para $n = 2$ o sistema (7.23) escreve-se:

$$\begin{cases} a_1 + a_2 = 2 \\ a_1 x_1 + a_2 x_2 = 0 \\ a_1 x_1^2 + a_2 x_2^2 = \frac{2}{3} \\ a_1 x_1^3 + a_2 x_2^3 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = -\frac{\sqrt{3}}{3} \\ x_2 = -x_1 \\ a_1 = 1 \\ a_2 = 1 \end{cases}$$

Portanto a fórmula de integração (7.22), neste caso, reduz-se a:

$$I_2(f) = f\left(-\frac{\sqrt{3}}{3}\right) + f\left(\frac{\sqrt{3}}{3}\right)$$

que é de grau $2n-1 = 2 \times 2 - 1 = 3$.

Obs.: Recordemos que a regra de Simpson, que também é de grau 3, utiliza 3 pontos de integração.

A obtenção destes valores $x_1, \ldots, x_n$ e $a_1, \ldots, a_n$ pelo método descrito anteriormente revela-se muito difícil para n muito grande.

Em geral, os pontos de integração $x_1, \ldots, x_n$ são os zeros de um certo polinómio pertencente a uma família de polinómios, chamados polinómios de Legendre.

Podemos então resumir o método de Gauss com n pontos ($n \leqslant 5$) à expressão

$$\int_{-1}^{1} g(x)\ dx = \sum_{i=1}^{n} c_i g(w_i) + \frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^2} \frac{g^{(2n)}(\xi)}{(2n)!}$$

onde

$n \longrightarrow$	pesos c_i	nós w_i
2	$c_1 = c_2 = 1$	$w_1 = -w_2 = -\frac{1}{\sqrt{3}}$
3	$c_1 = c_3 = \frac{5}{9},\ c_2 = \frac{8}{9}$	$w_1 = -w_3 = -\sqrt{\frac{3}{5}},\ w_2 = 0$
$4\longrightarrow$	$c_1 = c_4 = 0.347854845$	$w_1 = -w_4 = -0.861136312$
	$\longrightarrow c_2 = c_3 = 0.652145155$	$w_2 = -w_3 = -0.339981044$
$5\longrightarrow$	$c_1 = c_5 = 0.236926885$	$w_1 = -w_5 = -0.906179846$
	$\longrightarrow c_2 = c_4 = 0.478628670, c_3 = 0.568888889$	$w_2 = -w_4 = -0.538469310, w_3 = 0$

No caso do integral não ter limites de integração -1 e 1 temos de fazer a mudança de variável:

$$\int_a^b f(x)\ dx = \int_{-1}^{1} \frac{b-a}{2} f\left(\frac{b+a}{2} + \frac{b-a}{2} w\right) dw = \int_{-1}^{1} g(w)\ dw$$

7.5 Exercícios

1. A distância percorrida por um míssil entre $t = 8$ e $t = 30$ segundos é dada por

$$x = \int_8^{30} \left(2000 \ln\left[\frac{140000}{140000 - 2100t}\right] - 9.8t\right) dt.$$

   a) Determine uma aproximação para o valor do integral considerando $n = 2, 4, 6$.

   b) Determine o valor real para o erro.

   c) Determine o erro relativo.

2. Utilize a regra do trapézio para calcular uma aproximação para o valor do integral

$$\mathcal{I} = \int_1^2 \frac{e^x}{x} dx,$$

   com $n = 2, 4, 5, 6, 7, 8, 10$ e 11. Em cada caso determine uma estimativa para o erro.

3. Utilize a regra do trapézio para calcular uma aproximação para o valor do integral

$$\mathcal{I} = \int_0^1 e^{-x^2} dx,$$

com $n = 2, 4, 5, 6, 7, 8, 10$ e 11. Em cada caso determine uma estimativa para o erro.

4. Utilize a regra do trapézio para calcular uma aproximação para o valor do integral

$$\mathcal{I} = \int_1^2 \frac{e^x}{x} dx,$$

com $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$. Em cada caso determine uma estimativa para o erro.

5. Utilize a regra do trapézio para calcular uma aproximação para o valor do integral

$$\mathcal{I} = \int_0^1 e^{-x^2} dx,$$

com $n = 2, 4, 6, 8, 10, 12$. Em cada caso determine uma estimativa para o erro.

6. Calcule

$$\int_0^1 \frac{x}{x+1} dx,$$

utilizando a fórmula de Simpson com um erro inferior a 10^{-4}, considerando $n = 10$. Determine um limite inferior e um limite superior para o erro.

7. Considere $n = 6$. Determine uma aproximação para o valor dos seguintes integrais utilizando as regras de Simpson e do Trapézio dos seguintes integrais

a) $\int_0^1 \frac{dx}{x+1}$

b) $\int_0^1 \frac{dx}{1+x^2}$

c) $\int_0^1 \frac{dx}{1+x^3}$

d) $\int_1^2 x \ln(x)\, dx$

e) $\int_1^2 \frac{\log(x)}{x} dx$

f) $\int_1^2 \frac{\sin(x)}{x} dx$

g) $\int_1^\pi \frac{\log(x)}{x} dx$

h) $\int_1^2 \frac{\cos(x)}{x} dx$

i) $\int_1^{\frac{\pi}{2}} \frac{\cos(x)}{x} dx$

j) $\int_1^{\frac{\pi}{2}} \frac{\cos(x)}{x+1} dx$

k) $\int_0^1 e^{-x^1} dx$

8. Numa experiência laboratorial obteve-se, num determinado fenómeno físico a seguinte tabela de resultados:

x	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
$f(x_i)$	1.50	0.75	0.50	0.75	1.50	2.75	4.50	6.75	10.0

Utilizando a regra do Trapézio, determine uma aproximação para

$$\int_0^4 f(x)\, dx$$

Capítulo 8

Equações Diferenciais

8.1 Breve introdução histórica

Os trabalhos preliminares na área dos métodos numéricos para a solução de equações diferenciais são devidos, entre outros, a Isaac Newton (1643-1729) e Gottfried Wilhelm Leibniz (1643-1716), no século XVII, bem como a Lehonard Euler (1707-1783), no século XVIII. Mas foi sobretudo devido aos trabalhos deste último que foram impulsionados os estudos que conduziram aos métodos que hoje conhecemos. Euler deduziu um processo iterativo que permitia determinar, de forma aproximada, a solução de um problema de condição inicial num determinado ponto. A demonstração rigorosa de que, de facto, o processo por ele apresentado conduzia á solução da equação só foi apresentada mais tarde, em 1824, por Augustine Cauchy (1789-1857) e melhorada por Rudolf Lipschitz (1832-1908). No entanto, nem assim, o processo apresentado por Euler se tornou popular. A título de exemplo, Karl Weierstrass (1815-1897), famoso matemático alemão do século XIX trabalhou nestes assuntos sem conhecer os trabalhos de Cauchy e Lipschitz. Os finais do século XIX e princípios do século XX foram muito profíquos ao orescimento de métodos numéricos para a resolução de equações diferenciais. Com os desenvolvimentos efectuados na teoria do calor por Fourier e na mecânica celeste por Adams, Bessel, Cauchy, Gauss, Laguerre, Laplace, Legendre, Leverrier, Poincaré e outros vieram tornar imprescindível a existência de esquemas para determinar a solução numérica de equações diferenciais. A balística foi outra ciência que começou a exigir resultados nesta área. Poderemos dividir os métodos numéricos para determinar a soluç ao de equações diferenciais em duas grandes classes: por um lado, os chamados métodos de passo único aos quais pertence o método de Euler-Cauchy-Lipschitz; por outro os chamados métodos de passo múltiplo. Os sucessores do método de Euler-Cauchy-Lipschitz foram apresentados por K. Heun em 1900 e, sobretudo, por Carl Runge (1856-1927) em 1895 e 1908 e por Martin Wilhelm Kutta (1867-1944) em 1901, tendo sido considerados como generalizações das regras de integração. Estes métodos tornaram-se bastante populares devido às suas propriedades e à sua fácil utilização.1 De notar que o primeiro sistema de equações diferenciais a ser resolvido pelo ENIAC foi integrado pelo método de Heun.

Dos primeiros e mais conhecidos métodos de passo múltiplo para a resolução de equações diferenciais são os chamados métodos de Adams. John Couch Adams (1819-1892), famoso astrónomo britânico que descobriu, em co-autoria, o planeta Neptuno,

baseou-se nos métodos teóricos propostos por Cauchy para apresentar um método novo que usou na integração da equaç ao de Bashforth. Aliás, foi num trabalho de Francis Bashforth (1819-1912) de 1883 que o método proposto por Adams foi apresentado, sendo por isso também conhecido por método de Adams-Bashforth. Foi possivelmente a primeira Grande Guerra que veio dar um forte impulso ao florescimento dos métodos numéricos. A grande quantidade de cálculos e a complexidade dos problemas que a balística exige não poderiam ser efectuadas facilmente sem a ajuda destes processos alternativos. O primeiro contributo para o melhoramento dos métodos existentes foi dado pelo matemático americano Forest Ray Moulton (1872-1952) em 1925, propondo uma classe de métodos conhecida por Adams-Moulton. Em 1928 apareceu um trabalho da autoria de Richard Courant (1888-1972), Kurt Friedrichs (1901-1982) e Hans Lewy que revolucionou toda a teoria dos métodos numéricos para a resolução de equações diferenciais. No entanto foi só depois da segunda Grande Guerra e sobretudo depois do aparecimento do primeiro computador e dos trabalhos de Herman Goldstine (1903-2004) e John von Neumann (1903- 1957), em 1947, que estes métodos começaram a ser usados de forma sistemática. Conceitos como "erros de arredondamento", "número de condição"e "instabilidade numérica"começaram a surgir e a tornar-se de capital importância para o estudo da teoria subjacente. Os estudos sobre métodos numéricos para a resolução de equações diferenciais começaram a merecer a atenção de um número crescente de matemáticos e outros cientistas e hoje é uma das áreas mais importantes e profíquas da Matemática em geral e da Análise Numérica em particular.

As equações que envolvem a derivada de uma função a uma variável ocorrem em muitos ramos da matemática aplicada. Genericamente, qualquer situação que trate a "taxa de variação"de uma variável com respeito a outra conduz a uma equação diferencial. Neste capítulo iremos introduzir e descrever alguns métodos numéricos para integrar equações diferenciais ordinárias.

Antes de abordarmos o tema das equações diferenciais convém, antes de mais, fazer uma breve referência à diferenciação de funções.

8.2 Diferenciação

Para se calcular, pelo menos aproximadamente, o valor da derivada de uma função num ponto $x_0 \in D$, utiliza-se uma versão "simplificada"do conceito de derivada. Esta simplificação, obviamente implica a existência de erros.

O método mais utilizado para aproximar a derivada de uma função é o designada diferenciação de Euler,

$$\frac{dy}{dt} = \frac{y\left(t+\Delta t\right) - y\left(t\right)}{\Delta t} + \mathcal{O}\left(\Delta t\right), \tag{8.1}$$

que tem um erro da ordem Δt.

Um método mais preciso, consiste no cálculo de

$$\frac{dy}{dt} = \frac{y\left(t+\Delta t\right) - y\left(t-\Delta t\right)}{2\Delta t} + \mathcal{O}\left(\left(\Delta t\right)^2\right). \tag{8.2}$$

Para as derivadas de primeira ordem, existem três tipos básicos:

1. Diferenciação "progressiva"(Forward differentiation)

$$\frac{dy}{dx} = \frac{y(x+\Delta x) - y(x)}{\Delta x} + \mathcal{O}(\Delta t), \tag{8.3}$$

2. Diferenciação "regressiva"(Backward differentiation)

$$\frac{dy}{dx} = \frac{y(x) - y(x-\Delta x)}{\Delta x} + \mathcal{O}(\Delta t), \tag{8.4}$$

3. Diferenciação "centrada"(Centered differentiation)

$$\frac{dy}{dx} = \frac{y(x+\Delta x) - y(x-\Delta x)}{\Delta x} + \mathcal{O}\left((\Delta t)^2\right), \tag{8.5}$$

Para as derivadas de segunda ordem, o método mais comum consiste em calcular

$$\frac{d^2y}{dx^2} = \frac{y(x+\Delta x) - 2y(x) + y(x-\Delta x)}{(\Delta x)^2} + \mathcal{O}\left((\Delta t)^2\right), \tag{8.6}$$

8.3 Conceitos e definições

Uma equação diferencial é uma equação em que as incógnitas são funções e as suas derivadas.

Definição 8.3.1 *Chama-se equação diferencial de ordem n a uma equação que estabelece uma relação entre a variável independente x, a função incógnita (ou a variável dependente) $y = y(x)$ e as suas derivadas*

$$y', y'', y''', ..., y^{(n)},$$

isto é, uma equação do tipo

$$F(x, y(x), y', y'', y''', ..., y^{(n(}) = 0 \tag{8.7}$$

onde $y^{(n)} \equiv y^{(n)}(x) = \dfrac{d^n}{dx^n} y(x)$.

Se a função incógnita for uma função de uma única variável x, equação diferencial diz-se ordinária. Quando a função incógnita é função de duas ou mais variáveis, por exemplo, se tivermos $y = y(x,t)$, então uma equação do tipo

$$F\left(x, t, y(x,t), \frac{\partial y}{\partial x}, \frac{\partial y}{\partial t}, ..., \frac{\partial^n y}{\partial^k x \partial^l t}\right) = 0 \tag{8.8}$$

é designada por equação com derivadas parciais.
Os indices k e l nesta equação são números inteiros, tais que $k + l = n$.

Isto é, as equações diferenciais são classificadas quanto ao tipo, a ordem e a linearidade.

1. Quanto ao tipo, uma equação pode ser uma equação ordinária ou uma equação às derivadas parciais. Por exemplo, as equações da forma

$$F\left(x, y, y', y'', \ldots, y^{(n)}\right) = 0 \tag{8.9}$$

é uma equação ordinária e se a equação se pode escrever na forma

$$F\left(x, t, y(x,t), \frac{\partial y}{\partial x}, \frac{\partial y}{\partial t}, \ldots, \frac{\partial^n y}{\partial^k x \partial^l t}\right) = 0 \tag{8.10}$$

ela é uma equação às derivadas parciais.

2. Quanto à ordem de uma equação diferencial ela pode ser de 1^a, 2^a, $\ldots$, $n-$ésima ordem dependendo da derivada de maior ordem presente na equação. Por exemplo uma equação da forma

$$y^{(3)} + 2y'' - y' + 4y = 0 \tag{8.11}$$

é uma equação de terceira ordem enquanto que a equação

$$\begin{aligned} u_t &= \frac{\partial^2 u}{\partial x^2}, \\ &= \Delta u,\ u = u(t,x) \end{aligned} \tag{8.12}$$

é uma equação de segunda ordem.

3. Quanto à linearidade, uma equação pode ser linear ou não linear. Ela diz-se linear se as incógnitas aparecem de forma "linear"na equação. Por exemplo a equação

$$y^{(3)} + y' + 4y = 0 \tag{8.13}$$

é uma equação linear enquanto que as equações

$$y^{(3)} + 2y'' - y' + 4y^2 = 0 \tag{8.14}$$

e

$$u_t = \Delta u - u^2 = \frac{\partial^2 u}{\partial x^2} - u^2 \tag{8.15}$$

não são lineares.

Uma equação linear ordinária de ordem n com uma incógnita é uma equação que se pode escrever na forma

$$a_n(t)\frac{d^n y}{dt^n} + a_{n-1}(t)\frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_2(t)\frac{d^2 y}{dt^2} + a_1(t)\frac{dy}{dt} + a_0(t)\, y = 0. \tag{8.16}$$

Esta equação é uma equação de coeficientes variáveis.

8.4 Existência e unicidade

8.4.1 Solução geral da equação de 1^a ordem

Definição 8.4.1 *Chama-se solução geral da equação*

$$y'(x) = f(x, y) \tag{8.17}$$

à função

$$y = \phi(x, C) \tag{8.18}$$

dependente de uma constante arbitrária C, e tal que:

1. *Satisfaz a equação (8.17) para qualquer valor permitido da constante C;*
2. *Qualquer que seja a condição inicial*

$$y(x_0) = y_0 \tag{8.19}$$

é possível escolher um valor C_0 para a constante C de tal modo que a solução

$$y = \phi(x, C_0)$$

satisfaz a condição (8.19) considerada.

O ponto (x_0, y_0) considerado na condição (8.19) deve pertencer ao domínio D, no qual se verificam as condições de existência e unicidade da solução [1].

Chama-se solução particular da equação (8.17) a qualquer solução que se obtém da solução geral (8.18) quando se atribui um determinado valor á constante arbitrária C.

Exemplo 8.4.1 *Por exemplo, a função*

$$y \equiv y(x) = \sin x + \cos x$$

é uma solução da EDO

$$y'' + y = 0 \tag{8.20}$$

no intervalo $(-\infty, \infty)$.

De facto, derivando duas vezes a função $y(x)$ obtemos

$$y' = \cos x - \sin x, \quad y'' = -\sin x - \cos x$$

Substituindo as expressões de y'' e y na equação diferencial, obtemos a identidade

$$\sin x - \cos x + \sin x + \cos x = 0.$$

Exercício 8.4.1 *Mostre que a função*

$$y = c_1 \sin x + c_2 \cos x, \;\; c_1,\; c_2 \in \mathbb{R}$$

ainda é uma solução da equação (8.20).

[1] Estas condições estão indicadas no teorema de *Picard*.

Exemplo 8.4.2 *Consideremos a equação*

$$ay'' + by' + cy = 0, \;\; a, b, c \in \mathbb{R}, \;\; \textit{tais que } b^2 - 4ac = 0.$$

Vamos mostrar que $y(t) = e^{-\frac{b}{2a}t}$ *é solução desta equação. Temos então que:*

$$y'(t) = -\frac{b}{2a}e^{-\frac{b}{2a}t}, \;\; y''(t) = \frac{b^2}{4a^2}e^{-\frac{b}{2a}t}.$$

Substituindo $y(t)$, $y'(t)$ *e* $y''(t)$, *na equação obtemos*

$$\begin{aligned} ay'' + by' + cy &= a\frac{b^2}{4a^2}e^{-\frac{b}{2a}t} + b\left(-\frac{b}{2a}e^{-\frac{b}{2a}t}\right) + ce^{-\frac{b}{2a}t} \\ &= \left(\frac{b^2}{4a} - \frac{b^2}{2a} + c\right)e^{-\frac{b}{2a}t} \\ &= \frac{-b^2 + 4ac}{4a}e^{-\frac{b}{2a}t} = 0, \end{aligned}$$

pois por hipótese $b^2 - 4a = 0$. *Portanto,* $y(t) = e^{-\frac{b}{2a}t}$ *é a solução da equação.*

$$\frac{dy}{dt} = f(t, y) \Leftrightarrow y' = f(t, y). \tag{8.21}$$

Definição 8.4.2 *Uma solução (particular) de uma equação diferencial da primeira ordem (8.21) num intervalo* I *é uma função* $y(t)$ *definida nesse intervalo tal que a sua derivada* $y'(t)$ *está definida nesse intervalo e nele satisfaz (8.21).*

O problema

$$\begin{cases} y'(t) &= f(t, y) \\ y(t_0) &= y_0 \end{cases} \tag{8.22}$$

é designado por problema de Cauchy ou por problema de valores iniciais. Uma solução do problema (8.22) num intervalo I *é uma função* $y(t)$ *que está definida nesse intervalo, tal que a sua derivada também está definida nesse intervalo e que satisfaz (8.22).*

Quando resolvemos uma equação diferencial ordinária de primeira ordem, normalmente obtemos uma familia de soluções que dependem de uma constante arbitrária. Se toda a solução particular puder ser obtida a partir da familia de soluções que encontramos por uma escolha apropriada da constante dizemos que a familia de soluções é a solução geral da equação.

8.5 Equações ordinárias de primeira ordem $y' = f(x, y)$

As equações diferenciais de primeira ordem são as equações que se podem escrever na forma

$$F\left(x, y, y'\right) = 0.$$

8.5.1 Existência e unicidade da solução: o teorema de Picard

Consideremos o problema de Cauchy

$$y'(x) = f(x,y) \tag{8.23}$$

$$y(x_0) = y_0 \tag{8.24}$$

Teorema 8.5.1 *Sejam $f(x,y)$ e $\frac{\partial f}{\partial y}$ duas funções contínuas em x e y num rectângulo fechado R com lados paralelos aos eixos. Se (x_0,y_0) é um ponto interior de R então, existe um número $h>0$ de tal forma que o problema de valores iniciais*

$$y' = f(x,y),\ \ y(x_0) = y_0 \tag{8.25}$$

admite uma e uma só solução $y = y(x)$ no intervalo $|x-x_0| \leq h$. A solução é o limite da sucessão

$$y_n(x) = y_0 + \int_{x_0}^{x} f[t, y_{n-1}(t)]\,dt, n = 0,1,2,\ldots \tag{8.26}$$

Exemplo 8.5.1 *Aplicando o método das aproximações sucessivas, determine a solução do problema Cauchy:*

$$y'(x) = f(x,y) \equiv y, \quad y(0) = 1 \tag{8.27}$$

Resolução 8.5.1 *Vamos construir a sucessão $y_n(x)$. Fazendo $y_0(x) = y(x_0)$, obtemos a seguinte sucessão $y_n(x)$:*

$$\begin{aligned}
y_0(x) &\equiv 1,\\
y_1(x) &= 1 + \int_0^x y_0 dt = 1 + x,\\
y_2(x) &= 1 + \int_0^x y_1 dt = 1 + \int_0^x (1+t)dt = 1 + x + \frac{x^2}{2},\\
y_3(x) &= 1 + \int_0^x y_2 dt = 1 + \int_0^x (1 + t + \frac{t^2}{2})dt = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!},
\end{aligned}$$

Generalizando, obtêm-se

$$y_n(x) = 1 + \int_0^x y_{n-1}(t)dt = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\frac{x^n}{n!}$$

Torna-se assim claro que

$$y_n(x) \to e^x \quad \textit{quando} \quad n \to \infty.$$

Substituindo na equação, verifica-se imediatamente que a função

$$y(x) = \lim_{n\to\infty} y_n(x) = e^x$$

é a solução exacta do problema de Cauchy (8.23), (8.24). ■

Exemplo 8.5.2 *Determine uma aproximação para a solução do problema de Cauchy*

$$y' = xy, \;\; y(0) \equiv y_0 = 1.$$

Resolução 8.5.2 *Pela fórmula de Picard, temos*

$$y = y_0 + \int_{x_0}^{x} f(t, y(t))\, dt = 1 + \int_0^x ty(t)\, dt. \tag{8.28}$$

Para $y_0(t) = y_0 = 1$ *temos*

$$\begin{aligned}
y_1(x) &= 1 + \int_0^x ty_0(t)\, dt = 1 + \int_0^x tdt = 1 + \frac{x^2}{2} && (8.29)\\
y_2(x) &= 1 + \int_0^x ty_1(t)\, dt = 1 + \int_0^x t\left(1 + \frac{t^2}{2}\right) dt = 1 + \frac{x^2}{2} + \frac{x^4}{8} && (8.30)\\
y_3(x) &= 1 + \int_0^x t\left(1 + \frac{t^2}{2} + \frac{t^4}{8}\right) = 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} && (8.31)\\
&\vdots
\end{aligned}$$

Os termos (8.29)-(8.31) são os primeiros termos do desenvolvimento em série de McLaurin da função $y(x) = e^{\frac{x^2}{2}}$*, solução da equação.* ■

8.5.2 O método de Euler

Este método é o método mais simples para a resolução de problemas de valor inicial. Consideremos o problema de Cauchy,

$$\begin{cases} y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0 \end{cases} \tag{8.32}$$

Utilizando (8.1), aproximamos a derivada em ordem ao tempo

$$\frac{dy}{dt} \approx \frac{y(t_{n+1}) - y(t_n)}{\Delta t},$$

onde $\Delta t = t_{n+1} - t_n$

e obtemos a relação de recorrência:

$$y(t_{n+1}) = y(t_n) + \Delta t f(t_k, y_k), \;\; k = 0, 1, 2, \ldots \tag{8.33}$$

Exemplo 8.5.3 *Considere o problema de Cauchy,*

$$y' = x + y, \;\; y(0) = 1.$$

Determine uma aproximação para o valor de y *no intervalo* $[0, 1]$ *considerando* $n = 20$*.*

Resolução 8.5.3 *Se* $n = 20$ *então* $h = \Delta t = \frac{1-0}{20} = 0.05$. *Consequentemente, obtemos a tabela*

$$\begin{aligned} y_1 &= y_0 + \Delta t f(x_0, t_0) \\ &= 1 + \Delta t (x_0, y_0) \\ &= 1 + 0.05(0.1) = 1.05 \\ y_2 &= 1.05 + 0.05(0.05 + 1.05) \\ &= 1.105 \\ &\vdots \\ y_{20} &= 3.307 \end{aligned}$$

Reduzindo $\Delta t = h$ *para* 0.005, *o esforço computacional cresce mas é de esperar que a precisão aumente a cada passo de integração. Quando* $\Delta t = 0.005$ *o valor de* y *em* $t = 1$ *é de* 3.483. *O valor real é igual a* 3.437 *(com quatro algarismos significativos) pelo que o erro com* $h = 0.05$ *em* $y(1)$ *é igual a* 0.130. *Quando* $\Delta t = 0.005$ *o erro é igual a* 0.014. ■

Exemplo 8.5.4 *Utilizando o método de Euler, indique uma aproximação para a solução do problema de valores iniciais:*

$$\begin{cases} y' &= -y \\ y(0) &= 2 \end{cases}$$

Resolução 8.5.4 *A solução exacta da equação é* $y(x) = 2e^{-t}$.

Utilizando o Método De Euler, obtemos o seguinte esquema de recorrência

$$\begin{aligned} y(x_{n+1}) &= y_{x_n} + hf(x_n, y(y_n)) \\ y_{n+1} &= y_n + \frac{1}{10}(-y_n) = \frac{9}{10} y_n \end{aligned} \tag{8.34}$$

onde $y(x_0) = y_0 = 2$. *Cada ponto é definido pela fórmula*

$$x_{i+1} = x_0 + i\frac{b-a}{10}.$$

Como $\mathcal{I} = [0,1]$ *e vamos considerar* 10 *intervalos, obtemos a fórmula*

$$x_{i+1} = \frac{i}{10}, \text{ para } 0 \leq i \leq 10.$$

Portanto, vamos obter seguinte tabela

x_i	$y(x_i)$	$\tilde{y}(x_i)$	$\lvert y(x_i) - \tilde{y}(x_i)\rvert$
0.100	1.8096748	1.8000000	0.0096748
0.200	1.6374615	1.6200000	0.0174615
0.300	1.4816364	1.4580000	0.0236364
0.400	1.3406401	1.3122000	0.0284401
0.500	1.2130613	1.1809800	0.0320813
0.600	1.0976233	1.0628820	0.0347413
0.700	0.9931706	0.9565938	0.0365768
0.800	0.8986579	0.8609344	0.0377235
0.900	0.8131393	0.7748410	0.0382983
1.000	0.7357589	0.6973569	0.0384020

■

8.5.3 O método de Runge-Kutta

É a aproximação mais simples depois do método de Euler. O método de Euler pode ser refinado quando tomamos a inclinação média para extrapolar a função até o próximo ponto. O método de Runge-Kutta leva essa ideia mais longe e usa a média ponderada das inclinações no intervalo. Usando o método de Euler determina-se o valor da derivada y no meio do intervalo $x + \frac{\Delta x}{2}$ e depois usa-se essa derivada para calcular o novo valor de y.

O método de Runge-Kutta de segunda ordem é:

$$y_{n+1} = y_n + \frac{k}{2}(\kappa_1 + \kappa_2), \tag{8.35}$$

onde

$$\kappa_1 = f(x_n, y_n), \quad \kappa_2 = f(x_{n+1}, y_{\text{Euler}}) \tag{8.36}$$

onde $y_{\text{Euler}} = y_n + h\kappa_1 = y_n + hf(x_n, y_n)$.

A aproximação y_{n+1} no método de Runge-Kutta de quarta ordem é dada por

$$y_{n+1} = y_n + \frac{h}{6}(\kappa_1 + 2\kappa_2 + 2\kappa_3 + \kappa_4), \tag{8.37}$$

onde h representa o avanço em x e

$$\kappa_1 = f(x_n, y_n), \kappa_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{\kappa_1}{2}\right), \kappa_3 = f\left(x_n + \frac{h}{2}, y_n + \frac{\kappa_2}{2}\right), \kappa_4 = f(x_{n+1}, y_n + \kappa_3)$$

Exemplo 8.5.5 *Utilizando o método de Runge-Kutta de segunda ordem, indique uma aproximação para a solução do problema de Cauchy,*

$$\begin{cases} y' &= y + x \\ y(0) &= 1 \end{cases}$$

para $x \in [0, 1]$, com $n = 10$.

Compare os resultados obtidos com a solução exacta do problema, $y(x) = -x - 1 + 2e^x$.

Resolução 8.5.5 *Como $n = 10$, consideramos que o intervalo $[0,1]$ se encontra dividido em 10 subintervalos da forma com extremos dados por*

$$x_i = x_0 + i\frac{1-0}{10} = \frac{i}{10},\ \textit{para } 0 \leq i \leq 10.$$

A formula de recorrência é então

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{2}(\kappa_1 + \kappa_2), \\ &= y_n + \frac{\kappa_1 + \kappa_2}{20}, \end{aligned} \tag{8.38}$$

onde κ_1 e κ_2 são dados por

$$\kappa_1 = f(x_n, y_n) = x_n + y_n,\ \ \kappa_2 = f(x_n + h, y_n + h\kappa_1) = x_n + \frac{1}{10} + y_n + \frac{x_n + y_n}{10}, \tag{8.39}$$

onde $y_n = y(x_n)$.

No que se segue, a solução aproximada será representada por $\tilde{y}_i$.

Obtemos então a seguinte tabela:

x_i	$\tilde{y}_i$	$y_i = y(x_i))$	$\lvert\tilde{y} - y_i\rvert$	κ_1	κ_2
0.000	1.0000000	1.0000000	0.0000000		
0.100	1.1100000	1.1103418	0.0003418	1.0000000	1.2000000
0.200	1.2420500	1.2428055	0.0007555	1.2100000	1.4310000
0.300	1.3984653	1.3997176	0.0012524	1.4420500	1.6862550
0.400	1.5818041	1.5836494	0.0018453	1.6984653	1.9683118
0.500	1.7948935	1.7974425	0.0025490	1.9818041	2.2799845
0.600	2.0408574	2.0442376	0.0033802	2.2948935	2.6243829
0.700	2.3231474	2.3275054	0.0043580	2.6408574	3.0049431
0.800	2.6455778	2.6510819	0.0055040	3.0231474	3.4254621
0.900	3.0123635	3.0192062	0.0068427	3.4455778	3.8901356
1.000	3.4281617	3.4365637	0.0084020	3.9123635	4.4035999

Mostre que (8.38)-(8.39) é equivalente à relação de recorrência:

$$\begin{cases} y_0 = y(x_0) = 1 \\ y_{n+1} = \dfrac{21}{200}x_n + \dfrac{221}{200}y_n + + \dfrac{1}{200} \end{cases} \quad n\text{=0,}\ldots,\ 9. \tag{8.40}$$

■

Exemplo 8.5.6 *Utilizando o método de Runge-Kutta de quarta ordem, indique uma aproximação para a solução do problema de Cauchy,*

$$\begin{cases} y' = y + x \\ y(0) = 1 \end{cases}$$

para $x \in [0,1]$, *com* $n = 10$.

Compare os resultados obtidos com a solução exacta do problema, $y(x) = -x - 1 + 2e^x$.

Resolução 8.5.6 *Como* $n = 10$, *consideramos que o intervalo* $[0,1]$ *se encontra dividido em* 10 *subintervalos da forma com extremos dados por*

$$x_i = x_0 + i\frac{1-0}{10} = \frac{i}{10}, \text{ para } 0 \leq i \leq 10.$$

A formula de recorrência é então

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{2}(\kappa_1 + \kappa_2), \\ &= y_n + \frac{\kappa_1 + \kappa_2}{20}, \end{aligned} \tag{8.41}$$

onde κ_1 *e* κ_2 *são dados por*

$$\kappa_1 = f(x_n, y_n) = x_n + y_n, \quad \kappa_2 = f(x_n + h, y_n + h\kappa_1) = x_n + \frac{1}{10} + y_n + \frac{x_n + y_n}{10}, \tag{8.42}$$

onde, $y_n = y(x_n)$.

Obtemos então a seguinte tabela:

x_i	$\tilde{y}_i$	$y_i = y(x_i)$	$\lvert\tilde{y} - y_i\rvert$	κ_1	κ_2	κ_3	κ_4
0.000	1.0000000	1.0000000	0.0000000				
0.100	1.1103417	1.1103418	0.0000002	1.0000000	1.1000000	1.1050000	1.2105000
0.200	1.2428051	1.2428055	0.0000004	1.2103417	1.3208588	1.3263846	1.4429801
0.300	1.3997170	1.3997176	0.0000006	1.4428051	1.5649454	1.5710524	1.6999104
0.400	1.5836485	1.5836494	0.0000009	1.6997170	1.8347028	1.8414521	1.9838622
0.500	1.7974413	1.7974425	0.0000013	1.9836485	2.1328309	2.1402900	2.2976775
0.600	2.0442359	2.0442376	0.0000017	2.2974413	2.4623133	2.4705569	2.6444970
0.700	2.3275033	2.3275054	0.0000022	2.6442359	2.8264477	2.8355583	3.0277918
0.800	2.6510791	2.6510819	0.0000027	3.0275033	3.2288784	3.2389472	3.4513980
0.900	3.0192028	3.0192062	0.0000034	3.4510791	3.6736331	3.6847608	3.9195552
1.000	3.4365595	3.4365637	0.0000042	3.9192028	4.1651630	4.1774610	4.4369489

■

8.6 Exercícios

1. Considere o problema de valores iniciais:

$$\begin{cases} y' &= ty \\ y(0) &= 1 \end{cases}$$

Mostre que o problema admite uma e uma só solução.

2. Prove que o problema de valores iniciais:

$$\begin{cases} y' &= \frac{1}{1+y^2} \\ y(a) &= \alpha \end{cases}$$

admite uma e uma só solução para $t \in [a, b]$.

3. Utilize o método de Picard para obter a solução dos problemas de valores iniciais

a)

$$\begin{cases} y' &= y - 1 \\ y(0) &= 2 \end{cases}$$

b)

$$\begin{cases} y' &= y + 3t \\ y(0) &= 1 \end{cases}$$

c)

$$\begin{cases} y' &= -2y \\ y(0) &= 1 \end{cases}$$

4. Utilize o método de Euler para obter uma solução aproximada para cada um dos seguintes problemas de valores iniciais

a)

$$\begin{cases} y' &= y - 1 \\ y(0) &= 2 \end{cases}, t \in [0, 1],$$

com $h = 0.1$ e $h = 0.25$.

b)

$$\begin{cases} y' &= 2y + 3 \\ y(0) &= 1 \end{cases}, t \in [0, 1],$$

com $h = 0.1$ e $h = 0.25$.

c)

$$\begin{cases} y' &= -2ty \\ y(0) &= 1 \end{cases}, t \in [0, 1],$$

com $h = 0.1$ e $h = 0.25$.

5. Dado o problema de Cauchy

$$\begin{cases} y' &= \frac{t}{1+y^2} \\ y(0) &= 1 \end{cases} , t \in [0,1],$$

utilize o método de Runge-Kutta de segunda ordem para apresentar uma estimativa para os valores da solução, $y(t)$ no intervalo $[0,1]$. Utilizando o método das diferenças divididas, determine a função interpoladora de $y(t)$

Bibliografia

[1] Paulo Rebelo & Amilcar Miranda, *Álgebra Linear e Geometria Analítica segundo as aulas do Prof. Doutor Sampaio Martins,* Serviços gráficos da Universidade da Beira Interior, 2005.

[2] Paulo Rebelo & Amilcar Miranda, *Álgebra Linear e Numérica, Parte I,* Serviços gráficos da Universidade da Beira Interior, 2005.

[3] Maria Raquel Valença, *Análise Numérica,* Lisboa: Universidade Aberta, 1996.

[4] Heitor Pina, *Métodos Numéricos,* Mc Graw-Hill, 1998.

[5] Carlos Lemos & Heitor Pina, *Métodos Numéricos: Complementos e guia prático,* Instituto Superior Técnico 2006.

[6] José Duque, Sebenta de Matemática Computacional, Universidade da Beira Interior.

[7] Richard L. Burden, *Numerical Analysis,* Boston, Prindle, 1981.

[8] Ian Jacques, *Numerical Analysis,* Chapman Hall; London, 1987

[9] Donald Greenspan, *Numerical Analysis For Applied Mathematics,Science And Engineering,* Redwood; Addison-Wesley, 1988.

[10] Steven Chapra, *Numerical Methods For Engineers,* Mcgraw-Hill; New York, 1989.

[11] Mahinder Kumar Jain, *Numerical solution of differential equations,* New Delhi; Wiley Eastern, cop. 1979.

[12] R.I. Burden & J.D. Faires, *Numerical Analysis 7e,* PWS-Kent, Boston, 2001.

[13] Mario J. Miranda & Paul L. Fackler, *Applied Computational Economics and Finance,* The MIT Press, 2002.

[14] J.C. Butcher, *The Numerical Analysis of Ordinary Differential Equations,* John Wiley & Sons, Auckland, 1987.

[15] S.D. Conte & C. de Boor, *Elementary Numerical Analysis,* Mc Graw-Hill, NY, 1980.

[16] E. Hairer, S.P. Nørsett & G. Wanner, *Solving Ordinary Differential Equations I,* Springer Series in Comput. Mathematics, Vol. 8, Springer-Verlag, Heidelberg, 1987.

[17] E. Hairer & G. Wanner, *Solving Ordinary Differential Equations II,* Springer Series in Comput. Mathematics, Vol. 14, Springer-Verlag, Heidelberg, 1991.

[18] J.D. Lambert, *Numerical Methods for Ordinary Differential Systems,* John Wiley & Sons, Chichester, 1991.

[19] Carlos J. S. Alves, *Resumo da matéria teórica de Análise Numérica,* `www.math.ist.utl.pt/~calves/`.

[20] Adérito Araujo, *Métodos Numéricos: Complementos e guia prático,* `http://www.mat.uc.pt/~alma/aulas/anem/sebenta/`.