

TRATAMENTO MATEMÁTICO DAS OBSERVAÇÕES

Adérito Luís Martins Araújo
José Augusto Mendes Ferreira

Notas de apoio às aulas de Tratamento Matemático das Observações
da Licenciatura em Engenharia Geográfica, no ano lectivo de 2002/2003.



FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Capítulo 1

Preliminares

1.1 Introdução

Para modelar um fenómeno natural, social ou outro, é necessário percorrer um conjunto de tarefas bem definidas: recolher os dados; compreender como é que o fenómeno em estudo interage com o “resto do mundo”; escrever, em termos matemáticos, um sistema de leis que descrevam o comportamento que desejamos compreender.

No entanto, este procedimento ainda está longe de estar completo. O estabelecimento, puro e simples, das leis que descrevem o modelo não garante que estas possam ser usadas para descrever ou prever o fenómeno físico. Uma vez construído o modelo temos que o sujeitar a um considerado “fogo cruzado” de análise matemática (testar o modelo).

Passado este passo, e tendo já uma grande dose de confiança em relação ao modelo, o objectivo a ter em conta será encontrar a sua solução exacta ou, pelo menos, ter uma ideia bastante precisa em relação a essa solução. Apesar da análise matemática fornecer importantes informações qualitativas a respeito da solução, esta falha redondamente, na maioria dos casos, na resposta ao nosso objectivo. O próximo passo é pois a obtenção de um método numérico que permita construir uma solução aproximada para o problema. Um método numérico, nestas condições, é traduzido por um algoritmo que não é mais do que um completo e não ambíguo conjunto de passos que conduzem à solução do problema.

1.2 Erro absoluto e erro relativo

A introdução de erros num determinado processo de cálculo pode ter várias causas. É nosso objectivo analisar quais são essas causas e estudar mecanismos que nos permitam determinar limites superiores para os erros obtidos no final do processo de cálculo.

Para iniciar o nosso estudo, definamos dois tipos fundamentais de erros.

Definição 1.1 (Erro absoluto) *Seja $x \in \mathbb{R}^n$ um vector cujas componentes são desconhecidas e $\bar{x} \in \mathbb{R}^n$ um vector cujas componentes são aproximações para as componentes correspondentes de x . Chama-se erro absoluto de \bar{x} e representa-se por $e(\bar{x})$ a quantidade*

$$e(\bar{x}) = x - \bar{x}.$$

Na prática o valor do erro absoluto usa-se, geralmente, em norma pois, para a maioria dos problemas, não é relevante saber se o erro foi cometido por defeito ou por excesso. Vamos, então introduzir o conceito de norma vectorial.

Definição 1.2 (Norma) *Seja E um espaço vectorial (real ou complexo). A aplicação $\|\cdot\| : E \rightarrow \mathbb{R}_0^+$ que verifica*

1. $\forall x \in E, \quad \|x\| = 0 \Leftrightarrow x = 0,$
2. $\forall x \in E, \forall \lambda \in \mathbb{R} \text{ (ou } \mathbb{C}), \quad \|\lambda x\| = |\lambda| \|x\|,$
3. $\forall x, y \in E, \quad \|x + y\| \leq \|x\| + \|y\|,$

é designada por norma.

Como consequência da propriedade 3 da definição anterior temos

$$\|u\| = \|u - v + v\| \leq \|u - v\| + \|v\|$$

e portanto $\|u\| - \|v\| \leq \|u - v\|$. Logo vale a seguinte propriedade.

Teorema 1.3 *Seja E um espaço vectorial e $\|\cdot\|$ uma norma em E . Então*

$$|\|u\| - \|v\|| \leq \|u - v\|, \forall u, v \in E.$$

Existem várias funções que verificam as três propriedades das normas vectoriais. Entre elas destacam-se as dadas no Exercício 1.5.2.

Definição 1.4 (Erro relativo) *Seja $x \in \mathbb{R}^n$ um vector cujas componentes são desconhecidas e $\bar{x} \in \mathbb{R}^n$ um vector cujas componentes são aproximações para as componentes correspondentes de x . Chama-se erro relativo de \bar{x} e representa-se por $r(\bar{x})$ a quantidade*

$$r(\bar{x}) = e(\bar{x})/\|x\|.$$

Como na definição de erro relativo o valor de x não é conhecido, é usual considerar a aproximação $\|r(\bar{x})\| \approx \|e(\bar{x})\|/\|\bar{x}\|$. Melhor ainda, atendendo a que

$$\|x\| \geq \|\bar{x}\| - \|e(\bar{x})\|,$$

podemos considerar o majorante

$$\|r(\bar{x})\| \leq \frac{\|e(\bar{x})\|}{\|\bar{x}\| - \|e(\bar{x})\|}.$$

Observação 1.5 *O erro relativo, atendendo a que é uma quantidade adimensionada, é muitas vezes representado sob a forma de percentagem. Note-se também que o erro relativo nos dá uma maior informação quanto à precisão da aproximação que o erro absoluto.*

É com base nas definições de erro absoluto e erro relativo que iremos analisar os resultados numéricos que aparecerão como aproximações a valores que não conhecemos com exactidão.

1.3 Condicionamento de matrizes

Consideremos um sistema possível e determinado $Ax = b$ e seja \bar{b} o vector obtido a partir de b considerando perturbações numéricas nas suas componentes. Esta situação é frequente quando o vector dos termos independentes representa medições. Para analisar de que forma as perturbações dos termos independentes influenciam o resultado numérico, há necessidade de introduzir o conceito de norma de uma matriz.

Uma vez que o conjunto das matrizes reais de ordem n , que designaremos

por

$\mathcal{M}_{n \times n}(\mathbb{R})$, é um espaço vectorial, poderemos definir normas nesse espaço.

Definição 1.6 (Norma matricial) Consideremos $\|\cdot\|$ uma norma definida em \mathbb{R}^n . Então a aplicação $\|\cdot\| : \mathcal{M}_{n \times n}(\mathbb{R}) \rightarrow \mathbb{R}_0^+$ tal que, para todo o $A \in \mathcal{M}_{n \times n}(\mathbb{R})$

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

é designada norma de uma matriz subordinada à norma vectorial $\|\cdot\|$.

Podemos definir normas matriciais subordinadas às diferentes normas vectoriais dadas no Exercício 1.5.2. Consideremos a norma $\|\cdot\|_1$ para \mathbb{R}^n . Então, com $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, tem-se

$$\|A\|_1 = \sup_{\|x\|=1} \|Ax\|_1 = \sup_{\|x\|=1} \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}x_j \right|$$

e

$$\sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}x_j \right| \leq \sum_{j=1}^n \sum_{i=1}^n |a_{ij}| |x_j| \leq \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}| \|x\|_1$$

e portanto, concluímos

$$\|A\|_1 \leq \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|.$$

Temos então que (o máximo é atingido para um vector $x \in \mathbb{R}^n$ escolhido de forma conveniente)

$$\|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|,$$

o que mostra que a norma anterior é subordinada à norma $\|\cdot\|_1$.

Consideremos agora em \mathbb{R}^n a norma $\|\cdot\|_\infty$ e seja $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Então

$$\|A\|_\infty = \sup_{\|x\|_\infty=1} \|Ax\|_\infty.$$

Atendendo a que

$$\|Ax\|_\infty = \max_{i=1, \dots, n} \left| \sum_{j=1}^n a_{ij}x_j \right| \leq \left(\max_{i=1, \dots, n} \sum_{j=1}^n |a_{ij}| \right) \max_{j=1, \dots, n} |x_j|$$

vem

$$\|A\|_\infty \leq \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|.$$

Atendendo a que o máximo é atingido para um vector $x \in \mathbb{R}^n$ escolhido de forma conveniente concluimos que a norma $\|A\|_\infty$ é subordinada à norma vectorial $\|\cdot\|_\infty$.

Provemos agora alguns resultados importantes referentes a normas matriciais.

Teorema 1.7 *Seja $\|\cdot\|$ uma norma em \mathbb{R}^n . Então, para $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, tem-se*

$$\|Ax\| \leq \|A\| \|x\|, \forall x \in \mathbb{R}^n,$$

em que $\|A\|$ é a norma de A subordinada à norma $\|\cdot\|$.

Demonstração: Notamos que se a norma $\|\cdot\|$ é subordinada a uma norma vectorial então, para $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, temos

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \geq \frac{\|Ax\|}{\|x\|}, x \in \mathbb{R}^n, x \neq 0,$$

e portanto é válido o resultado. \square

Teorema 1.8 *Para $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$, tem-se $\|AB\| \leq \|A\| \|B\|$.*

Demonstração: Atendendo à desigualdade demonstrada no teorema anterior, temos

$$\|AB\| = \sup_{\|x\|=1} \|ABx\| \leq \|A\| \sup_{\|x\|=1} \|Bx\| = \|A\| \|B\|. \quad \square$$

Consideremos, de novo, o sistema possível e determinado $Ax = b$ e o vector \bar{b} obtido a partir de b considerando perturbações numéricas nas suas componentes. Seja $e(\bar{b})$ e $r(\bar{b})$ respectivamente os erros absoluto e relativo de \bar{b} . Vejamos de que modo estes erros influenciam os erros absoluto e relativo de \bar{x} , sendo \bar{x} a solução do sistema $A\bar{x} = \bar{b}$.

Se o sistema $Ax = b$ é possível e determinado, então A é invertível e portanto $x = A^{-1}b$. Consideremos agora o sistema em que o vector

dos termos independentes tem as componentes afectadas de erro, i.e, $A\bar{x} = \bar{b}$. Temos

$$r(x) = \frac{\|e(x)\|}{\|x\|} = \frac{\|A^{-1}(b - \bar{b})\|}{\|x\|} = \frac{\|A^{-1}e(b)\|}{\|x\|}.$$

Mas

$$\|b\| = \|Ax\| \leq \|A\|\|x\|$$

e portanto

$$\|x\| \geq \|A\|^{-1}\|b\|.$$

Utilizando esta desigualdade em $r(x)$, deduzimos

$$\|r(x)\| \leq \frac{\|A^{-1}\| \|e(b)\|}{\|A\|^{-1}\|b\|} = \|A^{-1}\| \|A\| \|r(b)\|.$$

Esta desigualdade permite determinar uma condição suficiente para garantir que pequenas variações nas componentes do vector dos termos independentes do sistema conduz a pequenas variações nas componentes do vector solução. Se o sistema apresenta a propriedade anterior dizemos que é um sistema estável e a matriz do sistema diz-se bem condicionada. Se o sistema de equações lineares não é estável então diz-se instável e a matriz do sistema diz-se mal condicionada.

A $\|A^{-1}\| \|A\|$ chamamos número de condição da matriz A e é denotado por $\text{cond}(A)$.

Teorema 1.9 *Se A é tal que $\text{cond}(A) \leq 1$ então o sistema que tem A como matriz é um sistema estável.*

1.4 O polinómio de Taylor

Seja f uma função real definida num intervalo $[a, b] \subseteq \mathbb{R}$. Um problema que frequentemente se coloca é o de determinar uma função g definida em $[a, b]$ tal que $\|f - g\|_\infty < \epsilon$, com $\epsilon > 0$ uma tolerância dada, onde $\|\cdot\|_\infty$ é a norma de Chebyshev definida no Exercício 1.5.10. A existência de solução para tal problema é dada pelo Teorema de Weierstrass que apresentamos sem demonstração.

Teorema 1.10 (Weierstrass) *Seja f uma função contínua definida em $[a, b]$. Então, para cada $\epsilon > 0$ existe um polinómio p definido em $[a, b]$ tal que*

$$\|f - p\|_\infty < \epsilon.$$

Notemos a grande importância deste resultado. De acordo com ele, podemos ter a certeza que dada uma função contínua f qualquer existe sempre um polinómio p que está tão próximo de f quanto se queira. Assim sendo, este resultado legitima a aproximação polinomial, isto é, a tarefa de, dada uma função, procurar um polinómio que a aproxime.

Notemos, no entanto, que o teorema não nos diz como podemos construir esse polinómio; ele apenas garante a existência.

Consideremos agora o seguinte teorema, devido a Brook Taylor.

Teorema 1.11 (Taylor) *Se f admite derivadas contínuas até à ordem n (inclusivé) em $[a, b]$, isto é, se $f \in C^n([a, b])$, e se $f^{(n+1)}$ existir em (a, b) então, para todo o $x, x_0 \in [a, b]$,*

$$f(x) = T_n(x; x_0) + R_n(x; x_0), \quad (1.1)$$

onde

$$T_n(x; x_0) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

e

$$R_n(x; x_0) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)^{n+1}, \quad \xi \in I\{x, x_0\},$$

sendo $I\{x, x_0\}$ o intervalo aberto definido por x e x_0 .

A (1.1) chamaremos fórmula de Taylor sendo $T_n(x; x_0)$ o polinómio de Taylor de f em torno do ponto x_0 e $R_n(x; x_0)$ o resto (de Lagrange) de ordem n (ou de grau $n+1$). Se $x_0 = 0$ a (1.1) chamaremos fórmula de Maclaurin.

Atente-se ao grande interesse prático deste resultado que afirma que, mediante certas condições, uma função qualquer poder ser escrita como a soma de um polinómio com um resto. Escolhendo valores de x e x_0 tais que

$$\lim_{n \rightarrow +\infty} R_n(x; x_0) = 0. \quad (1.2)$$

temos que, a partir de um valor de n suficientemente grande, a função dada pode ser aproximada pelo polinómio de Taylor. Assim, qualquer operação a efectuar sobre a função (derivação, integração, etc.) poderá ser feita sobre o polinómio. Uma função que verifica (1.2) para todo o $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$, ε positivo e escolhido de forma adequada, diz-se analítica em x_0 .

Observação 1.12 A escolha dos valores de x e x_0 deverá ser feita de modo a que eles pertençam ao intervalo de convergência da série

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

designada por série de Taylor. Neste curso não iremos dar ênfase a esta questão.

O objectivo fundamental será determinar qual menor valor de n que verifica

$$\max_{\xi \in I\{x, x_0\}} |R_n(x; x_0)| < \eta,$$

sendo $\eta > 0$ uma tolerância previamente fixada. Obtemos assim a aproximação

$$f(x) \approx T_n(x; x_0),$$

cujo erro não excede η . O valor de $R_n(x; x_0)$, sendo um erro absoluto uma vez que

$$|f(x) - T_n(x; x_0)| = |R_n(x; x_0)|,$$

é também designado erro de truncatura.

1.5 Exercícios

Exercício 1.5.1 A equação do segundo grau $ax^2 + bx + c = 0$ é usualmente resolvida pelas fórmulas

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \quad (1.3)$$

1. Prove que uma solução alternativa é dada por

$$x_1 = \frac{2c}{-b - \sqrt{b^2 - 4ac}}, \quad x_2 = \frac{2c}{-b + \sqrt{b^2 - 4ac}}. \quad (1.4)$$

2. Escreva um programa de computador que resolva equações de segundo grau de duas maneiras distintas: (i) usando as fórmulas (1.3); (ii) calculando uma raiz pela fórmula de (1.3) em que não se subtraem números do mesmo sinal e a outra raiz pela fórmula de (1.4) adequada.

3. Execute o programa construído em (b) quando: (i) $a = 1.0$, $b = -5.0$, $c = 6.0$; (ii) $a = 1.0$, $b = 12345678.03$, $c = 0.92$.

Exercício 1.5.2 Consideremos $E = \mathbb{R}^n$. Prove que as funções seguintes são normas

- $\|x\|_1 = \sum_{i=1}^n |x_i|$, (norma um)
- $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$, (norma euclidiana).
- $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$, (norma do máximo ou de Chebyshev),

Exercício 1.5.3 Mostre que a função estabelecida na Definição 1.6 é uma norma.

Exercício 1.5.4 Seja $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Prove que as seguintes aplicações verificam as propriedades que caracterizam um norma:

- $\|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|$;
- $\|A\|_\infty = \max_{i=1, \dots, n} \sum_{j=1}^n |a_{ij}|$.

Exercício 1.5.5 Seja A uma matriz real, não singular e de ordem n . Prove que se λ é um valor próprio de A então

$$\frac{1}{\|A^{-1}\|} \leq |\lambda| \leq \|A\|.$$

Exercício 1.5.6 Considere a matriz

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

Calcule $\|A\|_1$ e $\|A\|_\infty$.

Exercício 1.5.7 Ao resolver o sistema de equações lineares $Ax = b$, suponha que o termo independente b não é exacto, encontrando-se afectado de um erro δb . Qual o erro que vem para x resultante dessa inexactidão?

Exercício 1.5.8 Como vimos, número de condição de uma matriz A é definido por $\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$.

1. Mostre que uma medida do número de condição pode ser dada por λ_M/λ_m , onde λ_M e λ_m são, respectivamente, o maior e menor (em módulo) valores próprios de A .

2. Mostre que se A é singular $\text{cond}(A)$ é infinito e se A é não singular $\text{cond}(A) \geq 1$.

3. Resolva o sistema

$$\begin{cases} 2.000112x_1 + 1.414215x_2 = 0.521471 \\ 1.414215x_1 + 1.000105x_2 = 0.232279 \end{cases}$$

pelo método de eliminação de Gauss. Sabendo que a sua solução exacta é

$$(x_1, x_2) = (607.1248, -858.2826),$$

explique os resultados obtidos

Exercício 1.5.9 As matrizes dos sistemas

$$\begin{cases} x - y = 1 \\ x - 1.00001y = 0 \end{cases} \quad \text{e} \quad \begin{cases} x - y = 1 \\ x - 0.99999y = 0 \end{cases}$$

são aproximadamente iguais. Determine e compare as suas soluções.

Exercício 1.5.10 Num espaço de funções podemos também definir uma norma. Consideremos o conjunto das aplicações contínuas num intervalo $[a, b]$. Este conjunto é um espaço vectorial para a soma de funções e produto de um número real por uma função. Prove que, em $C[a, b]$, são normas as aplicações seguintes:

- $\|f\|_\infty = \max_{x \in [a, b]} |f(x)|$, (norma de Chebyshev);

- $\|f\|_2 = \sqrt{\int_a^b [f(x)]^2 dx}$, (norma L_2).

Exercício 1.5.11 Calcule um valor aproximado de $\cos 47^\circ$ utilizando o polinómio de Taylor com resto de grau 3 encontrado para a função $f(x) = \cos x$ em torno de $x_0 = \frac{\pi}{4}$. Indique o grau de precisão do resultado obtido.

Exercício 1.5.12 Utilizando o desenvolvimento em série de Taylor de $\ln(1+x)$ em torno de $x_0 = 0$, calcule $\ln\left(\frac{5}{4}\right)$ com duas casas decimais correctas.

Exercício 1.5.13 Obtenha uma aproximação com três casas decimais correctas para $\ln(1.25)$.

Exercício 1.5.14 Pretende-se calcular uma aproximação do real $A = \frac{8}{\pi} \sin \frac{\pi}{8}$ com três casas decimais correctas desenvolvendo $\frac{\sin x}{x}$ em série de Taylor.

1. Determine o menor número de termos a tomar naquele desenvolvimento de modo a obter a precisão referida.
2. Calcule A de acordo com a alínea anterior e apresente o erro absoluto cometido.

Exercício 1.5.15 Determine $\int_0^{\frac{1}{2}} \frac{\sin x}{x} dx$ com três casas decimais correctas.

Exercício 1.5.16 Desenvolva em série de Taylor a função $f(x) = \frac{1}{x} \int_0^x e^{-t^2} dt$ e indique um limite superior para o erro de truncatura.

Capítulo 2

Interpolação polinomial de funções de uma variável

2.1 Introdução

Seja f uma função real definida num intervalo $[a, b] \subseteq \mathbb{R}$ e conhecida nos pontos $x_0, x_1, \dots, x_n \in [a, b]$. Suponhamos ainda, sem perda de generalidade, que esses pontos formam uma partição Δ do intervalo $[a, b]$, isto é,

$$\Delta : \quad a = x_0 < x_1 < \dots < x_{n-1} < x_n = b. \quad (2.1)$$

Como poderemos calcular o valor de $f(x)$ e $x \in [a, b]$, tal que $x \neq x_i$, $i = 0, 1, \dots, n$? Este problema pode ser resolvido por interpolação. Em linhas gerais, o conceito de interpolação consiste em determinar uma função

$$g(x) = a_0\phi_0(x) + \dots + a_n\phi_n(x), \quad (2.2)$$

gerada por uma certa família de funções $\{\phi_k\}_{k=0}^n$, por forma a que

$$f(x_i) = g(x_i), \quad i = 0, 1, \dots, n. \quad (2.3)$$

A função g nestas condições é designada por função interpoladora de f nos pontos de suporte (ou pontos de interpolação) x_0, x_1, \dots, x_n . A condição (2.3) é designada condição de interpolação.

Observação 2.1 *Nada nos garante que o problema da interpolação tenha sempre solução. Por exemplo, fazendo $\phi_0(x) = 1$ e $\phi_1(x) = x^2$, não existe nenhuma função $g(x) = a_0 + a_1x^2$ que passe nos pontos $(1, 1)$ e $(-1, 0)$.*

O problema da interpolação tem uma grande importância prática, sobretudo no tratamento de funções para as quais se conhece apenas um conjunto finito de valores. Tal situação é muito frequente, por exemplo, no contexto das equações diferenciais. Quando se usam métodos numéricos para aproximar a solução de uma equação diferencial esta fica apenas conhecida num conjunto de pontos. A interpolação permite assim encontrar uma função que passa por esse conjunto de pontos e que pode funcionar como uma aproximação à solução da equação diferencial.

Outra utilidade prática da interpolação consiste na aproximação de uma função com uma expressão analítica complicada por uma outra função mais simples que coincida com a primeira num determinado conjunto de pontos. Assim, quando pretendermos operar com a primeira função podemos fazê-lo, de uma forma aproximada, recorrendo à função interpoladora.

2.2 Interpolação polinomial de Lagrange

Vamos considerar o problema de determinar um polinómio $P \in \mathcal{P}_n(\mathbb{R})$ (espaço vectorial dos polinómios de grau menor ou igual a n com coeficientes reais) que interpole a função f nos pontos da partição Δ . Neste caso, as funções geradoras $\{\phi_k\}_{k=0}^n$ terão de constituir uma base de $\mathcal{P}_n(\mathbb{R})$, espaço de dimensão $n + 1$. Um exemplo é dado pela base canónica

$$\phi_k(x) = x^k, \quad k = 0, 1, \dots, n. \quad (2.4)$$

Observação 2.2 *Iremos escrever $\mathcal{P}_n(\mathbb{R}) = \mathcal{P}_n([a, b])$ para evidenciar o facto de estarmos a trabalhar no intervalo $[a, b]$.*

Vamos começar por demonstrar o seguinte teorema.

Teorema 2.3 (Lagrange) *Seja f uma função definida nos pontos da partição (2.1) do intervalo $[a, b] \subseteq \mathbb{R}$. Existe um e um só polinómio P_n de grau menor ou igual a n interpolador de f nos pontos dados.*

Demonstração: Considerando em $\mathcal{P}_n([a, b])$ a base canónica (2.4) temos que $P_n \in \mathcal{P}_n([a, b])$ pode ser escrito na forma

$$P_n(x) = \sum_{j=0}^n c_j x^j.$$

Para que este polinómio verifique as condições de interpolação

$$f(x_i) = P_n(x_i), \quad i = 0, \dots, n,$$

os seus coeficientes serão determinados pela resolução do sistema linear

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}.$$

Como $x_i \neq x_j$ para $i \neq j$, temos que a matriz deste sistema é uma **matriz de Vandermonde** invertível (Exercício 2.2.4). Temos que este sistema é possível e determinado e, como tal, podemos dizer que existe um e um só polinómio P_n de grau menor ou igual a n que interpola f nos pontos da partição dada. \square

A determinação do polinómio interpolador por este processo é pouco eficiente e pouco estável. Quanto à eficiência, note-se que a resolução do sistema linear requer $(n+1)^3/3 + (n+1)^2 - (n+1)/3$ multiplicações/adições ($\mathcal{O}(n^3)$ operações). Quanto à estabilidade, é possível provar que a matriz de Vandermonde do sistema é muito mal condicionada: prova-se que o número de condição da matriz é tanto maior quanto maior for a sua ordem (dada pelo grau do polinómio). Na prática verifica-se que este método não permite ir além de valores de n da ordem da dezena quando se trabalha em aritmética com 6 ou 7 decimais.

2.2.1 Fórmula de Lagrange

Uma forma mais eficiente de determinar o polinómio interpolador de Lagrange de uma função f nos pontos da partição (2.1) é obtida à custa da chamada fórmula de Lagrange. Para a definir consideremos o conjunto $\{\ell_i\}_{i=0}^n$ dos **polinómios de Lagrange** de grau inferior ou igual a n dados por

$$\ell_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}. \quad (2.5)$$

Uma vez que $\{\ell_i\}_{i=0}^n$ constitui uma base de $\mathcal{P}_n([a, b])$ (Exercício 2.2.5), dado um polinómio $P_n \in \mathcal{P}_n([a, b])$, esse polinómio pode ser

escrito na forma

$$P_n(x) = \sum_{i=0}^n c_i \ell_i(x).$$

Para que este polinómio verifique as condições de interpolação (2.3) os coeficientes c_i terão que coincidir com os valores nodais $f(x_i)$, $i = 0, \dots, n$, pois $\ell_i(x_j) = \delta_{i,j}$, onde $\delta_{i,j}$ representa o símbolo de Kronecker

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Assim sendo, o polinómio interpolador de f nos pontos da partição dada pode ser escrito na forma

$$P_n(x) = \sum_{i=0}^n f(x_i) \ell_i(x). \quad (2.6)$$

As expressões (2.6) e (2.5) definem a fórmula de Lagrange para calcular o polinómio interpolador de f nos pontos x_0, x_1, \dots, x_n . O polinómio interpolador calculado é muitas vezes chamado **polinómio interpolador de Lagrange** de f nos pontos da partição (2.1).

Notemos que a fórmula de Lagrange pode ser escrita na forma

$$P_n(x) = \sum_{i=0}^n f(x_i) \frac{w(x)}{(x - x_i)w'(x_i)}, \quad (2.7)$$

sendo

$$w(x) = \prod_{j=0}^n (x - x_j). \quad (2.8)$$

De facto, atendendo a (2.8) temos que

$$w'(x) = \sum_{i=0}^n \prod_{\substack{j=0 \\ j \neq i}}^n (x - x_j) \Rightarrow w'(x_i) = \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j),$$

e como tal

$$\ell_i(x) = \frac{w(x)}{(x - x_i)w'(x_i)},$$

o que prova o pretendido.

Na prática, nunca se deve reduzir o polinómio interpolador de Lagrange à sua forma canónica uma vez que esta redução pode implicar perdas de precisão.

Para determinar o esforço computacional necessário à obtenção do polinómio interpolador pela fórmula de Lagrange, note-se que, supondo as constantes

$$F_i = \frac{f(x_i)}{w'(x_i)}, \quad i = 0, \dots, n,$$

calculadas *a priori*, o cálculo do valor do polinómio interpolador num determinado ponto pode ser dado por

$$P_n(x) = w(x) \left[\frac{F_0}{x - x_0} + \dots + \frac{F_n}{x - x_n} \right].$$

Este cálculo requer $n(n + 1)$ multiplicações e $n(n + 2)$ adições, isto é, $\mathcal{O}(n^2)$ operações, o que torna a fórmula de Lagrange muito mais eficiente que o processo matricial.

A fórmula de Lagrange possui, no entanto, o inconveniente de obrigar a refazer os cálculos dos polinómios (2.5) sempre que ocorra uma alteração nos pontos de suporte. Na prática esta situação acontece com frequência, por exemplo, quando pretendemos passar de P_n a P_{n+1} , pela adição de mais um ponto x_{n+1} ao suporte de interpolação, a fim de estudar o comportamento do erro.

2.2.2 Fórmula de Newton

Consideremos as seguintes funções

$$\phi_0(x) = 1, \quad \phi_i = \prod_{j=0}^{i-1} (x - x_j), \quad i = 1, \dots, n.$$

Atendendo a que o conjunto $\{\phi_i\}_{i=0}^n$ constitui uma base $\mathcal{P}_n([a, b])$ (prove), existem constantes $c_i, i = 0, \dots, n$, tais que o polinómio interpolador de Lagrange é dado por

$$P_n(x) = \sum_{i=0}^n c_i \phi_i(x). \quad (2.9)$$

Para determinar c_0 note-se que, se $P_n(x)$ poder ser escrito na forma (2.9), temos que

$$c_0 = P_n(x_0) = f(x_0).$$

De forma similar temos que c_1 pode ser determinado calculando P_n no ponto x_1 . Assim

$$f(x_0) + c_1(x - x_1) = P_n(x_1) = f(x_1) \Rightarrow c_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

Denotando por $f[x_0, x_1]$ a diferença dividida de primeira ordem $\frac{f(x_1) - f(x_0)}{x_1 - x_0}$ e prosseguindo de forma idêntica deduzimos que

$$c_2 = \frac{f(x_2) - f(x_0) - f[x_0, x_1](x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)}$$

que denotamos por $f[x_0, x_1, x_2]$. Podemos deste modo obter um processo recursivo para a determinação dos coeficientes do polinómio se atendermos à seguinte definição.

Definição 2.4 (Diferenças divididas) *Seja f uma função definida nos pontos da partição (2.1) do intervalo $[a, b] \subseteq \mathbb{R}$. A*

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$

chama-se diferença dividida de primeira ordem de f relativamente aos argumentos x_i e x_{i+1} . As diferenças divididas de ordem superior definem-se recursivamente. Assim, define-se diferença dividida de ordem k relativamente aos argumentos $x_i, x_{i+1}, \dots, x_{i+k}$, com $i + k < n$, por

$$f[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{f[x_{i+1}, x_{i+2}, \dots, x_{i+k}] - f[x_i, x_{i+1}, \dots, x_{i+k-1}]}{x_{i+k} - x_i}.$$

Usando a definição anterior pode demonstrar-se que

$$c_i = f[x_0, \dots, x_i], \quad i = 1, \dots, n.$$

Substituindo este valor na expressão (2.9) que define $P_n(x)$ obtemos

$$\begin{aligned} P_n(x) = & f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\ & + \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \cdots (x - x_{n-1}), \end{aligned} \tag{2.10}$$

conhecida por fórmula interpoladora de Newton das diferenças divididas. Abusivamente é usual designar por polinómio interpolador de Newton o polinómio calculado por (2.10).

Um resultado importante respeitante às diferenças divididas é o seguinte.

Teorema 2.5 *As diferenças divididas são invariantes para qualquer permutação dos índices de suporte.*

Demonstração: Com efeito, tem-se que

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{f(x_i) - f(x_{i+1})}{x_i - x_{i+1}} = f[x_{i+1}, x_i].$$

Por indução conclui-se facilmente (exercício) que o mesmo acontece para as diferenças divididas de qualquer ordem. \square

A demonstração do teorema anterior poderia ter sido feita atendendo ao seguinte exercício que se demonstra por indução.

Exercício 2.2.1 Seja P_n o polinómio interpolador de $f \in C^{n+1}([a, b])$ de grau inferior ou igual a n nos pontos da partição (2.1) do intervalo $[a, b]$ e w o polinómio nodal dado em (2.8). Mostre que se verifica a igualdade

$$f[x_0, x_1, \dots, x_n] = \sum_{i=0}^n \frac{f(x_i)}{w'(x_i)}.$$

Apresentámos três processos distintos para a construção do polinómio interpolador de Lagrange de grau n quando são conhecidos $n + 1$ valores de uma dada função. Dos processos apresentados aquele que se mostra menos eficiente é o método matricial pois não tem uma forma explícita de determinar os coeficientes do polinómio interpolador. Mais ainda, a determinação destes coeficientes é feita recorrendo à resolução de um sistema de equações lineares em que a matriz deste sistema pode ser mal condicionada.

O processo de construção mais eficiente é o método das diferenças divididas. O cálculo do polinómio interpolador usando (2.10) na forma encaixada (ver Exercício 2.2.2), supondo calculados os coeficientes $f(x_0)$, $f[x_0, x_1]$, \dots , $f[x_0, x_1, \dots, x_n]$, requer apenas $2n$ adições/subtracções e n multiplicações/divisões, isto é, $\mathcal{O}(n)$ operações.

2.2.3 Erro de interpolação

Por definição, o polinómio interpolador coincide com a função num dado conjunto de pontos de suporte. Interessa-nos saber, no entanto,

se para os outros pontos do domínio da função, o polinómio interpolador constitui uma boa ou uma má aproximação para a função. Nesse sentido temos o seguinte teorema.

Teorema 2.6 *Seja P_n o polinómio de grau menor ou igual a n interpolador da função f nos pontos da partição (2.1) do intervalo $[a, b]$. Se $f \in C^{n+1}([a, b])$ então, para cada $x \in [a, b]$, existe $\xi = \xi(x) \in]a, b[$ tal que*

$$e_n(x) := f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} w(x), \quad (2.11)$$

onde w é a função nodal dada em (2.8).

Demonstração: Se $x = x_i$, para algum i o resultado é, obviamente, válido. Se $x \neq x_i$, $i = 0, 1, \dots, n$, defina-se a função auxiliar

$$F(t) = f(t) - P_n(t) - \frac{w(t)}{w(x)}(f(x) - P_n(x)).$$

Ora, como $F(t) = 0$ possui $n+2$ raízes distintas em $[a, b]$, uma vez que $F(x_i) = 0$, $i = 0, 1, \dots, n$, e $F(x) = 0$, por aplicação do Teorema de Rolle, conclui-se que $F'(t) = 0$ possui pelo menos $n+1$ raízes distintas em $]a, b[$, $F''(t) = 0$ possui pelo menos n raízes distintas em $]a, b[$ e, sucessivamente, $F^{(n+1)}(t) = 0$ possui pelo menos uma raiz em $]a, b[$. Seja $t = \xi$ essa raiz. Uma vez que

$$F^{(n+1)}(t) = f^{(n+1)}(t) - \frac{(n+1)!}{w(x)}(f(x) - P_n(x)),$$

substituindo t por ξ , obtem-se (2.11). \square

Note-se a semelhança existente entre a fórmula do erro na interpolação e na fórmula de Taylor. A diferença está que, enquanto a primeira usa informação em vários pontos distintos, a segunda recorre apenas a um único ponto.

Observação 2.7 *Na prática a expressão (2.11) não é muito útil uma vez que o valor do ponto intremédio ξ não é conhecido. Vamos seguidamente considerar algumas formas alternativas de expressar o erro.*

1. *Pelo teorema anterior podemos determinar um majorante para o erro cometido ao substituir f pelo seu polinómio interpolador de*

Lagrange de grau n , P_n . De facto, de (2.11) sai que

$$\|f - P_n\|_\infty \leq \frac{\|f^{(n+1)}\|_\infty}{(n+1)!} \|w\|_\infty. \quad (2.12)$$

2. Atendendo a que $x, x_j \in [a, b]$, temos $|x - x_j| \leq (b - a)$ e portanto

$$\|f - P_n\|_\infty \leq \frac{\|f^{(n+1)}\|_\infty}{(n+1)!} (b - a)^{n+1}.$$

3. Atendendo à observação anterior, podemos determinar qual o grau do polinómio interpolador de Lagrange que satisfaz alguma restrição de proximidade. De facto, se considerarmos $M_{n+1} > 0$ tal que $\|f^{(n+1)}\|_\infty < M_{n+1}$ temos que, fixado $\epsilon > 0$, poderemos determinar o valor de n tal que

$$\frac{M_{n+1}}{(n+1)!} (b - a)^{n+1} < \epsilon.$$

Determinado o valor de n , e considerando pontos igualmente distânciados em $[a, b]$, construímos o polinómio interpolador.

No resultado seguinte é estabelecida uma estimativa mais precisa para o erro cometido ao aproximar uma função pelo seu polinómio interpolador.

Teorema 2.8 *Seja f uma função definida no intervalo $[a, b]$ onde se considera a partição (2.1). Seja P_n o polinómio interpolador de Lagrange para a função f e seja*

$$h = \max_{i=1, \dots, n} |x_i - x_{i-1}|.$$

Então

$$\|f - P_n\|_\infty \leq \|f^{(n+1)}\|_\infty \frac{h^{n+1}}{4(n+1)}. \quad (2.13)$$

Demonstração: Atendendo às observações anteriores, para concluir a estimativa apresentada provemos que

$$\|w\|_\infty \leq \frac{h^{n+1} n!}{4},$$

com w a função nodal (2.8). Vamos efectuar a demonstração por indução.

Para $n = 1$ temos que $w(x) = (x - x_0)(x - x_1)$. Assim,

$$w'(x) = 0 \Rightarrow x = \frac{x_0 + x_1}{2}.$$

Como tal,

$$\|w\|_\infty = \max \left\{ |w(a)|, \left| w \left(\frac{x_0 + x_1}{2} \right) \right|, |w(b)| \right\} = \left| w \left(\frac{x_0 + x_1}{2} \right) \right| \leq \frac{h^2}{4}.$$

Suponhamos que (2.13) se verifica para n e provemos a sua veracidade para $n + 1$, isto é, que

$$\max_{x \in [a, b]} \prod_{j=0}^{n+1} |x - x_j| \leq \frac{h^{n+2}(n+1)!}{4},$$

com $a = x_0$ e $x_{n+1} = b$. Dado $x \in [a, b]$ temos que $x \in [a, x_n]$ ou $x \in [x_n, b]$. Consideremos, sem perda de generalidade, a primeira hipótese. Então

$$\max_{x \in [a, b]} \prod_{j=0}^{n+1} |x - x_j| = \max_{x \in [a, b]} \prod_{j=0}^n |x - x_j| |x - b| \leq \frac{h^{n+1}n!}{4} (n+1)h = \frac{h^{n+2}(n+1)!}{4},$$

o que prova o pretendido. \square

Na determinação de uma estimativa para o erro cometido ao aproximar uma função pelo seu polinómio interpolador de Lagrange é usada a derivada da função que estamos a aproximar. Em geral esta função é apenas conhecida pontualmente. Vejamos seguidamente como podemos aproximar as derivadas que figuram na estimativa do erro utilizando apenas o valor da função nos pontos da partição do intervalo.

Teorema 2.9 *Seja f uma função admitindo derivadas até à ordem $n+1$ contínuas em $[a, b]$. Neste intervalo consideremos a partição (2.1) e seja P_n o polinómio interpolador de Lagrange para a função f na partição dada. Então para $x \in [a, b]$ existe $\xi = \xi(x) \in]a, b[$ tal que*

$$f[x_0, x_1, \dots, x_n]n! = f^{(n)}(\xi).$$

Demonstração: Consideremos o polinómio interpolador de Lagrange (2.10). Temos que, para $x \in [a, b]$, o erro de interpolação é dado por

$$e(x) = f(x) - P_n(x) = f(x) - \left(f(x_0) + \sum_{i=1}^n f[x_0, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j) \right).$$

A função erro tem $n + 1$ zeros em $[a, b]$, e portanto, pelo Teorema de Rolle, a função derivada de ordem n tem pelo menos um zero. Logo existe $\xi = \xi(x) \in]a, b[$ tal que $e^{(n)}(\xi) = 0$, isto é,

$$f^{(n)}(\xi) - f[x_0, \dots, x_n]n! = 0,$$

o que prova o pretendido. \square

Observação 2.10 *Pelo teorema anterior, para estimar a derivada de ordem $n + 1$ que figura em (2.11), podemos utilizar diferença dividida de ordem $n + 2$. De facto, se $x \in [a, b]$, existe $\xi = \xi(x) \in [a, b]$ tal que*

$$f[x_0, \dots, x_i, x, x_{i+1}, \dots, x_n](n + 1)! = f^{(n+1)}(\xi)$$

e portanto

$$f(x) - P_n(x) \simeq f[x_0, \dots, x_i, x, x_{i+1}, \dots, x_n]w(x).$$

Vamos agora analisar brevemente o problema da convergência do processo de interpolação. Para isso, consideremos uma função f definida num intervalo $[a, b]$ onde está definida uma partição uniforme (com $n + 1$ pontos) e seja P_n o seu polinómio interpolador de Lagrange. Provámos que

$$\|f - P_n\|_\infty \leq \|f^{(n+1)}\|_\infty \frac{h^{n+1}}{4(n + 1)},$$

em que $h = \frac{b-a}{n}$. Se existir uma constante positiva M tal que

$$\|f^{(n+1)}\|_\infty \leq M, \quad \forall n \in \mathbb{N},$$

então

$$\|f - P_n\|_\infty \leq M \frac{(b - a)^{n+1}}{4(n + 1)n^{n+1}}.$$

Atendendo a que

$$\lim_{n \rightarrow +\infty} \frac{(b - a)^{n+1}}{4(n + 1)n^{n+1}} = 0,$$

concluimos que

$$\|f - P_n\|_\infty \rightarrow 0, \quad n \rightarrow +\infty.$$

Neste caso, o processo de interpolação é convergente, isto é, o aumento do grau do polinómio implica um aumento de precisão. Existem, no entanto, funções para as quais não podemos concluir que um aumento do grau do polinómio provoque um aumento da proximidade do polinómio interpolador com a função interpolada. Isso acontece quando não é possível encontrar um majorante para as derivadas da função. Um exemplo que ilustra esta situação foi considerado por Carl Runge em 1901 e é o apresentado no Exercício 2.2.14. Esse exercício ilustra que, para a interpolação polinomial, podemos ter, em simultâneo, situações de convergência e de divergência.

2.2.4 Zeros dos polinómios de Chebyshev

Uma questão interessante consiste em saber como diminuir o erro de interpolação sem aumentar o número de pontos de suporte. A fórmula (2.12) mostra que o erro de interpolação depende tanto de $\|f^{(n+1)}\|_\infty$ como de $\|w\|_\infty$ (que depende da escolha dos pontos de interpolação). A questão interessante está em saber, para um dado n , qual a escolha dos pontos de interpolação que minimiza $\|w\|_\infty$. A resposta pode ser dada à custa dos chamados **polinómios de Chebyshev**.

Para $n = 0, 1, 2, \dots$ e $x \in [-1, 1]$ os polinómios de Chebyshev de grau n são definidos pela relação

$$T_n(x) = \cos(n \arccos x).$$

Uma forma simples de provar que T_n é, de facto, um polinómio, é atendendo à fórmula de recorrência (ver Exercício 2.2.15)

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n = 1, 2, \dots$$

Observação 2.11 *Da definição de polinómio de Chebyshev resulta imediatamente que*
 $|T_n(x)| \leq 1$, $n = 0, 1, 2, \dots$. Assim sendo, como $T_n(1) = 1$, temos que, em $[-1, 1]$, $\|T_n\|_\infty = 1$. Além disso, atendendo ao Exercício 2.2.16, os zeros dos polinómios de Chebyshev estão todos no intervalo $[-1, 1]$.

É fácil provar que o coeficiente do termo de maior grau de T_n é $a_n = 2^{n-1}$. Assim sendo, o polinómio $\tilde{T}_n := 2^{1-n}T_n$ é mónico, isto é, o seu coeficiente do termo de maior grau é igual à unidade. Designemos por $\tilde{\mathcal{P}}_n([a, b])$ a classe dos polinómios mónicos de grau menor ou igual a n em $[a, b]$.

Teorema 2.12 *O polinómio \tilde{T}_n é de todos os polinómios de $\tilde{\mathcal{P}}_n([-1, 1])$ o que tem menor norma, isto é,*

$$\|\tilde{T}_n\|_\infty \leq \|\tilde{P}\|_\infty, \quad \forall \tilde{P} \in \tilde{\mathcal{P}}_n([-1, 1]).$$

Demonstração: Sabemos que $\|\tilde{T}_n\|_\infty = 2^{1-n}$. Suponhamos que existe $\tilde{P} \in \tilde{\mathcal{P}}_n([-1, 1])$ tal que $\|\tilde{P}\|_\infty < 2^{1-n}$ e seja $Q = \tilde{T}_n - \tilde{P}$. Então o grau de Q é menor ou igual a $n - 1$. Por outro lado, para os valores de x'_k dados no Exercício 2.2.16,

$$Q(x'_k) = \tilde{T}_n(x'_k) - \tilde{P}(x'_k) = (-1)^k 2^{1-n} - \tilde{P}(x'_k).$$

Assim sendo, o polinómio Q tem n zeros pois tem sinais alternados em n intervalos e é uma função contínua. Logo Q é o polinómio nulo, o que prova o resultado. \square

Observação 2.13 *Se considerarmos o intervalo $[a, b]$ em vez do intervalo $[-1, 1]$ há que efectuar a mudança de variável*

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

Este resultado permite-nos afirmar, atendendo a que w dado por (2.8) é um polinómio mónico, que $\|w\|_\infty$ é mínimo quando se consideram os pontos de suporte

$$x_i = \frac{a+b}{2} + \frac{b-a}{2} \cos \frac{(2i+1)\pi}{2n+2}, \quad i = 0, \dots, n.$$

Neste caso o erro é dado por

$$\|f - P_n\|_\infty \leq \frac{(b-a)^n}{2^{n+1}(n+1)!} \|f^{(n+1)}\|_\infty.$$

O fenómeno de interpolação também é muito sensível a erros dos dados $y_i = f(x_i)$, $i = 0, \dots, n$, e a escolha criteriosa dos pontos de

suporte pode, também neste aspecto, ser importante. Suponhamos que o cálculo do polinómio interpolador é efectuado com os valores

$$\widehat{y}_i = y_i(1 + \epsilon_i), \quad |\epsilon_i| < \epsilon.$$

Assim, os polinómios que passam por (x_i, y_i) e (x_i, \widehat{y}_i) são dados, respectivamente, por

$$P_n(x) = \sum_{i=0}^n y_i \ell_i(x)$$

e por

$$\widehat{P}_n(x) = \sum_{i=0}^n \widehat{y}_i \ell_i(x).$$

Como tal,

$$|\widehat{P}_n(x) - P_n(x)| \leq \epsilon \max_{i=0, \dots, n} |y_i| \sum_{i=0}^n |\ell_i(x)|.$$

Temos então que a função

$$\sum_{i=0}^n |\ell_i(x)|$$

descreve o factor de amplificação dos erros dos dados. O seu valor máximo

$$\Lambda_n = \max_{x \in [a, b]} \sum_{i=0}^n |\ell_i(x)|$$

é chamado **constante de Lebesgue** associada aos pontos de interpolação dados e ao intervalo $[a, b]$. Esta constante pode ser calculada numericamente.

2.2.5 Exercícios

Exercício 2.2.2 Consideremos $P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ e suponhamos que pretendemos calcular $p(\bar{x})$. Ao usar $P_n(\bar{x}) = a_n \bar{x}^n + a_{n-1} \bar{x}^{n-1} + \dots + a_1 \bar{x} + a_0$ efectuamos n adições/subtracções e $2n - 1$ multiplicações/divisões. No entanto, se considerarmos

$$P_n(x) = a_0 + x(a_1 + x(a_2 + \dots + x(a_{n-1} + a_n x))),$$

designada por forma encaixada do polinómio, ao calcular $p(\bar{x})$ só efectuamos n adições/subtracções e n multiplicações/divisões. Esta forma é a

base do chamado método (ou algoritmo) de Horner, que consiste formalmente nas seguintes operações: $p \leftarrow a_n$; $p \leftarrow p\bar{x} + a_i$, $i = n - 1(-1)0$, sendo $P_n(\bar{x}) = p$.

1. Demonstre a chamada regra de Ruffini: O valor numérico de $P_n(\bar{x})$ de um polinómio P_n em \bar{x} é igual ao resto da divisão de $P_n(x)$ por $x - \bar{x}$.
2. Dividindo $P_n(x)$ por $x - \bar{x}$ obtém-se $P_n(x) = (x - \bar{x})q_{n-1}(x) + r$, onde q_{n-1} é um polinómio de grau inferior ou igual a $n - 1$ e r uma constante. Usando o algoritmo de Horner construa um algoritmo que permita determinar r e os coeficientes de q_{n-1} (algoritmo de Ruffini).
3. Calcule o valor de $p_5(x) = x^5 + x^4 - 4x^3 - x^2 - 12$ em $\bar{x} = \pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$.
4. Quais as soluções inteiras de um polinómio de grau n de coeficientes inteiros?

Exercício 2.2.3 Conhecem-se as coordenadas de cinco pontos de uma curva plana, que representa uma região de uma peça em corte. Determine o polinómio de Lagrange de grau 4 que interpola a referida curva sabendo que os pontos de coordenadas conhecidas são: $P_1 = (1, 2)$, $P_2 = (2, 1)$, $P_3 = (3, 1)$, $P_4 = (4, 2.5)$ e $P_5 = (5, 4)$. Determine ainda valores aproximados para as ordenadas dos pontos cujas abcissas são 0, 2.5 e 6.

Exercício 2.2.4 Sejam x_0, x_1, \dots, x_n , $n+1$ pontos distintos de \mathbb{R} . Mostre que a matriz de Vandermonde que lhe está associada $V(x_0, \dots, x_n)$ verifica

$$\det V(x_0, \dots, x_n) = \prod_{0 \leq j < k \leq n} (x_j - x_k).$$

Exercício 2.2.5 Prove que $\{\ell_i\}_{i=0}^n$ constitui uma base de $\mathcal{P}_n([a, b])$.

Exercício 2.2.6 Use o polinómio interpolador de Lagrange para determinar $\log 2.45$ sabendo que

x_i	2.2	2.3	2.4	2.5	2.6
$\log x_i$	0.34242	0.36173	0.38021	0.39794	0.41497

Determine uma estimativa para o erro cometido.

Exercício 2.2.7 Na seguinte tabela são dados diferentes valores para o peso específico p da água a diferentes temperaturas t (em graus centígrados):

t	0	1	2	3
p	0.999871	0.999928	0.999969	0.999991

Usando interpolação linear, quadrática e cúbica, determine uma aproximação para p quando $t = 4^\circ C$ usando a fórmula interpoladora de Lagrange e de Newton. Compare os resultados obtidos sabendo que o valor exacto é 1.000000.

Exercício 2.2.8 Escreva a fórmula interpoladora de Newton das diferenças divididas usando o algoritmo de Horner.

Exercício 2.2.9 Pretende-se construir uma tabela para a função $f(x) = e^x$, com $x \in [0, 1]$. Considere o valor de e com 5 casas decimais correctas e uma partição com pontos igualmente distanciados. Determine o diâmetro da partição a considerar de modo que o polinómio interpolador de Lagrange permita obter uma aproximação para f com um erro inferior a 10^{-6} .

Exercício 2.2.10 Considere a função $f(x) = \ln(x + 1)$, $x \in [1, 3]$. Determine o polinómio interpolador de Lagrange que aproxima f em $[1, 3]$ com um erro inferior a 10^{-2} .

Exercício 2.2.11 Determine uma aproximação para o instante na da passagem do perigeu da Lua em Março, 1999, a partir dos valores tabelados para as zero horas de cada dia; indique também a distância (em raios médios da Terra) da Terra à Lua nesse instante.

dia	19	20	21
distância	57.071	56.955	57.059

Exercício 2.2.12 Determine uma aproximação para a declinação aparente de Vénus para o dia 8 de Maio de 1999, às 18h30m45s, por interpolação cúbica a partir das Efemérides Astronómicas (onde está tabelada para cada dia, às zero horas)

dia	7	8	9	10
δ_i	+5°51'47".55	+6°22'25".20	+6°52'54".57	+6°23'14".96

Exercício 2.2.13 O censo da população dos Estados Unidos, entre 1930 e 1980, produziu os seguintes resultados:

Ano	1930	1940	1950	1960	1970	1980
População ($\times 10^3$)	123203	131669	150697	179323	203212	226505

Use um método de diferenças finitas apropriado para estimar a população nos anos de 1920, 1965, e 2000. Sabendo que a população no ano de 1920 era de 105711×10^3 , o que pode inferir quanto à precisão das aproximações obtidas para os anos de 1965 e 2000?

Exercício 2.2.14 Considere a função (de Runge) $f(x) = 1/(1 + 25x^2)$, $x \in [-1, 1]$.

1. Verifique graficamente que

$$\|f - P_3\|_\infty \leq \|f - P_8\|_\infty,$$

em que P_3 e P_8 são, respectivamente, os polinômios de Lagrange de grau 3 e 8 interpoladores de f em partições uniformes de $[-1, 1]$.

2. Verifique numericamente que, quando se considera a função de Runge, os polinômios interpoladores P_n divergem em $0.726 \leq |x| \leq 1$.

Exercício 2.2.15 Obtenha a fórmula de recorrência

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n = 1, 2, \dots,$$

e conclua que T_n é, de facto, um polinômio.

Exercício 2.2.16 Mostre que o polinômio de Chebyshev T_n tem os seus zeros localizados nos pontos

$$x_k = \cos \frac{(2k-1)\pi}{2n}, \quad k = 1, \dots, n,$$

e os extremos localizados em

$$x'_k = \cos \frac{k\pi}{n}, \quad k = 0, \dots, n,$$

nos quais $T_n(x'_k) = (-1)^k$.

Exercício 2.2.17 Mostre numericamente que, quando se consideram pontos igualmente distanciados no intervalo $[a, b]$, se tem

$$\Lambda_{20} \simeq 3 \times 10^4, \quad \Lambda_{40} \simeq 10^{10}$$

e quando se consideram os pontos de Chebyshev

$$\Lambda_n \leq 3, \quad (n \leq 30), \quad \Lambda_n \leq 4, \quad (n \leq 100).$$

2.3 O polinómio interpolador de Lagrange segmentado

2.3.1 Caso linear

Consideremos um intervalo $[a, b]$ e uma partição Δ dada por (2.1). Denotemos por $\mathcal{P}_1^0(\Delta)$ o conjunto de todas as funções contínuas em $[a, b]$ que, quando restringidas a cada um dos intervalos $[x_{i-1}, x_i]$, $i = 1, \dots, n$, da partição, coincidem com um polinómio de grau menor ou igual a um (polinómio que, em geral, varia com i). Se $S_1 \in \mathcal{P}_1^0(\Delta)$ dizemos que S_1 é uma **função linear por segmentos** ou um **polinómio segmentado linear** (na partição Δ).

Observação 2.14 *Note-se que, em geral, nos pontos $x_i \in \Delta$ as funções $S_1 \in \mathcal{P}_1^0(\Delta)$ apresentam descontinuidades da derivada.*

Consideremos agora o problema da interpolação. Seja f uma função conhecida nos pontos da partição Δ do intervalo $[a, b]$. Pelo que foi visto na secção anterior, existe um e um só polinómio segmentado linear $S_1 \in \mathcal{P}_1^0(\Delta)$ tal que

$$S_1(x_i) = f(x_i), \quad i = 0, 1, \dots, n.$$

Nestas condições, S_1 é chamado o **polinómio interpolador (de Lagrange) segmentado linear** de f nos pontos de Δ . Temos que

$$S_1(x) = \begin{cases} S_1^{(1)}(x) & x \in [x_0, x_1] \\ S_1^{(2)}(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_1^{(i)}(x) & x \in [x_{i-1}, x_i] \\ \vdots & \vdots \\ S_1^{(n)}(x) & x \in [x_{n-1}, x_n] \end{cases},$$

onde $S_1^{(i)}$ pode ser escrita na forma seguinte (fórmula de Newton)

$$S_1^{(i)}(x) = f(x_{i-1}) + f[x_{i-1}, x_i](x - x_{i-1}),$$

ou ainda (fórmula de Lagrange)

$$S_1^{(i)}(x) = f(x_{i-1}) \frac{x - x_i}{x_{i-1} - x_i} + f(x_i) \frac{x - x_{i-1}}{x_i - x_{i-1}}.$$

O que podemos dizer quanto ao erro que se comete ao aproximar f por uma função $S_1 \in \mathcal{P}_1^0(\Delta)$? Suponhamos que $x \in [x_{i-1}, x_i]$. Então, pelo que foi visto na secção anterior, temos que, nesse intervalo,

$$\|e^{(i)}\|_{i,\infty} := \|f - S_1^{(i)}\|_{i,\infty} := \max_{x \in [x_{i-1}, x_i]} |f(x) - S_1^{(i)}| = \frac{\|f^{(2)}\|_{i,\infty}}{2} \|w^{(i)}\|_{i,\infty},$$

onde (como vimos)

$$\|w^{(i)}\|_{i,\infty} := \max_{x \in [x_{i-1}, x_i]} |(x - x_{i-1})(x - x_i)| = \frac{1}{4}(x_i - x_{i-1})^2$$

e

$$\|f^{(2)}\|_{i,\infty} = \max_{x \in [x_{i-1}, x_i]} |f^{(2)}(x)|.$$

Temos então que, sendo $h_i = x_i - x_{i-1}$,

$$\|e^{(i)}\|_{i,\infty} = \frac{\|f^{(2)}\|_{i,\infty}}{8} h_i^2.$$

Considerando agora o erro no intervalo $[a, b]$,

$$\|e\|_\infty := \|f - S_1\|_\infty,$$

temos que

$$\|e\|_\infty = \max_{i=1,\dots,n} \|e^{(i)}\|_{i,\infty} = \frac{\|f^{(2)}\|_\infty}{8} h^2,$$

com $h = \max_{i=1,\dots,n} h_i$. Este limite superior para o erro permite demonstrar que o processo de interpolação linear por segmentos é convergente. De facto, se $f^{(2)}$ é limitada, à medida que o número de pontos da partição aumenta (h diminui) o erro tende para zero, ou seja, o polinómio segmentado linear tende para a função a interpolar uniformemente em $[a, b]$.

Observação 2.15 *A interpolação linear segmentada possui vantagens em relação à interpolação (global) de Lagrange. Note-se que, se n for muito grande, o cálculo do polinómio interpolador de Lagrange (global) P_n envolve muito mais operações que o cálculo do polinómio interpolador linear segmentado S_1 . Além disso, como foi visto, o facto de n aumentar não implica que o polinómio interpolador de Lagrange P_n tenda para a função a interpolar, mesmo que essa função seja infinitamente diferenciável. A desvantagem que o processo da interpolação*

segmentada linear apresenta relativamente à interpolação de Lagrange é que o polinómio P_n é infinitamente diferenciável enquanto que S_1 pode nem sequer possuir derivadas contínuas nos pontos da partição.

Como poderemos construir uma base no espaço $\mathcal{P}_1^0(\Delta)$? Para responder a esta questão, vamos determinar as funções $\{\phi_j\}_{j=0}^n$ tais que, para todo o $S_1 \in \mathcal{P}_1^0(\Delta)$,

$$S_1(x) = \sum_{j=0}^n f(x_j)\phi_j(x).$$

As funções $\{\phi_j\}_{j=0}^n$ nas condições pretendidas terão que ser lineares por segmentos e ser tais que $\phi_j(x_i) = \delta_{i,j}$, $i, j = 0, 1, \dots, n$, por forma a que S_1 seja interpolador de f nos pontos de Δ . Assim, não é difícil de verificar que os gráficos de ϕ_j , $j = 0, \dots, n$, serão os das funções chapéu dadas na Figura 2.1. Como é evidente, estas funções constituem uma base de $\mathcal{P}_1^0(\Delta)$.

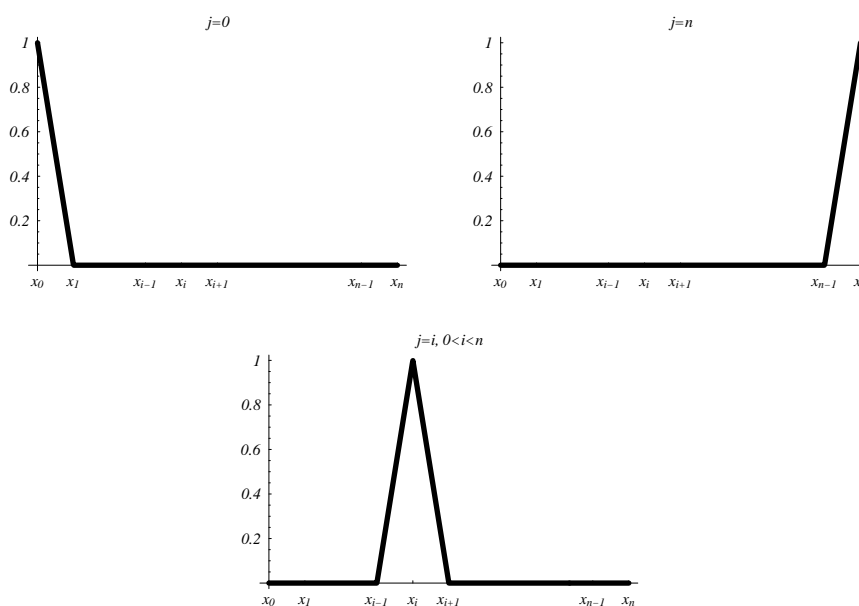


Figura 2.1: Gráficos das funções chapéu ϕ_j .

Não é difícil de ver que as funções chapéu têm as seguintes expressões analíticas (prove):

$$\phi_i(x) = \begin{cases} \frac{x-x_{i-1}}{h_i} & x \in [x_{i-1}, x_i] \\ -\frac{x-x_{i+1}}{h_{i+1}} & x \in [x_i, x_{i+1}] \\ 0 & x \in [a, x_{i-1}] \cup [x_{i+1}, b] \end{cases}, \quad i = 1, \dots, n-1,$$

e

$$\phi_0(x) = \begin{cases} \frac{x-x_1}{h_1} & x \in [a, x_1] \\ 0 & x \in [x_1, b] \end{cases}, \quad \phi_n(x) = \begin{cases} 0 & x \in [a, x_{n-1}] \\ -\frac{x-x_n}{h_n} & x \in [x_{n-1}, b] \end{cases},$$

onde $h_i = x_i - x_{i-1}$.

2.3.2 Caso quadrático

Seja $\mathcal{P}_2^0(\Delta)$ o conjunto das funções contínuas em $[a, b]$ que, quando restritas a um intervalo $[x_{i-1}, x_i]$, $i = 1, \dots, n$, da partição Δ dada por (2.1), coincidem com um polinómio do segundo grau. Se $S_2 \in \mathcal{P}_2^0(\Delta)$ dizemos que S_2 é uma função quadrática por segmentos.

Observação 2.16 *Note-se que, nos nós da partição, a função S_2 apresenta, em geral, descontinuidades na sua derivada.*

Seja f uma função conhecida nos pontos da partição Δ do intervalo $[a, b]$. Suponhamos ainda que conhecemos a função nos pontos intermédios $x_{i-1/2} \in]x_{i-1}, x_i[$, $i = 1, \dots, n$. Assim, existe um e um só $S_2 \in \mathcal{P}_2^0(\Delta)$ que interpola a função f tanto nos pontos da partição Δ como nos pontos intermédios considerados. Esse polinómio é dado por

$$S_2(x) = \begin{cases} S_2^{(1)}(x) & x \in [x_0, x_1] \\ S_2^{(2)}(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_2^{(i)}(x) & x \in [x_{i-1}, x_i] \\ \vdots & \vdots \\ S_2^{(n)}(x) & x \in [x_{n-1}, x_n] \end{cases},$$

onde $S_2^{(i)}$ pode ser escrita na forma seguinte (fórmula de Newton)

$$S_2^{(i)}(x) = f(x_{i-1}) + f[x_{i-1}, x_{i-1/2}](x-x_{i-1}) + f[x_{i-1}, x_{i-1/2}, x_i](x-x_{i-1})(x-x_{i-1/2}),$$

ou ainda (fórmula de Lagrange)

$$\begin{aligned} S_2^{(i)}(x) &= f(x_{i-1}) \frac{(x - x_{i-1/2})(x - x_i)}{(x_{i-1} - x_{i-1/2})(x_{i-1} - x_i)} \\ &\quad + f(x_{i-1/2}) \frac{(x - x_{i-1})(x - x_{i+1})}{(x_{i-1/2} - x_{i-1})(x_{i-1/2} - x_i)} \\ &\quad + f(x_i) \frac{(x - x_{i-1})(x - x_{i-1/2})}{(x_i - x_{i-1})(x_i - x_{i-1/2})}. \end{aligned}$$

Exemplo 2.17 Consideremos a função $f(x) = \sin \pi x$, $x \in [\frac{1}{2}, \frac{3}{2}]$, e, neste intervalo a partição $x_0 = \frac{1}{2}$, $x_1 = 1$ e $x_2 = \frac{3}{2}$. Determine o polinómio interpolador de Lagrange quadrático interpolador de f nos pontos da partição e nos pontos intermédios $x_{1/2} = \frac{3}{4}$ e $x_{3/2} = \frac{5}{4}$.

Resolução: É imediato que o polinómio pretendido é dado por

$$S_2(x) = \begin{cases} S_2^{(1)}(x) & x \in [\frac{1}{2}, 1] \\ S_2^{(2)}(x) & x \in [1, \frac{3}{2}], \end{cases}$$

em que

$$S_2^{(1)}(x) = \sin \pi \frac{(x - \frac{3}{4})(x - 1)}{(\frac{1}{2} - \frac{3}{4})(\frac{1}{2} - 1)} + \sin \frac{3\pi}{4} \frac{(x - \frac{1}{2})(x - 1)}{(\frac{3}{4} - 1)(\frac{3}{4} - \frac{1}{2})} + \sin \pi \frac{(x - \frac{1}{2})(x - \frac{3}{4})}{(1 - \frac{1}{2})(1 - \frac{3}{4})}, x \in [\frac{1}{2}, 1]$$

e

$$S_2^{(2)}(x) = \sin \pi \frac{(x - \frac{5}{4})(x - \frac{3}{2})}{(1 - \frac{5}{4})(1 - \frac{3}{2})} + \sin \frac{5\pi}{4} \frac{(x - 1)(x - \frac{3}{2})}{(\frac{5}{4} - 1)(\frac{5}{4} - \frac{3}{2})} + \sin \frac{3\pi}{2} \frac{(x - 1)(x - \frac{5}{4})}{(\frac{3}{2} - 1)(\frac{3}{2} - \frac{5}{4})}, x \in [1, \frac{3}{2}]$$

O que podemos dizer quanto ao erro cometido que ao aproximar f por $S_2 \in \mathcal{P}_2^0(\Delta)$? Tal como foi efectuado para o caso linear podemos provar que

$$\|e\|_\infty := \|f - S_2\|_\infty = \frac{\|f^{(3)}\|_\infty}{3!} \max_{i=1, \dots, n} \|w^{(i)}\|_{i, \infty},$$

com

$$\|w^{(i)}\|_{i, \infty} := \max_{x \in [x_{i-1}, x_i]} |(x - x_{i-1})(x - x_{i-1/2})(x - x_i)|.$$

Como é que poderemos construir uma base para o espaço $\mathcal{P}_2^0(\Delta)$? Vamos responder a esta questão determinando as funções $\{\phi_j\}_{j=0}^n$ e $\{\phi_{j-1/2}\}_{j=1}^n$ tais que, para todo o $S_2 \in \mathcal{P}_2^0(\Delta)$,

$$S_2(x) = f(x_0)\phi_0(x) + \sum_{j=1}^n (f(x_{j-1/2})\phi_{j-1/2}(x) + f(x_j)\phi_j(x)).$$

Consideremos $\phi_j, j = 0, \dots, n$ como sendo a função quadrática por segmentos que toma o valor 1 em $x = x_j$ e 0 nos restantes nós da partição bem como nos pontos intermédios. Para $j = 1, \dots, n$, consideremos $\phi_{j-1/2}$ como sendo a função quadrática por segmentos que toma o valor 1 em $x = x_{j-1/2}$ e se anula nos restantes pontos intermédios bem como nos pontos da partição. Assim, não é difícil de verificar que os gráficos de $\phi_j, j = 0, \dots, n$, e $\phi_{j-1/2}, j = 1, \dots, n$, são os das chamadas funções bigode, dadas nas Figuras 2.2 e 2.3, respectivamente. Como $\mathcal{P}_2^0(\Delta)$ é um espaço vectorial de dimensão $2n + 1$ (prove) temos que estas funções constituem uma base de $\mathcal{P}_2^0(\Delta)$.

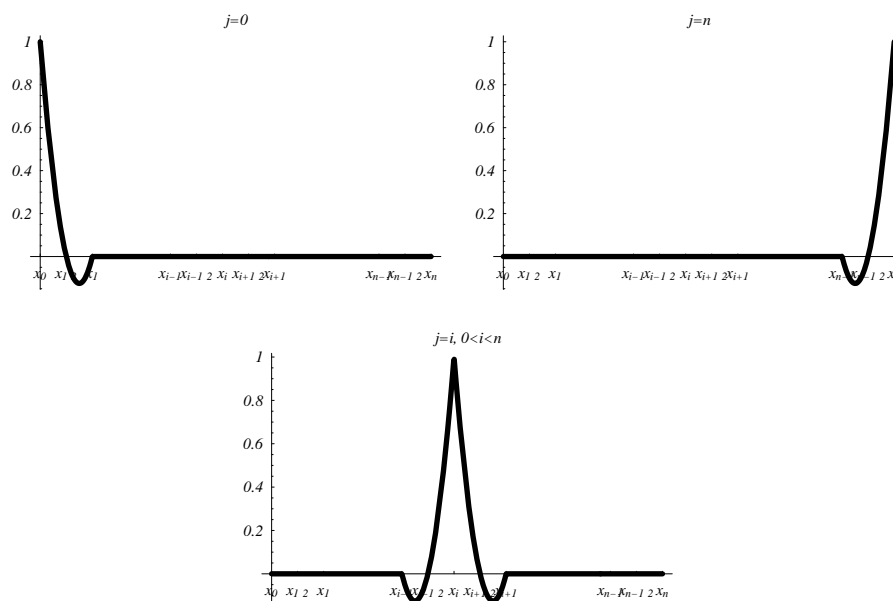


Figura 2.2: Gráficos das funções bigode ϕ_j .

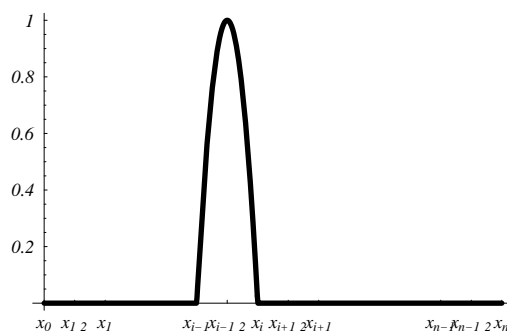


Figura 2.3: Gráficos das funções bigode $\phi_{j-1/2}$.

Tal como para o caso das funções chapéu, prova-se facilmente que as funções bigode têm as seguintes expressões analíticas:

$$\phi_i(x) = \begin{cases} \frac{(x-x_{i-1})(x-x_{i-1/2})}{(x_i-x_{i-1})(x_i-x_{i-1/2})} & x \in [x_{i-1}, x_i] \\ -\frac{(x-x_{i-1/2})(x-x_{i+1})}{(x_i-x_{i-1/2})(x_i-x_{i+1})} & x \in [x_i, x_{i+1}] \\ 0 & x \in [a, x_{i-1}] \cup [x_{i+1}, b] \end{cases}, \quad i = 1, \dots, n-1,$$

$$\phi_0(x) = \begin{cases} \frac{(x-x_{1/2})(x-x_1)}{(x_0-x_{1/2})(x_0-x_1)} & x \in [a, x_1] \\ 0 & x \in [x_1, b] \end{cases},$$

$$\phi_n(x) = \begin{cases} 0 & x \in [a, x_{n-1}] \\ \frac{(x-x_{n-1})(x-x_{n-1/2})}{(x_n-x_{n-1})(x_n-x_{n-1/2})} & x \in [x_{n-1}, b] \end{cases},$$

e

$$\phi_{i-1/2}(x) = \begin{cases} \frac{(x-x_{i-1})(x-x_i)}{(x_{i-1/2}-x_{i-1})(x_{i-1/2}-x_i)} & x \in [x_{i-1}, x_i] \\ 0 & x \in [a, x_{i-1}] \cup [x_{i+1}, b] \end{cases}.$$

2.3.3 Caso geral

Consideremos, de novo, a função f definida num intervalo $[a, b]$. Neste intervalo consideremos a partição Δ dada em (2.1) e ainda os pontos intermédios

$$x_{i-(m-1)/m} < x_{i-(m-2)/m} < \dots < x_{i-2/m} < x_{i-1/m},$$

pertencentes ao intervalo $]x_{i-1}, x_i[$, $i = 1, \dots, n$. Pelo que foi visto anteriormente, podemos concluir que existe um e um só $S_m \in \mathcal{P}_m^0(\Delta)$ – conjunto das funções contínuas em $[a, b]$ que coincidem com um polinómio de grau m quando restritas a um intervalo $]x_{i-1}, x_i[$, $i = 1, \dots, n$, da partição Δ – tal que

$$S_m(x_i) = f(x_i), \quad i = 0, \dots, n, \quad (2.14)$$

e

$$S_m(x_{i-(m-j)/m}) = f(x_{i-(m-j)/m}), \quad i = 1, \dots, n, \quad j = 1, \dots, m-1, \quad (2.15)$$

dado na forma

$$S_m(x) = \begin{cases} S_m^{(1)}(x) & x \in [x_0, x_1] \\ S_m^{(2)}(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_m^{(i)}(x) & x \in [x_{i-1}, x_i] \\ \vdots & \vdots \\ S_m^{(n)}(x) & x \in [x_{n-1}, x_n] \end{cases},$$

onde $S_m^{(i)}$ pode ser escrita na forma seguinte (fórmula de Newton)

$$\begin{aligned} S_m^{(i)}(x) = & f(x_{i-1}) + f[x_{i-1}, x_{i-(m-1)/m}](x - x_{i-1}) \\ & + \dots + f[x_{i-1}, x_{i-(m-1)/m}, \dots, x_i](x - x_{i-1})(x - x_{i-(m-1)/m}) \dots (x - x_{i-2}) \end{aligned}$$

O que poderemos dizer relativamente ao erro cometido ao aproximar a função f pelo polinómio interpolador de Lagrange segmentado $S_m(x)$? Atendendo ao que foi feito para os casos anteriores, é fácil concluir o seguinte resultado.

Teorema 2.18 *Seja f uma função com derivadas contínuas até à ordem $m+1$ em $]a, b[$, com $m \in \mathbb{N}$. Consideremos a partição Δ dada em (2.1) e ainda os pontos $x_{i-(m-j)/j}$, $j = 1, \dots, m-1$, em cada intervalo $]x_{i-1}, x_i[$, $i = 1, \dots, n$. Nestas condições, se $S_m \in \mathcal{P}_m^0(\Delta)$ verifica (2.14)–(2.15) então*

$$\|f - S_m\|_\infty \leq \frac{\|f^{(m+1)}\|_\infty}{(m+1)!} \max_{i=1, \dots, n} \|w^{(i)}\|_{i, \infty},$$

com

$$\|w^{(i)}\|_{i,\infty} = \max_{x \in [x_{i-1}, x_i]} \prod_{j=0}^m |x - x_{i-(m-j)/m}|.$$

2.3.4 Exercícios

Exercício 2.3.1 Mostre que a transformação linear $S_1 \rightarrow (S_1(x_0), \dots, S_1(x_n))$ que aplica $\mathcal{P}_1^0(\Delta)$ em \mathbb{R}^{n+1} é bijetiva e conclua que $\mathcal{P}_1^0(\Delta)$ é um espaço vectorial de dimensão $n + 1$.

Exercício 2.3.2 Determine o polinómio interpolador de Lagrange segmentado linear para a função f tal que

x_i	0	0.25	0.5	0.75	1
$f(x_i)$	1	1.25	1	1.5	1

Exercício 2.3.3 Calcule o polinómio interpolador de Lagrange segmentado quadrático para a função do exercício anterior.

Exercício 2.3.4 Determine o polinómio interpolador de Lagrange segmentado linear para uma função f supondo que esta é conhecida nos pontos

$$a = x_0 < x_1 < x_2 = b.$$

Exercício 2.3.5 Particularize o resultado do exercício anterior para a função $f(x) = \cos x$, $x \in [0, \pi]$ e aproxime $\cos(20^\circ)$, indicando um majorante para o erro cometido.

Exercício 2.3.6 Determine o polinómio interpolador de Lagrange segmentado de grau 2 para uma função f que se conhece nos pontos

$$a = x_0 < x_{1/2} < x_1 < x_{3/2} < x_2 < x_{5/2} < x_3 = b.$$

Exercício 2.3.7 Particularize o exercício anterior para

$$f(x) = \frac{1}{1+x^2}, \quad x \in [-1, 1].$$

Exercício 2.3.8 Determine o polinómio interpolador de Lagrange segmentado cúbico para uma função f supondo que esta é conhecida nos pontos

$$a = x_0 < x_{1/3} < x_{2/3} < x_1 < x_{4/3} < x_{5/3} < x_2 = b.$$

Exercício 2.3.9 Particularize o resultado do exercício anterior para a função $f(x) = \sin x$, $x \in [0, 2\pi]$, indicando um majorante para o erro cometido.

Exercício 2.3.10 Prove que se

$$x_{i-1/2} = \frac{x_{i-1} + x_i}{2}$$

o limite superior para o erro cometido na interpolação quadrática segmentada é dado por

$$\|f - S_2\|_\infty \leq \frac{\|f^{(3)}\|_\infty}{216} \sqrt{3} h^3,$$

com $h = |x_i - x_{i-1}|$, $i = 1, \dots, n$.

Exercício 2.3.11 Considere a seguinte tabela de valores da gravidade específica do ácido fosfórico como função da percentagem de H_3PO_4 .

$H_3PO_4(\%)$	Grav. esp.	$H_3PO_4(\%)$	Grav. esp.
0	1.0000	35	1.216
1	1.0038	40	1.254
2	1.0092	45	1.293
4	1.0200	50	1.335
6	1.0309	55	1.379
8	1.0420	60	1.426
10	1.0532	65	1.475
12	1.0647	70	1.526
14	1.0764	75	1.579
16	1.0884	80	1.633
18	1.1008	85	1.689
20	1.1134	90	1.746
22	1.1263	92	1.770
24	1.1395	94	1.794
26	1.1529	96	1.819
28	1.1665	98	1.844
30	1.1805	100	1.870

1. Aproxime os dados usando interpolação polinomial. Comente os resultados obtidos.
2. Use a função determinada por interpolação para tabelar a gravidade específica obtida para percentagens de 0, 5, 10, \dots , 100 de H_3PO_4 .

Exercício 2.3.12 Durante a sedimentação da reação de saponificação entre quantidades equimolares de hidróxido de sódio e acetato de etilo, a concentração c (g mole/litro) de cada reagente varia com o tempo t (min) de acordo com a equação

$$\frac{1}{c} = \frac{1}{c_0} + kt,$$

onde c_0 é a concentração inicial e k (litro/g mole min) é a constante de reacção. Foram obtidos os seguintes resultados em laboratório à temperatura de $77^\circ F$:

$1/c$	24.7	32.4	38.4	45.0	52.3	65.6	87.6	102	154	192
t	1	2	3	4	5	7	10	12	20	25

1. Obtenha uma estimativa para a concentração inicial.
2. Obtenha uma estimativa para a concentração ao fim de 15 minutos e compare-a com a solução obtida em laboratório (ao fim de 15 minutos obteve-se $1/c = 135$).

2.4 Interpolação de Hermite

O objectivo da interpolação de Hermite é o de representar uma função f por um polinómio que seja interpolador de f em alguns pontos do seu domínio e que a sua derivada seja interpolador da derivada de f nesses mesmos pontos. Isto é, supondo que f é diferenciável, vamos procurar um polinómio H tal que

$$\begin{aligned} f(x_i) &= H(x_i) \\ f'(x_i) &= H'(x_i) \end{aligned}, \quad i = 0, 1, \dots, n. \quad (2.16)$$

Quando tal situação acontece dizemos que f e H são funções que 2-osculam (osculam 2 vezes) os pontos x_i , $i = 0, 1, \dots, n$, ou que H é um polinómio 2-osculador de f nos pontos x_i , $i = 0, 1, \dots, n$.

2.4.1 Existência e unicidade

O próximo teorema estabelece a existência e unicidade do polinómio de grau inferior ou igual a $2n + 1$ que verifica (2.16). Além disso, indica-nos um processo que permite a sua determinação.

Teorema 2.19 *Seja $f \in C^{2n+2}([a, b])$ e x_0, x_1, \dots, x_n pontos distintos em $[a, b]$. Existe um e um só polinómio H_{2n+1} de grau menor ou igual a $2n + 1$ que verifica (2.16).*

Demonstração: Atendendo às condições impostas, o polinómio terá que ser de grau inferior ou igual a $2n + 1$. Para provar a sua existência vamos considerar as funções

$$h_i(x) = [1 - 2\ell'_i(x_i)(x - x_i)]\ell_i(x)^2 \quad \text{e} \quad \bar{h}_i(x) = (x - x_i)\ell_i(x)^2, \quad i = 0, \dots, n,$$

com ℓ_i , $i = 0, \dots, n$, os polinómios de Lagrange (2.6). Como se pode verificar facilmente

$$h_i(x_j) = \delta_{i,j}, \quad h'_i(x_j) = 0, \quad i, j = 0, \dots, n,$$

e

$$\bar{h}_i(x_j) = 0, \quad \bar{h}'_i(x_j) = \delta_{i,j}, \quad i, j = 0, \dots, n.$$

Assim, o polinómio

$$H_{2n+1}(x) = \sum_{i=0}^n [f(x_i)h_i(x) + f'(x_i)\bar{h}_i(x)]$$

tem grau inferior ou igual a $2n + 1$ e verifica (2.16).

Falta apenas provar a unicidade. Seja Q_{2n+1} outro polinómio de grau inferior ou igual a $2n + 1$ que verifica (2.16) e

$$R_{2n+1}(x) = H_{2n+1}(x) - Q_{2n+1}(x).$$

Como $R_{2n+1}(x_i) = R'_{2n+1}(x_i) = 0$, para $i = 0, \dots, n$, temos que este polinómio de grau inferior ou igual a $2n + 1$ tem $2n + 2$ zeros o que implica que terá que ser o polinómio nulo. Assim sendo, provámos a unicidade pretendida. \square

O único polinómio de grau menor ou igual a $2n + 1$ que verifica as condições (2.16) é também chamado **polinómio interpolador de Hermite** de f nos pontos x_0, x_1, \dots, x_n .

Observação 2.20 *Note-se que, tal como na interpolação de Lagrange, se m for o número de condições impostas para a determinação do polinómio interpolador, o seu grau é $m - 1$.*

A obtenção do polinómio interpolador de Hermite pode ser feita de várias maneiras. Vamos apresenta-la neste curso numa forma que generaliza o polinómio interpolador de Newton das diferenças divididas.

Consideremos a mudança de variável $z_0 = x_0, z_1 = x_0, z_2 = x_1, z_3 = x_1, \dots, z_{2n} = x_n, z_{2n+1} = x_n$. Uma vez que

$$z_{2i} = z_{2i+1} = x_i, \quad i = 0, \dots, n,$$

não podemos definir as diferenças divididas

$$f[z_{2i}, z_{2i+1}] = f[x_i, x_i].$$

No entanto, atendendo a que

$$\lim_{x \rightarrow x_i} f[x, x_i] = \lim_{x \rightarrow x_i} \frac{f(x) - f(x_i)}{x - x_i} = f'(x_i),$$

podemos definir as diferenças divididas generalizadas para pontos não distintos na forma

$$f[x_i, x_i] := f'(x_i).$$

Pelo Teorema do Valor Médio de Lagrange generalizado podemos ainda definir

$$f[\underbrace{x_i, x_i, \dots, x_i}_{r+1 \text{ vezes}}] = \frac{f^{(r)}(x_i)}{r!}. \quad (2.17)$$

Com esta notação pode verificar-se facilmente que o polinómio interpolador de Hermite de grau $2n + 1$ nos pontos da partição (2.1) é dado por (verifique para $n = 1$)

$$\begin{aligned} H_{2n+1}(x) &= f(z_0) + \sum_{i=1}^{2n+1} f[z_0, z_1, \dots, z_i] \prod_{j=0}^{i-1} (x - z_j) \\ &= f(x_0) + f'(x_0)(x - x_0) \\ &\quad + f[x_0, x_0, x_1](x - x_0)^2 + f[x_0, x_0, x_1, x_1](x - x_0)^2(x - x_1) \\ &\quad + \dots + f[x_0, x_0, \dots, x_n, x_n](x - x_0)^2(x - x_1)^2 \dots (x - x_{n-1})^2(x - x_n) \end{aligned}$$

O polinómio interpolador de Hermite pode assim ser determinado recorrendo à tabela das diferenças divididas generalizadas, tabela essa onde cada ponto aparece repetido duas vezes.

Exemplo 2.21 Determine o polinómio interpolador de Hermite de grau mínimo para a função $f(x) = \sin x$ em $[0, \frac{\pi}{2}]$.

Resolução: Temos

x_i	$f(x_i)$	$f[\cdot, \cdot]$	$f[\cdot, \cdot, \cdot]$	$f[\cdot, \cdot, \cdot, \cdot]$
0	0			
		1		
0	0	$\frac{2}{\pi}$	$\frac{4-2\pi}{\pi^2}$	
$\frac{\pi}{2}$	1		$-\frac{4}{\pi^2}$	$-\frac{16+4\pi}{\pi^3}$
$\frac{\pi}{2}$	1	0		

Logo

$$H_3(x) = x + \frac{4 - 2\pi}{\pi^2} x^2 - \frac{16 + 4\pi}{\pi^3} x^2 (x - \frac{\pi}{2}) = x[1 + x[-0.231 - 0.921(x - \frac{\pi}{2})]].$$

2.4.2 Erro de interpolação

O estudo do erro na interpolação de Hermite consiste na generalização do estudo efectuado para a interpolação de Lagrange de acordo com o seguinte teorema.

Teorema 2.22 *Seja H_{2n+1} o polinómio, de grau menor ou igual a $2n+1$ interpolador de Hermite da função f nos pontos distintos $x_0, x_1, \dots, x_n \in [a, b]$. Se $f \in C^{2n+2}([a, b])$ então para cada $x \in [a, b]$ existe $\xi = \xi(x) \in]a, b[$ tal que*

$$e(x) = f(x) - H_{2n+1}(x) = \frac{f^{(2n+2)}(\xi)}{(2n+2)!} w^2(x), \tag{2.18}$$

onde w é a função dada por (2.8).

Demonstração: Se $x = x_i$, para algum i o resultado está provado. Se $x \neq x_i, i = 0, 1, \dots, n$, definamos a função auxiliar

$$F(t) = f(t) - H_{2n+1}(t) - \frac{w(t)^2}{w(x)^2} (f(x) - H_{2n+1}(x)).$$

Como $F(t) = 0$ possui $2n + 3$ raízes ($n + 1$ zeros duplos $x_i, i = 0, \dots, n$ e uma raiz simples x) temos, por aplicação do Teorema de

Rolle generalizado, que $F^{(2n+2)}(t) = 0$ possui, pelo menos, uma raiz em $]a, b[$. Seja ξ essa raiz. Uma vez que

$$F^{(2n+2)}(t) = f^{(2n+2)}(t) - \frac{(2n+2)!}{w(x)^2} (f(x) - H_{2n+1}(x)),$$

substituindo t por ξ obtém-se o resultado pretendido. \square

Observação 2.23 *Tal como no caso da interpolação de Lagrange podemos efectuar as seguintes observações.*

1. *Pelo teorema anterior, podemos determinar um majorante para o erro cometido ao substituir f pelo seu polinómio interpolador de Hermite de grau n , H_{2n+1} . De facto, de (2.18) sai que:*

$$\|f - H_{2n+1}\|_{\infty} \leq \frac{\|f^{(2n+2)}\|_{\infty}}{(2n+2)!} \|w^2\|_{\infty}.$$

2. *Atendendo a que $x, x_j \in [a, b]$, temos que $|x - x_j| \leq (b - a)$ e portanto*

$$\|f - H_{2n+1}\|_{\infty} \leq \frac{\|f^{(2n+2)}\|_{\infty}}{(2n+2)!} (b - a)^{2n+2}.$$

3. *Uma vez que*

$$\|w\|_{\infty} \leq \frac{h^{n+1}n!}{4},$$

com $h = \max_{i=1, \dots, n} (x_i - x_{i-1})$, temos que

$$\|f - H_{2n+1}\|_{\infty} \leq \|f^{(2n+2)}\|_{\infty} \frac{h^{2n+2}(n!)^2}{16(2n+2)!}.$$

Observamos que dependendo do comportamento de $\|f^{(2n+2)}\|_{\infty}$ podemos, ou não, concluir que o aumento do grau do polinómio interpolador de Hermite implica uma diminuição do erro cometido ao aproximar a função por este polinómio. Uma forma de minimizar o erro consiste na utilização de polinómios interpoladores de Hermite segmentados.

2.4.3 Interpolação de Hermite segmentada

Consideremos um intervalo $[a, b]$ e uma partição dada por (2.1). Designemos por **polinómio segmentado cúbico** (ou **função cúbica por segmentos**) na partição Δ dada por (2.1), uma função S contínua em $[a, b]$ que, quando restringida a cada um dos intervalos $[x_{i-1}, x_i]$, $i = 1, \dots, n$, da partição, coincide com um polinómio de grau menor ou igual a três. Neste caso, dizemos que $S \in \mathcal{P}_3^1(\Delta)$.

Seja f uma função conhecida nos pontos da partição (2.1). Como se sabe, existe um e um só polinómio segmentado cúbico S_H tal que

$$\begin{aligned} S_H(x_i) &= f(x_i) \\ S'_H(x_i) &= f'(x_i) \end{aligned}, \quad i = 0, 1, \dots, n.$$

Nestas condições, S_H é chamado o **polinómio interpolador (de Hermite) segmentado cúbico** de f nos pontos de (2.1). Temos que

$$S_H(x) = \begin{cases} S_H^{(1)}(x) & x \in [x_0, x_1] \\ S_H^{(2)}(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_H^{(i)}(x) & x \in [x_{i-1}, x_i] \\ \vdots & \vdots \\ S_H^{(n)}(x) & x \in [x_{n-1}, x_n] \end{cases},$$

onde $S_H^{(i)}$ pode ser escrita na forma seguinte

$$\begin{aligned} S_H^{(i)}(x) &= f(x_{i-1}) + f'(x_{i-1})(x - x_{i-1}) + f[x_{i-1}, x_{i-1}, x_i](x - x_{i-1})^2 \\ &\quad + f[x_{i-1}, x_{i-1}, x_i, x_i](x - x_{i-1})^2(x - x_i). \end{aligned}$$

O que foi feito para $S_H \in \mathcal{P}_3^1(\Delta)$ poderia ser generalizado para $S_H \in \mathcal{P}_{2m+1}^1(\Delta)$, desde que se conhecesse a função f e a sua derivada nos pontos de Δ e nos pontos intermédios $x_{i-(m-j)/m} \in]x_{i-1}, x_i[$, com $j = 1, \dots, m-1$, $i = 1, \dots, n$. (Determine a expressão analítica de $S_H \in \mathcal{P}_{2m+1}^1(\Delta)$ nas condições descritas.)

Vamos agora estudar o erro que se comete ao aproximar f por $S_H \in \mathcal{P}_{2m+1}^1(\Delta)$ apresentando o seguinte teorema do erro, cuja demonstração resulta imediatamente do Teorema 2.22.

Teorema 2.24 *Seja f uma função com derivadas contínuas até à ordem $2m+2$ em $]a, b[$, com $m \in \mathbb{N}$. Consideremos a partição Δ dada em*

(2.1) e ainda os pontos $x_{i-(m-j)/j}$, $j = 1, \dots, m-1$, em cada intervalo $]x_{i-1}, x_i[$, $i = 1, \dots, n$. Nestas condições, se $S_H \in \mathcal{P}_{2m+1}^1(\Delta)$ oscula duas vezes com f os pontos de Δ bem como os pontos intermédios $x_{i-(m-j)/j}$, $j = 1, \dots, m-1$, $i = 1, \dots, n$, então

$$\|f - S_H\|_\infty \leq \frac{\|f^{(2m+2)}\|_\infty}{(2m+2)!} \max_{i=1, \dots, n} \|(w^{(i)})^2\|_{i, \infty},$$

com

$$\|(w^{(i)})^2\|_{i, \infty} = \max_{x \in [x_{i-1}, x_i]} \prod_{j=0}^m |x - x_{i-(m-j)/m}|^2.$$

Como é que poderemos construir uma base para o espaço $\mathcal{P}_3^1(\Delta)$? Para responder a essa questão, comecemos por notar que $\mathcal{P}_3^1(\Delta)$ é um espaço vectorial de dimensão $2n+2$. Determinemos agora as funções $\{\phi_j\}_{j=0}^n$ e $\{\bar{\phi}_j\}_{j=1}^n$ tais que, para todo o $S_H \in \mathcal{P}_3^1(\Delta)$,

$$S_H(x) = \sum_{i=1}^n (f(x_i)\phi_i(x) + f'(x_i)\bar{\phi}_i(x)).$$

Consideremos $\phi_i, \bar{\phi}_i \in \mathcal{P}_3^1(\Delta)$, $i = 0, \dots, n$, tais que, para todo $i, j = 0, \dots, n$, $\phi_i(x_j) = \delta_{ij}$, $\phi_i'(x_j) = 0$, $\bar{\phi}_i(x_j) = 0$, $\bar{\phi}_i'(x_j) = \delta_{ij}$. Prove que as funções $\phi_i, \bar{\phi}_i \in \mathcal{P}_3^1(\Delta)$, $i = 0, \dots, n$, pretendidas têm as seguintes expressões analíticas (prove):

$$\phi_i(x) = \begin{cases} \frac{(x_i - x_{i-1}) - 2(x - x_i)}{(x_i - x_{i-1})^3} (x - x_{i-1})^2 & x \in [x_{i-1}, x_i] \\ \frac{(x_i - x_{i+1}) - 2(x - x_i)}{(x_i - x_{i+1})^3} (x - x_{i+1})^2 & x \in [x_i, x_{i+1}] \\ 0 & x \in [a, x_{i-1}] \cup [x_{i+1}, b] \end{cases}, \quad i = 1, \dots, n$$

$$\phi_0(x) = \begin{cases} \frac{(x_0 - x_1) - 2(x - x_0)}{(x_0 - x_1)^3} (x - x_1)^2 & x \in [a, x_1] \\ 0 & x \in [x_1, b] \end{cases},$$

$$\phi_n(x) = \begin{cases} 0 & x \in [a, x_{n-1}] \\ \frac{(x_n - x_{n-1}) - 2(x - x_n)}{(x_n - x_{n-1})^3} (x - x_{n-1})^2 & x \in [x_{n-1}, b] \end{cases},$$

e

$$\bar{\phi}_i(x) = \begin{cases} \frac{(x-x_i)(x-x_{i-1})^2}{(x_i-x_{i-1})^2} & x \in [x_{i-1}, x_i] \\ \frac{(x-x_i)(x-x_{i+1})^2}{(x_i-x_{i+1})^2} & x \in [x_i, x_{i+1}] \\ 0 & x \in [a, x_{i-1}] \cup [x_{i+1}, b] \end{cases}, \quad i = 1, \dots, n-1,$$

$$\bar{\phi}_0(x) = \begin{cases} \frac{(x-x_0)(x-x_1)^2}{(x_0-x_1)^2} & x \in [a, x_1] \\ 0 & x \in [x_1, b] \end{cases},$$

$$\bar{\phi}_n(x) = \begin{cases} 0 & x \in [a, x_{n-1}] \\ \frac{(x-x_n)(x-x_{n-1})^2}{(x_n-x_{n-1})^2} & x \in [x_{n-1}, b] \end{cases}.$$

2.4.4 Polinómios osculadores

Para finalizar esta secção vamos generalizar o raciocínio efectado na obtenção do polinómio interpolador de Hermite para determinar polinómios que osculem os pontos de suporte mais do que duas vezes.

Suponhamos que, dada uma função f suficientemente diferenciável, queremos determinar um polinómio H_N que verifique

$$\frac{d^j f}{dx^j}(x_i) = \frac{d^j H_N}{dx^j}(x_i), \quad i = 0, \dots, n, \quad j = 0, \dots, r_i. \quad (2.19)$$

Quando tal situação acontece dizemos que f e H_N são funções que r_i -osculam (osculam r_i vezes) o ponto x_i , $i = 0, 1, \dots, n$. Apresentemos o seguinte teorema sem demonstração.

Teorema 2.25 *Existe um único polinómio H_N , de grau menor ou igual a N , com*

$$N = n + \sum_{j=0}^n r_j,$$

que satisfaz (2.19).

A determinação do polinómio referido no teorema anterior pode ser feita de forma análoga à do polinómio interpolador de Hermite. Não iremos considerar o caso geral mas sim um exemplo elucidativo.

Exemplo 2.26 Seja f uma função definida num intervalo $[a, b]$ e suponhamos são dados os valores $f(a), f(b), f'(a), f'(b)$ e $f''(a)$. Determine o polinómio $H_4(x)$ que verifica

$$H_4(a) = f(a), H_4(b) = f(b), H_4'(a) = f'(a), H_4'(b) = f'(b), H_4''(a) = f''(a).$$

Resolução: O polinómio $H_4(x)$ vai pertencer ao espaço dos polinómios de grau menor ou igual a 4. Neste espaço as funções

$$1, (x - a), (x - a)^2, (x - a)^3, (x - a)^3(x - b),$$

são linearmente independentes e portanto

$$H_4(x) = C_0 + C_1(x - a) + C_2(x - a)^2 + C_3(x - a)^3 + C_4(x - a)^3(x - b).$$

Determinemos as constantes atendendo às condições impostas a este polinómio. Atendendo a que $H_4(a) = f(a)$ vem $C_0 = f(a)$. Atendendo a que $H_4'(a) = f'(a)$ obtemos $C_1 = f'(a) = f[a, a]$. Atendendo a que $f''(a) = H_4''(a)$ vem $C_2 = f''(a)/2$ e que pode ser representado por $f[a, a, a]$.

De facto, comecemos por notar que

$$\begin{aligned} f[a, a + h, a + h + k] &= \frac{f[a + h, a + h + k] - f[a, a + h]}{h + k} \\ &= \frac{\frac{f(a + h + k) - f(a + h)}{k}}{h + k}, \end{aligned}$$

e portanto

$$\frac{f(a + h + k) - f(a + h)}{k} - \frac{f(a + h) - f(a)}{h} = f[a, a + h, a + h + k](h + k).$$

Mas,

$$f(a + h) - f(a) = f'(a)h + h\mathcal{O}(h), \quad f(a + h + k) - f(a + h) = f'(a + h)k + k\mathcal{O}(k),$$

em que $\mathcal{O}(h)$ e $\mathcal{O}(k)$ denotam quantidades que convergem para zero com h e k respectivamente. Substituindo vem

$$f[a, a + h, a + h + k](h + k) = f'(a + h) + \mathcal{O}(k) - f'(a) - \mathcal{O}(h).$$

Consideremos, na igualdade anterior o limite quando $h \rightarrow 0, k \rightarrow 0$. Existindo este limite, em particular existe quando $h = k$ e deduzimos

$$\lim_{h \rightarrow 0} \lim_{k \rightarrow 0} f[a, a + h, a + h + k] = \frac{f''(a)}{2}$$

e portanto podemos considerar

$$f[a, a, a] = \frac{f''(a)}{2}.$$

De $H_4(b) = f(b)$ deduzimos uma expressão para C_3 e, facilmente se obtém,

$$C_3 = f[a, a, a, b].$$

Considerando $H_4'(b) = f'(b)$ vem

$$C_4 = \frac{f'(b) - f[a, a] - 2f[a, a, a](b - a) - 3f[a, a, a, b](b - a)^2}{(b - a)^3},$$

e, tem-se

$$C_4 = f[a, a, a, b, b].$$

Substituindo, vem finalmente

$$\begin{aligned} H_4(x) = & f(a) + f[a, a](x - a) + f[a, a, a](x - a)^2 \\ & + f[a, a, a, b](x - a)^3 + f[a, a, a, b, b](x - a)^3(x - b). \end{aligned}$$

O polinómio anterior pode ser construído utilizando a tabela das diferenças divididas para os seguintes pontos $(y_i, g(y_i))$, $i = 0, \dots, 4$, com

$$y_0 = a, y_1 = a, y_2 = a, y_3 = b, y_4 = b$$

e

$$g(y_0) = g(y_1) = g(y_2) = f(a), \quad g(y_3) = g(y_4) = f(b)$$

e atendendo a que

$$\begin{aligned} g[y_0, y_1] &= f[a, a] = f'(a) \\ g[y_0, y_1, y_2] &= f[a, a, a] = f''(a)/2 \\ g[y_2, y_3] &= f[b, b] = f'(b) \\ g[y_3, y_4] &= f[b, b] = f'(b) \end{aligned} .$$

De acordo com o efectuado no exercício anterior, apresentamos, de seguida, um exemplo numérico.

Exemplo 2.27 Determine polinómio interpolador de uma função f para o suporte $f(0) = -1$, $f'(0) = -2$, $f(1) = 0$, $f'(1) = 10$ e $f''(1) = 40$.

Resolução: O polinómio pretendido pode ser determinado com a ajuda de seguinte tabela

x_i	$f(x_i)$	$f[\cdot, \cdot]$	$f[\cdot, \cdot, \cdot]$	$f[\cdot, \cdot, \cdot, \cdot]$	$f[\cdot, \cdot, \cdot, \cdot, \cdot]$
0	-1				
		-2			
0	-1		3		
		1		6	
1	0		9		5
		10		11	
1	0		20		
		10			
1	0				

Assim

$$H_4(x) = -1 - 2x + 3x^2 + 6x^2(x-1) + 5x^2(x-1)^2 = -1 + x[-2 + x[3 + (x-1)[6 + 5(x-1)]]$$

Observação 2.28 Note-se que o polinómio de Maclaurin de f de grau n oscula, com f , $n + 1$ vezes a origem. De facto, sendo o polinómio de Maclaurin de f dado por

$$P_n(x) = f(0) + f'(0)x + \dots + \frac{f^{(n)}(0)}{n!}x^n,$$

temos que

$$\begin{aligned} f(0) &= P_n(0) \\ f'(0) &= P'_n(0) \\ &\dots \\ f^{(n)}(0) &= P_n^{(n)}(0) \end{aligned}$$

A questão natural que se coloca é a seguinte: poderemos determinar um polinómio segmentado que apresente as derivadas de primeira e segunda ordens contínuas nos pontos da partição Δ , isto é, poderemos determinar $S \in \mathcal{P}_m^2(\Delta)$, com $m \in \mathbb{N}$, interpolador de f nos pontos de Δ ? A resposta à questão anterior é afirmativa. De facto, considerando apenas o caso em que $S \in \mathcal{P}_5^2(\Delta)$, temos que o polinómio que verifica as condições

$$S(x_i) = f(x_i), \quad S'(x_i) = f'(x_i), \quad S''(x_i) = f''(x_i),$$

é dado por

$$S(x) = \begin{cases} S^{(1)}(x) & x \in [x_0, x_1] \\ S^{(2)}(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S^{(i)}(x) & x \in [x_{i-1}, x_i] \\ \vdots & \vdots \\ S^{(n)}(x) & x \in [x_{n-1}, x_n] \end{cases},$$

com

$$\begin{aligned} S^{(i)}(x) = & f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 \\ & + f[x_0, x_0, x_0, x_1](x - x_0)^3 + f[x_0, x_0, x_0, x_1, x_1](x - x_0)^3(x - x_1) \\ & + f[x_0, x_0, x_0, x_1, x_1, x_1](x - x_0)^3(x - x_1)^2. \end{aligned}$$

O grau mínimo para o polinómio segmentado que apresenta primeira e segunda derivadas contínuas nos pontos da partição é 5. Será, no entanto, que existe um polinómio de grau 3 que tenha as derivadas de primeira e segunda ordens contínuas nos pontos da partição? Esta questão é o alvo de estudo da secção seguinte.

2.4.5 Exercícios

Exercício 2.4.1 Construa um algoritmo para determinar o valor do polinómio interpolador de Hermite num determinado ponto do seu domínio.

Exercício 2.4.2 Determine o polinómio interpolador de Hermite de grau mínimo para a função $f(x) = \sin x$ quando se considera $x \in [0, \pi/2]$.

Exercício 2.4.3 Determine, de dois modos distintos, o polinómio interpolador de Hermite para os dados

x_i	0	0.25	0.5
$f(x_i)$	0.75	1	0.25
$f'(x_i)$	0.25	0.5	0.75

Exercício 2.4.4 Considere a função $f(x) = 3xe^x - e^{2x}$. Determine uma aproximação para $f(1.03)$ usando o polinómio interpolador de Hermite considerando os pontos $x_0 = 0$ e $x_1 = 1.05$. Determine uma estimativa para o erro cometido.

Exercício 2.4.5 Considere $f(x) = e^x$. Determine o polinómio interpolador de Hermite de grau 5 – H_5 – usando os pontos $x_0 = 0$, $x_1 = 1$ e $x_2 = 2$. Compare $H_5(0.25)$ com $f(0.25)$ e com $P_2(0.25)$ em que P_2 é o polinómio interpolador de Lagrange de grau 2, para a função f , determinado nos mesmos pontos.

Exercício 2.4.6 Determine o polinómio de grau mínimo que faça a concordância entre a recta

$$y = -2 + \frac{1}{2}(8 - x), \quad \text{no ponto } (8, -2),$$

e a circunferência

$$(x - 1)^2 + (y + 2)^2 = 1, \quad \text{no ponto } (1, -1).$$

Nota: Duas curvas dizem-se concordantes de tiverem a mesma tangente no ponto de união.

Exercício 2.4.7 Considere $f(x) = e^x$. Determine o polinómio interpolador de Hermite de grau 5 – H_5 – usando os pontos $x_0 = 0$, $x_1 = 1$ e $x_2 = 2$. Compare $H_5(0.25)$ com $f(0.25)$ e com $P_2(0.25)$ em que P_2 é o polinómio interpolador de Lagrange de grau 2, para a função f , determinado nos mesmos pontos.

Exercício 2.4.8 Mostre que o erro que se comete ao aproximar $f \in C^4([a, b])$ pelo seu polinómio interpolador segmentado de Hermite cúbico na partição (2.1) é dado por

$$\max_{x \in [a, b]} |f(x) - S_H(x)| \leq \frac{M_4}{384} h^4,$$

onde

$$M_4 = \max_{x \in [a, b]} |f^{(4)}(x)|$$

e $h = \max_{i=1, \dots, n} (x_i - x_{i-1})$.

Exercício 2.4.9 Considere as funções f e g das quais se conhecem os valores

x_i	0	1
$f(x_i)$	-1	0
$f'(x_i)$	-2	10
$f''(x_i)$		40

x_i	0	0.25
$g(x_i)$	0.75	1
$g'(x_i)$	0.25	0.5
$g''(x_i)$	0.25	0.5

1. Determine o polinómio P que oscula com f duas vezes o ponto $x_0 = 0$ e três vezes o ponto $x_1 = 1$.
2. Determine o polinómio P que oscula com g duas vezes os pontos $x_0 = 0$ e $x_1 = 0.25$.

Exercício 2.4.10 Determine o polinómio interpolador de Hermite segmentado cúbico de f sabendo que

x_i	0	0.25	0.5
$f(x_i)$	0.5	1	0.5
$f'(x_i)$	0.5	0	-0.5

Exercício 2.4.11 Suponha que conhece uma função e a sua derivada nos pontos

$$x_0 < x_{1/2} < x_1 < x_{3/2} < x_2.$$

Determine o polinómio interpolador de Hermite segmentado de grau 5 de dois modos distintos.

Exercício 2.4.12 Particularize o exercício anterior para a função $f(x) = \cos x$ considerando os pontos

$$x_0 = 0, \quad x_{1/2} = \frac{\pi}{4}, \quad x_1 = \frac{\pi}{2}, \quad x_{3/2} = \frac{5\pi}{4}, \quad x_2 = \pi.$$

2.5 Interpolação com funções *spline*

2.5.1 Abordagem clássica

O termo inglês *spline* pode ser traduzido pelo vocábulo “virote”. Um virote é um instrumento usado pelos desenhadores para unir um conjunto de pontos do plano.

Seja f uma função definida num intervalo $[a, b]$ onde consideramos a partição Δ dada por (2.1). Matematicamente, o problema de unir pontos do plano com um virote pode ser traduzido da seguinte forma: determinar a função $S : [a, b] \rightarrow \mathbb{R}$, com $a = x_0$, $b = x_n$, que satisfaz:

$$[S1] \quad S(x_i) = f(x_i), \quad i = 0, \dots, n;$$

$$[S2] \quad S \in C^2([a, b]);$$

[S3] o princípio de Maupertius da energia mínima, isto é,

$$\int_a^b (S''(x))^2 dx \leq \int_a^b (g''(x))^2 dx,$$

para toda a função g que satisfaz [S1] e [S2].

Temos os seguinte resultado.

Teorema 2.29 *Sejam $S, g : [a, b] \rightarrow \mathbb{R}$ duas funções que verificam [S1] e [S2]. Suponhamos que*

$$S''(b)[g'(b) - S'(b)] = S''(a)[g'(a) - S'(a)]$$

e que S é um polinómio de grau 3 em cada sub-intervalo da partição dada. Então temos que

$$\int_a^b (S''(x))^2 dx \leq \int_a^b (g''(x))^2 dx.$$

Demonstração: Temos que

$$\int_a^b (g''(x))^2 dx - \int_a^b (S''(x))^2 dx = \int_a^b (g''(x) - S''(x))^2 dx + 2 \int_a^b S''(x)(g''(x) - S''(x)) dx$$

Mas, integrando por partes o último integral do segundo membro vem

$$\begin{aligned} \int_a^b S''(x)(g''(x) - S''(x)) dx &= S''(b)[g'(b) - S'(b)] - S''(a)[g'(a) - S'(a)] \\ &\quad - \sum_{i=1}^n \int_{x_{i-1}}^{x_i} S'''(x)(g''(x) - S''(x)) dx. \end{aligned}$$

Ora, atendendo às hipóteses do teorema,

$$\int_a^b S''(x)(g''(x) - S''(x)) dx = 0$$

e, como tal,

$$\int_a^b (S''(x))^2 dx = \int_a^b (g''(x))^2 dx - \int_a^b (g''(x) - S''(x))^2 dx,$$

o que permite concluir o pretendido. \square

Este teorema mostra que os candidatos à resolução de [S1]–[S3] são as funções pertencentes a $C^2([a, b])$ que são polinómios de grau 3 em cada intervalo da partição, isto é, são funções $S \in \mathcal{P}_3^2(\Delta)$. Essas funções serão designadas por funções *spline* cúbicas.

Definição 2.30 (Spline) Uma função spline de ordem m é um polinómio segmentado de grau $m-1$ continuamente derivável até à ordem $m-2$. Por outras palavras, dada uma partição Δ da forma (2.1), S é uma função spline de ordem m se $S \in \mathcal{P}_{m-1}^{m-2}(\Delta)$, ou seja, se

$$\frac{d^k}{dx^k} S^{(i+1)}(x) = \frac{d^k}{dx^k} S^{(i)}(x), \quad k = 0, \dots, m-2, i = 1, \dots, n-1,$$

com $S^{(i)}(x) = S(x)$, $x \in [x_{i-1}, x_i]$, $i = 1, \dots, n$.

As funções spline mais populares são as cúbicas ($m = 4$). Pelas razões apresentadas, serão essas que iremos considerar.

Note-se que, em cada intervalo $[x_{i-1}, x_i]$ a função $S \in \mathcal{P}_3^2(\Delta)$ que interpola f nos pontos da partição (2.1) é um polinómio de grau 3 e, como tal, é definido à custa de 4 parâmetros. Assim, para determinar S de forma única temos que especificar $4n$ parâmetros. Para isso teremos que definir $4n$ equações. Atendendo à definição de função spline temos impostas as seguintes equações: $n+1$ equações de interpolação; $n+1$ equações de ligação de S ; $n+1$ equações de ligação de S' e $n+1$ equações de ligação de S'' . No total temos assim $4n-2$ equações. Para determinar S temos que considerar mais duas condições suplementares. As formas mais usuais de definir essas condições são as seguintes: $S'(a) = f'(a)$ e $S'(b) = f'(b)$ (Spline completa); $S''(a) = f''(a)$ e $S''(b) = f''(b)$ (Spline natural); $S''(a) = 0$ e $S''(b) = 0$ (Spline livre); $S'(a) = S'(b)$ e $S''(a) = S''(b)$ (Spline periódica).

O seguinte teorema, que apresentamos sem demonstração, estabelece a existência e unicidade da função spline cúbica interpoladora.

Teorema 2.31 Seja f uma função definida em $[a, b]$. A função spline cúbica completa (natural, livre, periódica) interpoladora de f nos pontos da partição Δ dada em (2.1) existe e é única.

2.5.2 Funções de base

Seja f uma função definida num intervalo $[a, b]$ onde consideramos a partição Δ dada por (2.1). Suponhamos que os pontos da partição são igualmente espaçados e seja $h = x_i - x_{i-1}$, $i = 1, \dots, n$. Consideremos os seguintes pontos auxiliares

$$x_{-1} = a - h, x_{-2} = a - 2h, x_{-3} = a - 3h$$

$$x_{n+1} = b + h, x_{n+2} = b + 2h, x_{n+3} = b + 3h.$$

Os polinómios segmentados

$$B_i(x) = \frac{1}{h^3} \begin{cases} (x - x_{i-2})^3 & x \in [x_{i-2}, x_{i-1}] \\ h^3 + 3h^2(x - x_{i-1}) + 3h(x - x_{i-1})^2 - 3(x - x_{i-1})^3, & x \in]x_{i-1}, x_i] \\ h^3 + 3h^2(x_{i+1} - x) + 3h(x_{i+1} - x)^2 - 3(x_{i+1} - x)^3, & x \in]x_i, x_{i+1}] \\ (x_{i+2} - x)^3 & x \in [x_{i+1}, x_{i+2}] \\ 0 & \text{nos restantes} \end{cases}$$

em que $i = -1, \dots, n + 1$, são designadas por funções *spline-B*.

Teorema 2.32 *As funções B_i , $i = -1, \dots, n + 1$, são tais que $B_i \in C^2([a, b])$ e*

	...	x_{i-2}	x_{i-1}	x_i	x_{i+1}	x_{i+2}
B_i	0	0	1	4	1	0
B'_i	0	0	$3/h$	0	$-3/h$	0
B''_i	0	0	$6/h^2$	$-12/h^2$	$6/h^2$	0

Demonstração: Fica ao cuidado do aluno. \square

O espaço gerado pelas funções B_i , $i = -1, \dots, n + 1$, será denotado por $\mathcal{B}_3(\Delta)$. Este espaço é constituído por polinómios do tipo

$$S(x) = \sum_{i=-1}^{n+1} C_i B_i(x),$$

em que C_s denotam constantes, e portanto as funções de $\mathcal{B}_3(\Delta)$ são polinómios segmentados cúbicos e que têm no intervalo $[x_i, x_{i+1}[$ a seguinte expressão

$$S(x) = C_{i-1} B_{i-1}(x) + C_i B_i(x) + C_{i+1} B_{i+1}(x) + C_{i+2} B_{i+2}(x),$$

ou seja,

$$\begin{aligned}
 S(x) = & \frac{1}{h^3} [C_{i-1}(x_{i+1} - x)^3 \\
 & + C_i(h^3 + 3h(x_{i+1} - x)^2 + 3h^2(x_{i+1} - x) - 3(x_{i+1} - x)^3) \\
 & + C_{i+1}(h^3 + 3h(x - x_i)^2 + 3h^2(x - x_i) - 3(x - x_i)^3) \\
 & + C_{i+2}(x - x_i)^3].
 \end{aligned} \tag{2.20}$$

Pretendemos determinar um polinómio em $\mathcal{B}_3(\Delta)$ que seja interpolador para a função f nos pontos da partição. Este polinómio é designado *spline* cúbico interpolador. Começemos por notar que as condições de interpolação $S(x_i) = f(x_i), i = 0, \dots, n$, para a determinação das constantes que definem o *spline* cúbico interpolador são $n + 1$ e portanto necessitamos de mais duas condições. No resultado seguinte estão especificadas duas condições que permitem definir o *spline* cúbico interpolador.

Teorema 2.33 *Seja f uma função definida em $[a, b]$ onde considerado $n + 1$ pontos igualmente distanciados $x_i, i = 0, \dots, n$, em que $x_i = x_{i-1} + h, i = 1, \dots, n, x_0 = a, h = (b - a)/n$. Seja \mathcal{B} o espaço gerado pelos splines cúbicos $B_i, i = -1, \dots, n + 1$. Dados $f(x_i), i = 0, \dots, n, f'(a)$ e $f'(b)$ existe um e um só polinómio $S(x)$ em \mathcal{B} que satisfaz a*

$$S(x_i) = f(x_i), i = 0, \dots, n, \quad S'(a) = f'(a), S'(b) = f'(b). \tag{2.21}$$

Demonstração: O polinómio *spline* cúbico interpolador pertence a \mathcal{B} e portanto existem constantes $c_j, j = -1, \dots, n + 1$, tais

$$S(x) = \sum_{j=-1}^{n+1} C_j B_j(x).$$

Estas constantes C_j são determinadas atendendo às condições (2.21). De facto, atendendo a estas condições temos

$$\begin{aligned} f'(a) &= \sum_{j=-1}^{n+1} C_j B'_j(x_0) \\ f(x_i) &= \sum_{j=-1}^{n+1} C_j B_j(x_i), \quad i = 0, \dots, n \\ f'(b) &= \sum_{j=-1}^{n+1} C_j B'_j(x_b). \end{aligned}$$

Atendendo às propriedades dos *splines* cúbicos, o sistema anterior é equivalente a

$$\begin{aligned} f'(a) &= B'_{-1}(x_0)C_0 + B'_0(x_0)C_0 + B'_1(x_0)C_1 \\ f(x_i) &= B_{i-1}(x_i)C_{i-1} + B_i(x_i)C_i + B_{i+1}(x_i)C_{i+1}, \quad i = 0, \dots, n \\ f'(b) &= B'_{n-1}(x_n)C_{n-1} + B'_n(x_n)C_n + B'_{n+1}(x_n)C_{n+1}, \end{aligned}$$

ou ainda, a

$$\begin{aligned} f'(a) &= -\frac{3}{h}C_0 + 0C_0 + \frac{3}{h}C_1 \\ f(x_i) &= 1C_{i-1} + 4C_i + 1C_{i+1}, \quad i = 0, \dots, n \\ f'(b) &= -\frac{3}{h}C_{n-1} + 0C_n + \frac{3}{h}C_{n+1}. \end{aligned}$$

Uma vez que a matriz do sistema

$$\mathbf{B} = \begin{bmatrix} -\frac{3}{h} & 0 & \frac{3}{h} & \dots & 0 & 0 & 0 \\ 1 & 4 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & 4 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 4 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & 4 & 1 \\ 0 & 0 & 0 & \dots & -\frac{3}{h} & 0 & \frac{3}{h} \end{bmatrix},$$

é não singular concluímos que existe um e um só *spline* cúbico interpolador que verifica as condições (2.21). \square

A determinação o *spline* cúbico que verifica as condições (2.21) é feita resolvendo o sistema $\mathbf{BC} = F$ em que \mathbf{B} está definida na demonstração do teorema anterior e

$$C = \begin{bmatrix} C_{-1} \\ C_0 \\ C_1 \\ \vdots \\ C_{n-1} \\ C_n \\ C_{n+1} \end{bmatrix}, \quad F = \begin{bmatrix} f'(a) \\ f(a) \\ f(x_1) \\ \vdots \\ f(x_{n-1}) \\ f(b) \\ f'(b) \end{bmatrix}.$$

O *spline* cúbico $S(x)$ é interpolador para f nos pontos da partição e ainda interpola f' nos extremos do intervalo. Conhecidos os coeficientes do *spline* cúbico interpolador ficamos com $S(x)$ determinado nos intervalos da partição e em $[x_i, x_{i+1}]$ temos que $S(x)$ é dado por (2.20).

Exemplo 2.34 Determine o *spline* cúbico para a função f dada pela tabela

x_i	2.2	2.4	2.6
$f(x_i)$	0.5207843	0.5104147	0.4813306
$f'(x_i)$	-0.014878		-0.1883635

Resolução: Consideremos os pontos auxiliares $x_{-1} = 2, x_{-2} = 1.8, x_3 = 2.8, x_4 = 3$, e sejam $B_i(x), i = -1, \dots, 3$, os *splines* cúbicos. O sistema que permite determinar os coeficientes do *spline* cúbico interpolador

$$S(x) = C_{-1}B_{-1}(x) + C_0B_0(x) + C_1B_1(x) + C_2B_2(x) + C_3B_3(x)$$

é

$$\begin{bmatrix} -15 & 0 & 15 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & -15 & 0 & 15 \end{bmatrix} \begin{bmatrix} C_{-1} \\ C_0 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} -0.014878 \\ 0.5207843 \\ 0.5104147 \\ 0.4813306 \\ -0.1883635 \end{bmatrix},$$

que tem por solução

$$C_{-1} = 0.0856896, C_0 = 0.087376, C_1 = 0.0855905, C_2 = 0.0806768, C_3 = 0.07$$

Em [2.2, 2.4] temos

$$\begin{aligned} S(x) &= (0.0856896(2.4 - x)^3 + 0.087376(0.2^3 + 3 \times 0.2^2(2.4 - x) \\ &+ 3 \times 0.2(2.4 - x)^2 - 3(2.4 - x)^3)0.0855905(0.2^3 + 3 \times 0.2^2(x - 2.2) \\ &+ 3 \times 0.2(x - 2.2)^2 - 3(x - 2.2)^3) + 0.0806768(x - 2.2)^3)/0.2^3 \\ &= -1.19371 + 1.76806x - 0.543945x^2 + 0.0429625x^3. \end{aligned}$$

e em [2.4, 2.6]

$$\begin{aligned} S(x) &= (0.087376(2.6 - x)^3 + 0.0855905(0.2^3 + 3 \times 0.2^2(2.6 - x) \\ &+ 3 \times 0.2(2.6 - x)^2 - 3(2.6 - x)^3) + 0.0806768(0.2^3 + 3 \times 0.2^2(x - 2.4) \\ &+ 3 \times 0.2(x - 2.4)^2 - 3(x - 2.4)^3) + 0.0730329(x - 2.4)^3)/0.2^3 \\ &= -1.28754 + 1.88534x - 0.592815x^2 + 0.04975x^3. \end{aligned}$$

O *spline* cúbico interpolador, é, como vimos, deduzido exigindo que este polinómio seja interpolador da função em todos os pontos da partição e que seja interpolador da derivada nos extremos do intervalo. Estas duas últimas condições permitiram construir o *spline* cúbico interpolador e podem ser substituídas por condições na segunda derivada do polinómio interpolador. No teorema seguinte é estabelecida a existência de *spline* interpolador que interpola a função e a sua segunda derivada nos extremos do intervalo. Este polinómio interpolador é designado *spline* cúbico natural. Se admitirmos as seguintes condições $S''(x_0) = S''(x_n) = 0$ então este *spline* é designado *spline* cúbico livre.

Teorema 2.35 *Seja f uma função definida em $[a, b]$ onde considerado $n + 1$ pontos igualmente distanciados $x_i, i = 0, \dots, n$, em que $x_i = x_{i-1} + h, i = 1, \dots, n, x_0 = a, h = (b - a)/n$. Seja \mathcal{B} o espaço gerado pelos splines cúbicos $B_i, i = -1, \dots, n + 1$. Dados $f(x_i), i = 0, \dots, n, f''(a)$ e $f''(b)$ existe um e um só polinómio $S(x)$ em \mathcal{B} que satisfaz a*

$$S(x_i) = f(x_i), i = 0, \dots, n, \quad S''(a) = f''(a), S''(b) = f''(b). \quad (2.22)$$

A demonstração do teorema anterior é análoga à do Teorema 2.33. Notamos que a matriz que permite determinar os coeficientes do *spline*

cúbico interpolador $S(x)$ é

$$\begin{bmatrix} \frac{6}{h^2} & -\frac{12}{h^2} & \frac{6}{h^2} & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 4 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & \frac{6}{h^2} & -\frac{12}{h^2} & \frac{6}{h^2} \end{bmatrix} \begin{bmatrix} C_{-1} \\ C_0 \\ C_1 \\ \vdots \\ C_{n-1} \\ C_n \\ C_{n+1} \end{bmatrix} = \begin{bmatrix} f''(a) \\ f(a) \\ f(x_1) \\ \vdots \\ f(x_{n-1}) \\ f(b) \\ f''(b) \end{bmatrix}.$$

2.5.3 Estudo do erro

O *spline* cúbico interpolador de uma função coincide com a função nos pontos da partição e a sua derivada coincide com a derivada da função nos extremos do intervalo de partição. No resultado seguinte, estabelecido sem demonstração, é apresentado o comportamento do erro que se comete ao aproximar uma função pelo seu *spline* cúbico interpolador.

Teorema 2.36 *Seja f uma função definida em $[a, b]$ onde considerado $n + 1$ pontos igualmente distanciados $x_i, i = 0, \dots, n$, em que $x_i = x_{i-1} + h, i = 1, \dots, n, x_0 = a, h = (b - a)/n$. O erro cometido ao aproximar f pelo seu *spline* cúbico interpolador verifica a*

$$\|S - f\|_\infty \leq \frac{5}{384} \|f^{(4)}\|_\infty h^4.$$

2.5.4 Exercícios

Exercício 2.5.1 A estrela S da Ursa Maior apresenta uma variação para a sua magnitude aparente m , em função do ângulo de fase θ (em graus), de acordo com os dados da seguinte tabela:

θ	-60	-20	20
m	9.40	11.39	10.84

Usando um *spline* cúbico livre, determine uma aproximação para o ângulo de fase pertencente ao intervalo $[-20, 20]$ em que a magnitude aparente da estrela é máxima.

Exercício 2.5.2 Pretende-se interpolar a função f definida por $f(x) = \ln x$, com $x \in [2, 2.5]$, por um *spline* cúbico completo S numa malha uniforme.

1. Calcule o número mínimo de pontos a usar para garantir que $\|f - S\|_{\infty} \leq 0.5 \times 10^{-4}$.
2. Determine uma aproximação para $f(2.3)$ usando o *spline* cúbico completo interpolador de f nos pontos obtidos na alínea anterior.

Exercício 2.5.3 Deslocando-se um receptor de GPS num veículo ao longo do eixo de uma estrada, em Angola, obtiveram-se as coordenadas locais:

latitude ϕ	26' 56".1	26' 50".4	27' 02".7	26' 58".3
longitude λ	5' 36"	5' 56"	6' 16"	6' 36"

Aproximando o eixo da estrada por um *spline* natural determine:

1. a latitude da estrada quando a longitude é $\lambda = 6'$;
2. as coordenadas da estrada no ponto mais perto do equador, supondo que isso acontece entre 6' 16" e 6' 36" de longitude.

Exercício 2.5.4 1. Usando um *spline* cúbico livre, determine uma aproximação para a declinação aparente de Vénus para o dia 9 de Maio de 1999, às 18h30m45s, a partir das Efemérides Astronómicas (onde está tabelada para cada dia, às zero horas)

dia	8	9	10
δ_i	+6°22'25".20	+6°52'54".57	+6°23'14".96

2. A partir da função obtida na alínea anterior, determine uma aproximação para o instante em que a declinação aparente de Vénus no dia 9 de Maio de 1999 foi máxima.

Exercício 2.5.5 Um carro percorre uma rua, em linha recta, tendo sido efectuados os seguintes registos:

tempo (t) em segundos	0	5	10
distância (d) em metros	0	90	150
velocidade (v) em km/hora	40		40

Usando o *spline* cúbico completo interpolador da função distância nos pontos dados, indique, justificando, uma aproximação para:

1. o primeiro instante em que o carro excedeu o limite de velocidade permitido dentro das localidades;
2. o instante em que o carro atingiu a velocidade máxima nos primeiros 5 segundos.

2.6 Derivação numérica

Acontece frequentemente sermos confrontados com a necessidade de determinar valores da derivada de uma função num conjunto de pontos conhecendo o valor da função apenas nesses pontos. Na impossibilidade de obter esses valores de forma exacta, vamos considerar a sua aproximação através do valor da derivada do polinómio interpolador da função nos referidos pontos.

Para o estudo que iremos efetuar nesta secção, consideremos uma função $f \in C^{n+1}([a, b])$ conhecida num conjunto de pontos da partição uniforme

$$a = x_0 < x_1 < \cdots < x_n = b, \quad (2.23)$$

com $x_i - x_{i-1} = h$, $i = 1, \dots, n$.

2.6.1 Aproximação da primeira derivada

Queremos aproximar a derivada de f num dos pontos x_k , $k \in \{0, 1, \dots, n\}$, da partição (2.23). Usando a fórmula interpoladora de Lagrange temos que, para $x \in]a, b[$,

$$f(x) = \sum_{i=0}^n f(x_i) \ell_i(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} w(x),$$

sendo ℓ_i , $i = 0, \dots, n$, os polinómios de Lagrange dados por (2.5), w a função dada por (2.8) e $\xi \in]a, b[$ um valor que depende de x . Derivando esta expressão obtemos

$$f'(x) = \sum_{i=0}^n f(x_i) \ell'_i(x) + \left(\frac{f^{(n+1)}(\xi)}{(n+1)!} w(x) \right)'$$

Podemos, assim, considerar a aproximação

$$f'(x) \approx \sum_{i=0}^n f(x_i) \ell'_i(x),$$

com erro dado por

$$e(x) := \left(\frac{f^{(n+1)}(\xi)}{(n+1)!} w(x) \right)' = \frac{(f^{(n+1)}(\xi))'}{(n+1)!} w(x) + \frac{w'(x)}{(n+1)!} f^{(n+1)}(\xi)$$

Ora, como se sabe,

$$w'(x) = \sum_{l=0}^{n-1} \prod_{j=0, j \neq l}^{n-1} (x - x_j);$$

a dificuldade reside no cálculo de $(f^{(n+1)}(\xi))'$, uma vez que ξ depende de x . No entanto, para um ponto $x = x_k$, $k \in \{0, 1, \dots, n\}$, da partição (2.23) temos que $w(x_k) = 0$ e como tal

$$f'(x_k) = \sum_{i=0}^n f(x_i) \ell'_i(x_k) + \frac{f^{(n+1)}(\xi)}{(n+1)!} w'(x_k), \quad (2.24)$$

Como $x_k - x_j = (k - j)h$ temos que a fórmula que nos permite aproximar a derivada é

$$f'(x_k) \approx \sum_{i=0}^n f(x_k + (i - k)h) \ell'_i(x_k), \quad (2.25)$$

e o erro cometido é dado por

$$e(x_k) = h^n \frac{f^{(n+1)}(\xi)}{(n+1)!} \sum_{l=0}^{n-1} \prod_{j=0, j \neq l}^{n-1} (k - j).$$

Podemos então concluir que

$$|e(x_k)| \leq C_{n+1} h^n,$$

onde C_{n+1} é um valor que não depende de h . Por este facto dizemos que a fórmula usada para aproximar a derivada da função é de ordem n . É também usual usar a notação $e(x_k) = \mathcal{O}(h^n)$.

Observação 2.37 *Se existir uma constante positiva C tal que*

$$C_{n+1} \leq C, \quad \forall n \in \mathbb{N},$$

concluimos que o aumento do número de pontos de interpolação implica uma diminuição do erro cometido na aproximação (2.25).

Atendendo a que, pela fórmula interpoladora de Newton das diferenças progressivas, o polinómio interpolador de f nos pontos da partição (2.23) é dado por

$$f(x_0) + \sum_{i=1}^n \frac{\Delta^i f(x_0)}{i! h^i} \prod_{j=0}^{i-1} (x - x_j),$$

resulta que a fórmula (2.25) pode ser escrita na forma (prove)

$$f'(x_k) \approx \sum_{i=0}^n \frac{\Delta^i f(x_0)}{i!h} \left(\sum_{l=0}^{i-1} \prod_{j=0, j \neq l}^{i-1} (k-j) \right). \quad (2.26)$$

A fórmula de diferenças finitas (2.26), convenhamos, não é nada simpática. Vamos particularizá-la deduzindo várias fórmulas de diferenças finitas para aproximar a derivada de uma função f .

Fórmulas com dois pontos

Usando a fórmula interpoladora de Newton temos que, para $x \in [x_k, x_{k+1}]$, $k = 0, \dots, n-1$,

$$f(x) = f(x_k) + \frac{f(x_k) - f(x_{k+1})}{h}(x-x_k) + \frac{f''(\xi)}{2}(x-x_k)(x-x_{k+1}), \quad \xi \in]x_k, x_{k+1}[.$$

Derivando sai que

$$f'(x_k) = \frac{f(x_k) - f(x_{k+1})}{h} - h \frac{f''(\xi)}{2}, \quad \xi \in]x_k, x_{k+1}[,$$

e

$$f'(x_{k+1}) = \frac{f(x_k) - f(x_{k+1})}{h} + h \frac{f''(\xi)}{2} \quad \xi \in]x_k, x_{k+1}[.$$

Obtemos assim duas fórmulas de diferenças finitas de primeira ordem para aproximar a primeira derivada de uma função num ponto. A

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

é usual chamar fórmula de diferenças progressivas e a

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

costuma chamar-se fórmula de diferenças regressivas.

Fórmulas com três pontos

Para obter fórmulas mais precisas para aproximar a primeira derivada de uma função num ponto, vamos considerar fórmulas com mais pontos. As fórmulas de diferenças progressivas, centradas e regressivas com três pontos são as seguintes:

1. $f'(x_k) = \frac{1}{2h} [-3f(x_k) + 4f(x_{k+1}) - f(x_{k+2})] + \frac{h^2}{3} f'''(\xi_0);$
2. $f'(x_k) = \frac{1}{2h} [-f(x_{k-1}) + f(x_{k+1})] - \frac{h^2}{6} f'''(\xi_1);$
3. $f'(x_k) = \frac{1}{2h} [f(x_{k-2}) - 4f(x_{k-1}) + 3f(x_k)] + \frac{h^2}{3} f'''(\xi_2);$

Vamos só deduzir a segunda fórmula. Assim, temos que

$$f(x) = f(x_{k-1}) + \frac{\Delta f(x_{k-1})}{h}(x - x_{k-1}) + \frac{\Delta^2 f(x_{k-1})}{2h^2}(x - x_{k-1})(x - x_k) + \frac{f'''(\xi)}{6}(x - x_{k-1})(x - x_k)(x - x_{k+1}), \quad \xi_1 \in]x_{k-1}, x_{k+1}[.$$

Derivando sai que

$$f'(x_k) = \frac{\Delta f(x_{k-1})}{h} + \frac{\Delta^2 f(x_{k-1})}{2h^2} [(x_k - x_{k-1}) + (x_k - x_k)] + \frac{f'''(\xi_1)}{6} (x_k - x_{k-1})(x_k - x_{k+1}), \quad \xi_1 \in]x_{k-1}, x_{k+1}[.$$

Atendendo a que

x_i	$f(x_i)$	$\Delta f(x_i)$	$\Delta^2 f(x_i)$
x_{k-1}	$f(x_{k-1})$		
		$f(x_k) - f(x_{k-1})$	
x_k	$f(x_k)$		$f(x_{k+1}) - 2f(x_k) + f(x_{k-1})$
		$f(x_{k+1}) - f(x_k)$	
x_{k+1}	$f(x_{k+1})$		

temos

$$f'(x_k) = \frac{f(x_k) - f(x_{k-1})}{h} + \frac{f(x_{k+1}) - 2f(x_k) + f(x_{k-1}))}{2h} - h^2 \frac{f'''(\xi_1)}{6},$$

com $\xi_1 \in]x_{k-1}, x_{k+1}[$, ou seja

$$f'(x_k) = \frac{1}{2h} [-f(x_{k-1}) + f(x_{k+1})] - h^2 \frac{f'''(\xi)}{6}, \quad \xi_1 \in]x_{k-1}, x_{k+1}[.$$

Exemplo 2.38 Considere os seguintes valores da função $f(x) = xe^x$:

x_i	1.8	1.9	2.0	2.1	2.2
$f(x_i)$	10.889365	12.703199	14.778112	17.148957	19.855030

Aproxime o valor de $f'(2.0) = 22.167168$ usando as fórmulas de diferenças finitas dadas no exercício anterior e compare os erros cometidos.

Resolução: Vamos considerar as três fórmulas separadamente.

- Fórmula progressiva de segunda ordem com $h = 0.1$.

$$f'(2.0) \approx \frac{1}{0.2}[-3f(2.0) + 4f(2.1) - f(2.2)] = 22.032310.$$

O erro cometido é aproximadamente 1.35×10^{-1} .

- Fórmula regressiva de segunda ordem com $h = 0.1$.

$$f'(2.0) \approx \frac{1}{0.2}[f(1.8) - 4f(1.9) + 3f(2.0)] = 22.054525.$$

O erro cometido é aproximadamente 1.13×10^{-1} .

- Fórmula centrada de segunda ordem com $h = 0.1$.

$$f'(2.0) \approx \frac{1}{0.2}[f(2.1) - f(1.9)] = 22.228790.$$

O erro cometido é aproximadamente -6.16×10^{-2} .

Note-se que o erro cometido quando se usa a fórmula de diferenças centradas.

2.6.2 Aproximação da segunda derivada. Algumas fórmulas

Queremos aproximar a segunda derivada de f num dos pontos x_k , $k \in \{0, 1, \dots, n\}$, da partição (2.23). Poderíamos, tal como para a primeira derivada, usar o polinómio interpolador na dedução das fórmulas para a segunda derivada. A obtenção de estimativas para o erro é, no entanto, mais complicada. Um processo alternativo para a dedução das fórmulas de derivação (e respectivo erro) faz uso da série de Taylor da função.

Desenvolvendo f em série de Taylor em torno do ponto x_k temos:

$$f(x_{k+1}) = f(x_k) + f'(x_k)h + \frac{h^2}{2}f''(x_k) + \frac{h^3}{6}f'''(x_k) + \frac{h^4}{24}f^{(4)}(\xi_1), \quad \xi_1 \in]x_k, x_{k+1}[;$$

$$f(x_{k-1}) = f(x_k) - f'(x_k)h + \frac{h^2}{2}f''(x_k) - \frac{h^3}{6}f'''(x_k) + \frac{h^4}{24}f^{(4)}(\xi_2), \quad \xi_2 \in]x_{k-1}, x_k[.$$

Se adicionarmos estas duas expressões obtemos

$$f''(x_k) = \frac{1}{h^2}[f(x_{k-1}) - 2f(x_k) + f(x_{k+1})] - \frac{h^2}{24}(f^{(4)}(\xi_1) + f^{(4)}(\xi_2)).$$

Admitindo que $f^{(4)}$ é contínua em $[x_{k-1}, x_{k+1}]$, o Teorema de Bolzano permite concluir que existe um $\xi \in]x_{k-1}, x_{k+1}[$ tal que

$$f^{(4)}(\xi) = \frac{1}{2} \left(f^{(4)}(\xi_1) + f^{(4)}(\xi_2) \right).$$

Assim

$$f''(x_k) = \frac{1}{h^2} [f(x_{k-1}) - 2f(x_k) + f(x_{k+1})] - \frac{h^2}{12} f^{(4)}(\xi). \quad (2.27)$$

Esta fórmula é conhecida como fórmula de diferenças centradas de segunda ordem para aproximar a segunda derivada. Por um raciocínio semelhante poderiam ser obtidas outras fórmulas de diferenças finitas para aproximar a segunda derivada, não só centradas como também progressivas e regressivas.

Exemplo 2.39 Considere, de novo, os valores da função $f(x) = xe^x$ dados na tabela do Exercício 2.38. Aproxime o valor de $f''(2.0) = 29.556224$ usando a fórmula de diferenças finitas centradas de segunda ordem.

Resolução: Temos que

$$f''(2.0) \approx \frac{1}{0.01} [f(1.9) - 2f(2.0) + f(2.1)] = 29.593200.$$

O erro cometido é aproximadamente -3.7×10^{-2} .

2.6.3 Aproximação de derivadas de ordem superior

O estudo efectuado pode ser generalizado para obter fórmulas de diferenças finitas para aproximar derivadas de ordem superior. Essas fórmulas podem ser obtidas quer por interpolação quer recorrendo à série de Taylor. Um algoritmo para obter fórmulas de diferenças finitas de qualquer ordem para aproximar qualquer derivada de uma função pode ser visto em Fornberg (1988).

2.6.4 Exercícios

Exercício 2.6.1 Elabore um algoritmo que permita obter o valor aproximado da primeira derivada de uma função num ponto à custa da fórmula 2.26.

Exercício 2.6.2 Num circuito eléctrico com voltagem aplicada $E(t)$ e indutância L , a primeira Lei de Kirchoff dá-nos a relação

$$E(t) = LI'(t) + RI(t),$$

onde R é a resistência no circuito e $I(t)$ a corrente no instante t . Suponhamos que medimos a corrente para vários valores de $t = t_i$, $i = 1, \dots, 5$, obtendo

t_i	1.00	1.01	1.02	1.03	1.04
$I(t_i)$	3.10	3.12	3.14	3.18	3.24

onde tempo é medido em segundos, a corrente em amperes, a indutância é uma constante dada por $L = 0.98$ henries e a resistência é 0.142 ohms. Aproxime a voltagem E nos valores de t dados na tabela.

Exercício 2.6.3 Mostre, a partir do polinómio interpolador de Lagrange da função f nos pontos x_0 , x_1 e x_2 , tais que $x_1 - x_0 = h$ e $x_2 - x_1 = \alpha h$, que

$$f''(x) \approx \frac{2}{h^2} \left[\frac{f(x_0)}{1 + \alpha} - \frac{f(x_1)}{\alpha} + \frac{f(x_2)}{\alpha(1 + \alpha)} \right].$$

Verifique que quando $\alpha = 1$ se recupera a fórmula das diferenças centrais.

Exercício 2.6.4 Prove que

$$f''(x_k) = \frac{1}{12h^2} [-f(x_{k-2}) + 16f(x_{k-1}) - 30f(x_k) + 16f(x_{k+1}) - f(x_{k+2})] + \frac{h^4}{90} f^{(4)}(\xi),$$

com $\xi \in (x_{k-2}, x_{k+2})$.

Exercício 2.6.5 Prove que:

$$1. f'''(x_k) = \frac{1}{2h^3} [-f(x_{k-2}) + 2f(x_{k-1}) - 2f(x_{k+1}) + f(x_{k+2})] - \frac{h^2}{4} f^{(5)}(\xi_1);$$

$$2. f^{(4)}(x_k) = \frac{1}{h^4} [f(x_{k-2}) - 4f(x_{k-1}) + 6f(x_k) - 4f(x_{k+1}) + f(x_{k+2})] - \frac{h^2}{6} f^{(6)}(\xi_2)$$

Capítulo 3

Interpolação de funções definidas em \mathbb{R}^2

3.1 Introdução

No capítulo anterior tratámos o problema de interpolação para uma função definida num intervalo real e considerámos vários tipos de aproximações dependentes do tipo de regularidade desejada para o polinómio. Neste capítulo iremos considerar a determinação de um polinómio de duas variáveis que seja interpolador de uma função conhecida num conjunto de pontos de \mathbb{R}^2 .

Seja $\Omega = [a, b] \times [c, d]$ um subconjunto de \mathbb{R}^2 . No intervalo $[a, b]$ consideremos a partição

$$a = x_0 < x_1 < \cdots < x_i < \cdots < x_{n-1} < x_n = b$$

e em $[c, d]$

$$c = y_0 < y_1 < \cdots < y_j < \cdots < y_{m-1} < y_m = d.$$

As duas partições anteriores induzem em Ω o seguinte conjunto de pontos

$$\{(x_i, y_j), i = 0, \dots, n, j = 0, \dots, m\}, \quad (3.1)$$

que designamos **rede rectangular**.

Seja f uma função definida em Ω e suponhamos que f é conhecida nos pontos da rede rectangular definida. O nosso objectivo é determinar um polinómio P de duas variáveis que verifique as condições

$$P(x_i, y_j) = f(x_i, y_j), \quad i = 0, \dots, n, j = 0, \dots, m.$$

3.2 Polinómio interpolador de Lagrange

Por polinómio em x e y queremos designar uma função do tipo

$$\sum_{i=0}^n \sum_{j=0}^m b_{ij} x^i y^j, \quad (x, y) \in \mathbb{R}^2.$$

As condições a que o polinómio P deve verificar são $(n+1) \times (m+1)$ e portanto se pensarmos num polinómio em x e y que permite resolver o problema anterior ele poderá apresentar $(n+1) \times (m+1)$ coeficientes.

Consideremos as funções $\ell_i(x)$ e $\ell_j(y)$ definidas considerando as partições em $[a, b]$ e $[c, d]$. Com estas funções definamos $\ell_{ij}(x, y) = \ell_i(x)\ell_j(y)$, $i = 0, \dots, n$, $j = 0, \dots, m$, que verificam a

$$\ell_{ij}(x_k, y_t) = \delta_{kt} = \begin{cases} 1 & k = t \\ 0 & k \neq t \end{cases}.$$

O papel desempenhados por estas funções é análogo ao das funções da base canónica considerada quando introduzimos o polinómio interpolador de Lagrange para uma função definida num intervalo real. Consideremos o seguinte polinómio de grau n em x e m em y

$$\begin{aligned} P(x, y) &= \sum_{i=0}^n \sum_{j=0}^m f(x_i, y_j) \ell_{ij}(x, y) \\ &= f(x_0, y_0) + f(x_1, y_0) \ell_1(x) + \dots + f(x_n, y_0) \ell_n(x) \\ &+ f(x_0, y_1) \ell_1(y) + f(x_1, y_1) \ell_1(x) \ell_1(y) + \dots + f(x_n, y_1) \ell_n(x) \ell_1(y) \\ &+ \dots + f(x_0, y_m) \ell_m(y) + f(x_1, y_m) \ell_1(x) \ell_m(y) + \dots + f(x_n, y_m) \ell_n(x) \ell_m(y) \end{aligned}$$

Atendendo às propriedades das funções $\ell_{ij}(x, y)$, o polinómio $P(x, y)$ satisfaz a

$$P(x_k, y_t) = f(x_k, y_t),$$

e, além disso, é o único polinómio de grau n em x e m em y que resolve o problema de interpolação que nos propoemos resolver. De facto, seja

$$Q(x, y) = \sum_{i=0}^n \sum_{j=0}^m b_{ij} x^i y^j$$

um polinómio de grau n em x e m em y que verifica a $Q(x_i, y_j) = f(x_i, y_j)$. Fixemos y_t na partição de $[c, d]$ e consideremos o polinómio em x

$$Q(x, y_t) = \sum_{i=0}^n \sum_{j=0}^m b_{ij} x^i y_t^j = \sum_{i=0}^n q_{it} x^i$$

em que

$$q_{it} = \sum_{j=0}^m b_{ij} y_t^j.$$

Atendendo às condições para Q , os coeficientes deste polinómio devem satisfazer a

$$\sum_{i=0}^n q_{it} x_k^i = f(x_k, y_t), \quad k = 0, \dots, n,$$

isto é, o seguinte sistema

$$\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \dots & x_n^n \end{bmatrix} \begin{bmatrix} q_{0t} \\ q_{1t} \\ q_{2t} \\ \vdots \\ q_{nt} \end{bmatrix} = \begin{bmatrix} f(x_0, y_t) \\ f(x_1, y_t) \\ f(x_2, y_t) \\ \vdots \\ f(x_n, y_t) \end{bmatrix}.$$

Atendendo a que este sistema é possível e determinado, existe para cada t , uma única solução $q_{it}, i = 0, \dots, n$.

Finalmente, para os coeficientes b_{ij} e para cada $i = 0, \dots, n$, temos o seguinte sistema

$$\sum_{j=0}^m b_{ij} y_t^j = q_{it}, \quad t = 0, \dots, m,$$

isto é,

$$\begin{bmatrix} 1 & y_0 & y_0^2 & y_0^3 & \dots & y_0^m \\ 1 & y_1 & y_1^2 & y_1^3 & \dots & y_1^m \\ 1 & y_2 & y_2^2 & y_2^3 & \dots & y_2^m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_m & y_m^2 & y_m^3 & \dots & y_m^m \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ \vdots \\ b_{im} \end{bmatrix} = \begin{bmatrix} q_{i0} \\ q_{i1} \\ q_{i2} \\ \vdots \\ q_{im} \end{bmatrix},$$

que é também um sistema possível e determinado. Provámos deste modo a unicidade do polinómio interpolador $P(x, y)$. Este polinómio é designado **polinómio interpolador de Lagange**. Provámos então o seguinte teorema.

Teorema 3.1 *Seja f uma função definida no rectângulo $\Omega = [a, b] \times [c, d]$ onde consideramos a rede rectangular $\{(x_i, y_j), i = 0, \dots, n, j = 0, \dots, m\}$. Dados $f(x_i, y_j), i = 0, \dots, n, j = 0, \dots, m$, o único polinómio $P(x, y)$ de grau n em x e m em y que verifica $P(x_i, y_j) = f(x_i, y_j), i = 0, \dots, n, j = 0, \dots, m$, é o polinómio interpolador de Lagrange*

$$P(x, y) = \sum_{i=0}^n \sum_{j=0}^m f(x_i, y_j) \ell_{ij}(x, y).$$

Exemplo 3.2 Consideremos a função $f(x, y) = \exp -(x^2 + y^2)$ com $(x, y) \in [-2, 2] \times [-2, 2]$. Considere o rectângulo $[0, 0.1] \times [0, 0.12]$. Em cada um dos intervalos $[0, 0.1]$ e $[0, 0.12]$ definamos as partições de pontos igualmente distanciados e de espaçamentos, respectivamente, $h = 0.05$ e $k = 0.04$. Utilizando a rede rectangular induzida no rectângulo $[0, 0.1] \times [0, 0.12]$ pelas partições anteriores definamos o polinómio interpolador de Lagrange de grau 2 em x e 3 em y ,

$$\begin{aligned} P(x, y) = & \left(f(0, 0) \frac{(x - 0.05)(x - 0.1)}{0.005} + f(0.05, 0) \frac{x(x - 0.1)}{-0.0025} \right. \\ & \left. + f(0.1, 0) \frac{(x - 0.05)x}{0.0025} \right) \frac{(y - 0.04)(y - 0.08)(y - 0.12)}{-0.000384} \\ & + \left(f(0, 0.04) \frac{(x - 0.05)(x - 0.1)}{0.005} + f(0.05, 0.04) \frac{x(x - 0.1)}{-0.0025} \right. \\ & \left. + f(0.1, 0.04) \frac{(x - 0.05)x}{0.0025} \right) \frac{y(y - 0.08)(y - 0.12)}{0.000128} \\ & + \left(f(0, 0.08) \frac{(x - 0.05)(x - 0.1)}{0.005} + f(0.05, 0.08) \frac{x(x - 0.1)}{-0.0025} \right. \\ & \left. + f(0.1, 0.08) \frac{(x - 0.05)x}{0.0025} \right) \frac{y(y - 0.04)(y - 0.12)}{-0.000128} \\ & + \left(f(0, 0.12) \frac{(x - 0.05)(x - 0.1)}{0.005} + f(0.05, 0.12) \frac{x(x - 0.1)}{-0.0025} \right. \\ & \left. + f(0.1, 0.12) \frac{(x - 0.05)x}{0.0025} \right) \frac{y(y - 0.08)(y - 0.04)}{-0.000384} \end{aligned}$$

Estimemos seguidamente o erro cometido ao aproximar uma função f definida em $[a, b] \times [c, d]$ pelo seu polinómio interpolador de Lagrange.

Em $[a, b] \times [c, d]$ consideremos a rede rectangular (3.1) e sejam

$$h_x = \max_{i=1, \dots, n} (x_i - x_{i-1}), \quad h_y = \max_{j=1, \dots, m} (y_j - y_{j-1}).$$

Seja $p(x, y)$ o polinómio interpolador de Lagrange de grau n em x e m em y que é interpolador de f na rede rectangular anterior. O nosso objectivo é determinar uma estimativa para

$$\|f - P\|_\infty = \max_{(x,y) \in [a,b] \times [c,d]} |f(x, y) - p(x, y)|.$$

Notemos que

$$\|f - P\|_\infty \leq \|f - \tilde{P}\|_\infty + \|\tilde{P} - P\|_\infty, \quad (3.2)$$

em que

$$\tilde{P}(x, y) = \sum_{i=0}^n f(x_i, y) \ell_i(x).$$

Então,

$$f(x, y) - \tilde{P}(x, y) = \frac{\partial^{n+1} f}{\partial x^{n+1}}(\eta, y) \prod_{i=0}^n (x - x_i),$$

e portanto

$$\|f - \tilde{P}\|_\infty \leq \frac{h_x^{n+1}}{4(n+1)!} \left\| \frac{\partial^{n+1} f}{\partial x^{n+1}} \right\|_\infty.$$

Relativamente à segunda parcela do segundo membro de (3.2), notamos que

$$\begin{aligned} \tilde{P}(x, y) - P(x, y) &= \sum_{i=0}^n (f(x_i, y) - \sum_{j=0}^m f(x_i, y_j) \ell_j(y)) \ell_i(x) \\ &= \frac{1}{4(m+1)!} \prod_{j=0}^m (y - y_j) \sum_{i=0}^n \frac{\partial^{m+1} f}{\partial y^{m+1}}(x_i, \eta_2) \ell_i(x) \\ &= \frac{1}{4(m+1)!} \prod_{j=0}^m (y - y_j) \left[\frac{\partial^{m+1} f}{\partial y^{m+1}}(x_i, \eta_2) + \sum_{i=0}^n \frac{\partial^{m+1} f}{\partial y^{m+1}}(x_i, \eta_2) \ell_i \right. \\ &\quad \left. - \frac{\partial^{m+1} f}{\partial y^{m+1}}(x_i, \eta_2) \right] \\ &= \frac{1}{4(m+1)!} \prod_{j=0}^m (y - y_j) \left[\frac{\partial^{m+1}}{\partial y^{m+1}}(x_i, \eta_2) + \frac{1}{4(n+1)!} \frac{\partial^{(m+1)(n+1)}}{\partial x^{n+1} y^{m+1}} \right] \end{aligned}$$

Logo, temos

$$\|\tilde{P} - P\|_\infty \leq \frac{h_y^{m+1}}{4(m+1)} \left[\left\| \frac{\partial^{m+1}}{\partial y^{m+1}} \right\|_\infty + \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}} \right\|_\infty \right]$$

Substituindo em (3.2) obtemos

$$\|f - P\|_\infty \leq \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{n+1} f}{\partial x^{n+1}} \right\|_\infty + \frac{h_y^{m+1}}{4(m+1)} \left[\left\| \frac{\partial^{m+1}}{\partial y^{m+1}} \right\|_\infty + \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}} \right\|_\infty \right]$$

Provamos o seguinte teorema.

Teorema 3.3 *Seja f uma função que tem, em $[a, b] \times [c, d]$, as seguintes derivadas parciais contínuas*

$$\frac{\partial^{m+1}}{\partial y^{m+1}}, \quad \frac{\partial^{n+1}}{\partial x^{n+1}}, \quad \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}}.$$

Em $[a, b] \times [c, d]$ consideremos a rede rectangular (3.1) e sejam

$$h_x = \max_{i=1, \dots, n} (x_i - x_{i-1}), \quad h_y = \max_{j=1, \dots, m} (y_j - y_{j-1}).$$

Se $P(x, y)$ é o polinómio interpolador de Lagrange de grau n em x e m em y que verifica

$$P(x_i, y_j) = f(x_i, y_j), \quad i = 0, \dots, n, j = 0, \dots, m,$$

então

$$\|f - P\|_\infty \leq \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{n+1} f}{\partial x^{n+1}} \right\|_\infty + \frac{h_y^{m+1}}{4(m+1)} \left[\left\| \frac{\partial^{m+1}}{\partial y^{m+1}} \right\|_\infty + \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}} \right\|_\infty \right]$$

3.3 Interpolação de Lagrange segmentada

A interpolação de funções de duas variáveis, tal como para funções de uma variável real, é particularmente vantajosa quando é feita considerando polinómios de grau baixo. De modo análogo ao caso de funções de uma variável, o polinómio interpolador de Lagrange para funções de duas variáveis pode ser considerado por “segmentos”, isto é, com expressões diferentes em subrectangulos do domínio $[a, b] \times [c, d]$, surgindo de modo natural o chamado polinómio de Lagrange segmentado.

Consideremos em $\Omega = [a, b] \times [c, d]$ a rede rectangular (3.1) em que n e m são divisíveis, respectivamente, por n_1 e m_1 . Seja f uma função definida em Ω e suponhamos que conhecemos $f(x_i, y_j), i = 0, \dots, j = 0, \dots, m$. Chamamos polinómio interpolador de Lagrange segmentado de grau n_1 em x e m_1 em y ao polinómio

$$S(x, y) = \left\{ \begin{array}{ll} P_{0,0}(x, y) & x \in [x_0, x_{n_1}] \times [y_0, y_{m_1}] \\ P_{1,0}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_0, y_{m_1}] \\ \vdots & \vdots \\ P_{\frac{n}{n_1}-1,0}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_0, y_{m_1}] \\ P_{0,1}(x, y) & x \in [x_0, x_{n_1}] \times [y_{m_1}, y_{2m_1}] \\ P_{1,1}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_{m_1}, y_{2m_1}] \\ \vdots & \vdots \\ P_{\frac{n}{n_1}-1,1}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_{m_1}, y_{2m_1}] \\ \vdots & \vdots \\ P_{0,\frac{m}{m_1}-1}(x, y) & x \in [x_0, x_{n_1}] \times [y_{m-m_1}, y_m] \\ P_{1,\frac{m}{m_1}-1}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_{m-m_1}, y_m] \\ \vdots & \vdots \\ P_{\frac{n}{n_1}-1,\frac{m}{m_1}-1}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_{m-m_1}, y_m] \end{array} \right. ,$$

em que $P_{i,j}$ é o polinómio de Lagrange de grau n_1 em x e m_1 em y para a rede rectangular do rectângulo $[x_{in_1}, x_{(i+1)n_1}] \times [y_{jm_1}, y_{(j+1)m_1}]$ para $i = 0, \dots, \frac{n}{n_1} - 1, j = 0, \dots, \frac{m}{m_1} - 1$.

Exemplo 3.4 Considere a função f definida num rectângulo $[a, b] \times [c, d]$ onde consideramos a rede rectangular $\{(x_i, y_j), i = 0, 1, 2, j = 0, 1\}$. Então o polinómio interpolador de Lagrange segmentado linear em x e

em y é dado por

$$S(x, y) = \begin{cases} \left(f(x_0, y_0) \frac{x - x_1}{x_0 - x_1} + f(x_1, y_0) \frac{x - x_0}{x_1 - x_0} \right) \frac{y - y_1}{y_0 - y_1} \\ + \left(f(x_0, y_1) \frac{x - x_1}{x_0 - x_1} + f(x_1, y_1) \frac{x - x_0}{x_1 - x_0} \right) \frac{y - y_0}{y_1 - y_0} & (x, y) \in [x_0, x_1] \times [y_0, y_1] \\ \left(f(x_1, y_0) \frac{x - x_2}{x_1 - x_2} + f(x_2, y_0) \frac{x - x_1}{x_2 - x_1} \right) \frac{y - y_1}{y_0 - y_1} \\ + \left(f(x_1, y_1) \frac{x - x_2}{x_1 - x_2} + f(x_2, y_1) \frac{x - x_1}{x_2 - x_1} \right) \frac{y - y_0}{y_1 - y_0} & (x, y) \in [x_1, x_2] \times [y_0, y_1] \\ \left(f(x_1, y_1) \frac{x - x_2}{x_1 - x_2} + f(x_2, y_1) \frac{x - x_1}{x_2 - x_1} \right) \frac{y - y_2}{y_1 - y_2} \\ + \left(f(x_1, y_2) \frac{x - x_2}{x_1 - x_2} + f(x_2, y_2) \frac{x - x_1}{x_2 - x_1} \right) \frac{y - y_1}{y_2 - y_1} & (x, y) \in [x_1, x_2] \times [y_1, y_2] \end{cases}$$

A determinação de uma estimativa para o erro cometido ao aproximar uma função pelo seu polinómio interpolador de Lagrange segmentado é feita de modo análogo caso unidimensional.

Seja $S(x, y)$ o polinómio interpolador de Lagrange segmentado de grau n_1 em x e m_1 em y para uma função f definida num rectângulo $\Omega = [a, b] \times [c, d]$ onde consideramos a rede rectangular (3.1) em que n e m são divisíveis, respectivamente, por n_1 e m_1 . Atendendo a que

$$\|f - S\|_\infty = \max_{i=0, \dots, \frac{n}{n_1}-1, j=0, \dots, \frac{m}{m_1}-1} \|f - P_{i,j}\|_\infty \quad (3.3)$$

e, pelo Teorema 3.3, vale a seguinte estimativa,

$$\|f - P_{i,j}\|_\infty \leq \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{n+1} f}{\partial x^{n+1}} \right\|_{\infty, i, j} + \frac{h_y^{m+1}}{4(m+1)} \left[\left\| \frac{\partial^{m+1} f}{\partial y^{m+1}} \right\|_{\infty, i, j} + \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}} \right\|_{\infty, i, j} \right]$$

Considerando a última desigualdade em (3.3) obtemos a seguinte es-

timativa

$$\begin{aligned} \|f - S\|_\infty \leq & \max_{i=0, \dots, \frac{n}{n_1}-1, j=0, \dots, \frac{m}{m_1}-1} \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{n+1} f}{\partial x^{n+1}} \right\|_{\infty, i, j} \\ & + \frac{h_y^{m+1}}{4(m+1)} \left[\left\| \frac{\partial^{m+1}}{\partial y^{m+1}} \right\|_{\infty, i, j} + \frac{h_x^{n+1}}{4(n+1)} \left\| \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}} \right\|_{\infty, i, j} \right] \end{aligned} \quad (3.4)$$

Demonstrámos o seguinte resultado.

Teorema 3.5 *Seja f uma função que tem, em $[a, b] \times [c, d]$, as seguintes derivadas parciais contínuas*

$$\frac{\partial^{m+1}}{\partial y^{m+1}}, \quad \frac{\partial^{n+1}}{\partial x^{n+1}}, \quad \frac{\partial^{(m+1)(n+1)} f}{\partial x^{n+1} \partial y^{m+1}}.$$

Em $[a, b] \times [c, d]$ consideremos a rede rectangular (3.1) e sejam

$$h_x = \max_{i=1, \dots, n} (x_i - x_{i-1}), \quad h_y = \max_{j=1, \dots, m} (y_j - y_{j-1}).$$

Se $S(x, y)$ é o polinómio interpolador de Lagrange segmentado de grau n_1 em x e m_1 em y para uma função f em que n e m são divisíveis, respectivamente, por n_1 e m_1 , então é válida a estimativa (3.4).

3.4 O polinómio interpolador de Hermite

O polinómio interpolador de Lagrange é uma função contínua que coincide com a função e as suas derivadas parciais pode não coincidir com as correspondentes derivadas de f . O polinómio interpolador de Lagrange segmentado não tem, em geral, as derivadas parciais contínuas. Vamos generalizar o polinómio interpolador de Hermite estudado no caso unidimensional para funções de duas variáveis.

Seja f uma função definida num rectângulo $[a, b] \times [c, d]$ onde consideramos uma rede rectangular (3.1). Suponhamos que conhecemos

$$f(x_i, y_j), \quad \frac{\partial f}{\partial x}(x_i, y_j), \quad \frac{\partial f}{\partial y}(x_i, y_j), \quad \frac{\partial^2 f}{\partial x \partial y}(x_i, y_j)$$

nos pontos da rede rectangular. O polinómio

$$\begin{aligned} H(x, y) &= \sum_{i=0}^n \sum_{j=0}^m f(x_i, y_j) h_i(x) h_j(y) \\ &+ \sum_{i=0}^n \sum_{j=0}^m \frac{\partial f}{\partial x}(x_i, y_j) \bar{h}_i(x) h_j(y) + \sum_{i=0}^n \sum_{j=0}^m \frac{\partial f}{\partial y}(x_i, y_j) h_i(x) \bar{h}_j(y) \\ &+ \sum_{i=0}^n \sum_{j=0}^m \frac{\partial^2 f}{\partial x \partial y}(x_i, y_j) \bar{h}_i(x) \bar{h}_j(y) \end{aligned}$$

em que

$$\begin{aligned} h_i(x) &= [1 - 2(x - x_i)\ell'_i(x_i)]\ell_i(x)^2, \\ h_j(y) &= [1 - 2(y - y_j)\ell'_j(y_j)]\ell_j(y)^2, \\ \bar{h}_i(x) &= (x - x_i)\ell_i(x)^2, \\ \bar{h}_j(y) &= (y - y_j)\ell_j(y)^2, \end{aligned}$$

e

$$\ell_i(x) = \prod_{k=0, k \neq i}^n \frac{x - x_k}{x_i - x_k}, \quad \ell_j(y) = \prod_{k=0, k \neq j}^m \frac{y - y_k}{y_j - y_k},$$

é designado polinómio interpolador de Hermite de grau $2n + 1$ em x e $2m + 1$ em y . Este polinómio verifica as seguintes propriedades, para $i = 0, 1, \dots, n$ e $j = 0, 1, \dots, m$:

$$\begin{aligned} H(x_i, y_j) &= f(x_i, y_j), \\ \frac{\partial H}{\partial x}(x_i, y_j) &= \frac{\partial f}{\partial x}(x_i, y_j), \\ \frac{\partial H}{\partial y}(x_i, y_j) &= \frac{\partial f}{\partial y}(x_i, y_j), \\ \frac{\partial^2 H}{\partial y \partial x}(x_i, y_j) &= \frac{\partial^2 f}{\partial y \partial x}(x_i, y_j). \end{aligned}$$

Tal como para o polinómio interpolador de Lagrange, o polinómio interpolador de Hermite pode ser considerado segmentado. Consideremos em $\Omega = [a, b] \times [c, d]$ a rede rectangular (3.1) em que n e m são divisíveis, respectivamente, por n_1 e m_1 . Seja f uma função definida em Ω e suponhamos que conhecemos

$$f(x_i, y_j), \frac{\partial f}{\partial x}(x_i, y_j), \frac{\partial f}{\partial y}(x_i, y_j), \frac{\partial^2 f}{\partial x \partial y}(x_i, y_j), \quad i = 0, \dots, \quad j = 0, \dots, m.$$

Chamamos polinómio interpolador de Hermite segmentado de grau n_1 em x e m_1 em y ao polinómio

$$S(x, y) = \left\{ \begin{array}{ll} H_{0,0}(x, y) & x \in [x_0, x_{n_1}] \times [y_0, y_{m_1}] \\ H_{1,0}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_0, y_{m_1}] \\ \vdots & \vdots \\ H_{\frac{n}{n_1}-1,0}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_0, y_{m_1}] \\ H_{0,1}(x, y) & x \in [x_0, x_{n_1}] \times [y_{m_1}, y_{2m_1}] \\ H_{1,1}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_{m_1}, y_{2m_1}] \\ \vdots & \vdots \\ H_{\frac{n}{n_1}-1,1}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_{m_1}, y_{2m_1}] \\ \vdots & \vdots \\ H_{0,\frac{m}{m_1}-1}(x, y) & x \in [x_0, x_{n_1}] \times [y_{m-m_1}, y_m] \\ H_{1,\frac{m}{m_1}-1}(x, y) & x \in [x_{n_1}, x_{2n_1}] \times [y_{m-m_1}, y_m] \\ \vdots & \vdots \\ H_{\frac{n}{n_1}-1,\frac{m}{m_1}-1}(x, y) & x \in [x_{n-n_1}, x_n] \times [y_{m-m_1}, y_m] \end{array} \right. ,$$

em que $H_{i,j}$ é o polinómio interpolador de Hermite de grau $2n_1 + 1$ em x e $2m_1 + 1$ em y para a rede rectangular definida no rectângulo $[x_{in_1}, x_{(i+1)n_1}] \times [y_{jm_1}, y_{(j+1)m_1}]$ para $i = 0, \dots, \frac{n}{n_1} - 1, j = 0, \dots, \frac{m}{m_1} - 1$.

Capítulo 4

Interpolação de curvas paramétricas

4.1 Interpolação paramétrica

Nenhuma das técnicas usadas no Capítulo 1 permite gerar curvas que não possam ser expressas como uma função real de variável real. Por exemplo, a curva expressa na Figura 4.1 não é o gráfico de nenhuma função uma vez que, por exemplo, para a abscissa $x = 0$ correspondem três pontos sobre a curva.

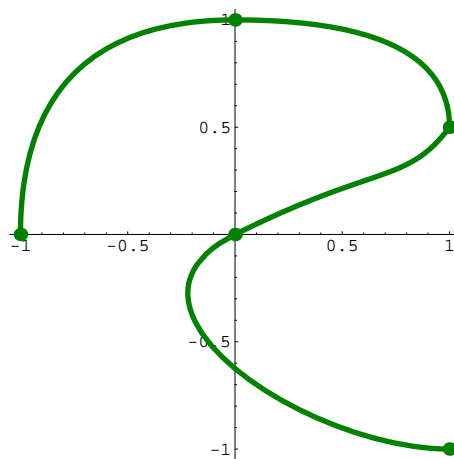


Figura 4.1: Curva paramétrica.

A forma de contornar o problema consiste em escrever a curva \mathcal{C} , de coordenadas (x, y) , na sua forma paramétrica

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}, \quad t \in [t_0, T].$$

Poderemos considerar, sem perda de generalidade, $[t_0, T] = [0, 1]$.

A aproximação da curva que une os pontos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, por esta ordem, é feita considerando uma partição do intervalo $[0, 1]$, geralmente uniforme,

$$0 = t_0 < t_1 < \dots < t_{n-1} < t_n = 1,$$

tal que

$$\begin{cases} x_i = x(t_i) \\ y_i = y(t_i) \end{cases}, \quad i = 0, 1, \dots, n,$$

e calculando os polinômios (globais ou segmentados) P e Q tais que $P(t_i) = x(t_i)$ e $Q(t_i) = y(t_i)$, $i = 0, 1, \dots, n$.

Exemplo 4.1 Usando interpolação de Lagrange, determine uma aproximação para a curva dada na Figura 4.1.

Resolução: Com os dados da figura e considerando a partição $t_i = 0.25i$, $i = 0, 1, 3, 4$, podemos construir a tabela:

i	0	1	2	3	4
t_i	0	0.25	0.5	0.75	1
x_i	-1	0	1	0	1
y_i	0	1	0.5	0	-1

A curva de equação paramétrica $\begin{cases} x = x(t) \\ y = y(t) \end{cases}$, $t \in [0, 1]$, pode

ser aproximada pela curva de equação paramétrica $\begin{cases} x = P(t) \\ y = Q(t) \end{cases}$, $t \in$

$[0, 1]$, com P e Q calculados usando interpolação de Lagrange. Iremos calcular apenas o polinômio P deixando o cálculo de Q ao cuidado do aluno.

Começemos por considerar a seguinte tabela de diferenças finitas.

t_i	$P(t_i)$	$\Delta P(t_i)$	$\Delta^2 P(t_i)$	$\Delta^3 P(t_i)$	$\Delta^4 P(t_i)$
0	-1				
		1			
0.25	0		0		
		1		-2	
0.5	1		-2		6
		-1		4	
0.75	0		2		
		1			
1	1				

Assim, temos que

$$p(t) = -1 + 4t(1 + 16(t - 0.25)(t - 0.5)(-1/3 + (t - 0.75))).$$

4.2 Interpolação cúbica de Hermite segmentada

Em muitas situações (em computação gráfica, por exemplo), é necessário gerar rapidamente curvas suaves que possam ser modificadas de forma fácil. Além disso, a alteração de uma porção dessas curvas não deve afectar as restantes partes da curva.

Uma das escolhas mais usuais para gerar curvas paramétricas são os polinómios cúbicos de Hermite segmentados. Como foi visto, cada porção desses polinómios é determinada pela especificação dos seus pontos extremos e das derivadas (tangentes à curva) nesses pontos.

Suponhamos que queremos calcular a curva que passa nos pontos $P_i = (x_i, y_i)$, $i = 0, 1, \dots, n$, usando polinómios cúbicos de Hermite segmentados. Para isso teremos que, para cada ponto P_i , especificar os valores de $T_i = (x'(t_i), y'(t_i))$, $i = 0, 1, \dots, n$, onde $x_i = x(t_i)$ e $y_i = y(t_i)$.

Vamos simplificar o problema. Pretendemos determinar a curva (polinómio cúbico) que passe pelos pontos $P_0 = (x_0, y_0) = (x(t_0), y(t_0))$ e $P_1 = (x_1, y_1) = (x(t_1), y(t_1))$ e tenha tangentes (inclinações)

$$\frac{dy}{dx}(0) = \alpha_0 \quad \text{em } (x_0, y_0),$$

e

$$\frac{dy}{dx}(1) = \alpha_1 \quad \text{em } (x_1, y_1).$$

Esta curva é chamada curva de Ferguson-Coons. Como

$$\frac{dy}{dx}(0) = \frac{y'(0)}{x'(0)} \quad \text{e} \quad \frac{dy}{dx}(1) = \frac{y'(1)}{x'(1)}$$

temos que α_0 e α_1 não são alterados se multiplicarmos $x'(0)$ e $y'(0)$ pelo mesmo parâmetro λ e $x'(1)$ e $y'(1)$ pelo mesmo parâmetro μ . A multiplicação por um parâmetro não altera as tangentes à curva mas altera a sua forma. Iremos ver, com um exemplo, que, à medida que o parâmetro λ cresce a curva aproxima-se da tangente à curva em P_0 (o mesmo para μ e P_1).

Exemplo 4.2 Determinar a curva que passa pelos pontos $P_0 = (x_0, y_0)$ e $P_1 = (x_1, y_1)$ com tangentes $\alpha_0 = \alpha_1 = 1$.

Resolução: Podemos construir a tabela:

t	0	1
x	x_0	x_1
y	y_0	y_1
x'	λ	$-\mu$
y'	λ	$-\mu$

A curva que iremos calcular é da forma $\begin{cases} x = x(t) \\ y = y(t) \end{cases}, t \in [0, 1]$. Iremos calcular apenas o polinómio x deixando o cálculo de y ao cuidado do aluno.

Começemos por considerar a seguinte tabela de diferenças finitas.

t_i	$x(t_i)$	$x[\cdot, \cdot]$	$x[\cdot, \cdot, \cdot]$	$x[\cdot, \cdot, \cdot, \cdot]$
0	x_0			
		λ		
0	x_0		$x_1 - x_0 - \lambda$	
		$x_1 - x_0$		$\lambda - \mu + 2(x_0 - x_1)$
1	x_1		$-\mu - x_1 + x_0$	
		$-\mu$		
1	x_1			

Assim, temos que

$$x(t) = x_0 + t(\lambda + t(x_1 - x_0 - \lambda + (\lambda - \mu + 2(x_0 - x_1))(t - 1))).$$

4.3 Curvas de Bézier

Consideremos, de novo, o problema de calcular a curva de Ferguson-Coons que passa pelos pontos $P_0 = (x_0, y_0) = (x(t_0), y(t_0))$ e $P_1 = (x_1, y_1) = (x(t_1), y(t_1))$ sendo dados os valores

$$\alpha = \frac{y'(0)}{x'(0)} \quad \text{e} \quad \alpha_1 = \frac{y'(1)}{x'(1)}.$$

A especificação dos valores dos declives α_0 e α_1 pode não ser óbvia. Em computação gráfica essa especificação é feita, de forma simples, através de dois pontos auxiliares (pontos de guia), tal como indicado na Figura 4.2.

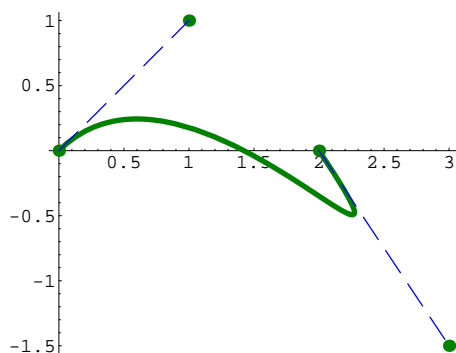


Figura 4.2: Curva de Ferguson-Coons obtida à custa de dois pontos de guia.

Assim, dados dois pontos auxiliares P_2 e P_3 , vamos considerar

$$\begin{cases} T_0 = \lambda(P_2 - P_0) \\ T_1 = \mu(P_1 - P_3) \end{cases} \Leftrightarrow \begin{cases} (x'(0), y'(0)) = \lambda(x_2 - x_0, y_2 - y_0) \\ (x'(1), y'(1)) = \mu(x_1 - x_3, y_1 - y_3) \end{cases},$$

com λ e μ factores de normalização. No que se segue vamos considerar $\lambda = \mu$. Assim, a curva de Ferguson-Coons é dada por

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}, \quad t \in [0, 1],$$

com (prove)

$$x(t) = x_0 + t(\lambda(x_2 - x_0) + t(x_1 - x_0 - \lambda(x_2 - x_0)) + (t-1)(\lambda(-x_3 + x_2 + x_1 - x_0) - 2(x_1 - x_0)))$$

e

$$y(t) = y_0 + t(\lambda(y_2 - y_0) + t(y_1 - y_0 - \lambda(y_2 - y_0)) + (t-1)(\lambda(-y_3 + y_2 + y_1 - y_0) - 2(y_1 - y_0)))$$

Uma propriedade importante para o controlo da curva de Ferguson-Coons reside no facto dessa curva estar contida no invólucro convexo definido pelos pontos P_0 , P_1 , P_2 e P_3 se e só se $\lambda \in [0, 3]$.

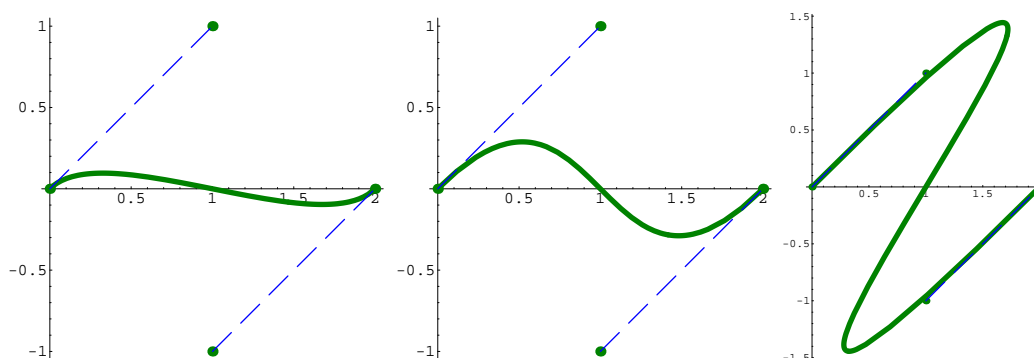


Figura 4.3: Gráficos das curvas de Ferguson-Coons com $\lambda = 1, 3, 15$.

O invólucro convexo de um conjunto de pontos é definido como sendo o menor convexo que contém esses pontos. Sabe-se que um ponto P pertence ao invólucro convexo definido pelos pontos P_0, \dots, P_n se e só se P puder ser escrito como uma combinação linear convexa desses pontos, isto é, se

$$P = \sum_{i=0}^n \phi_i P_i,$$

com $\phi_i \geq 0$, $i = 0, \dots, n$, e $\sum_{i=0}^n \phi_i = 1$. Tal facto resulta imediatamente dos Exercícios 4.4.2 e 4.4.3. Quando $\lambda = 3$ a curva de Ferguson-Coons também se chama curva de Bézier.

4.4 Exercícios

Exercício 4.4.1 Considere a carta dada na Figura 4.4. Usando curvas de Bézier:

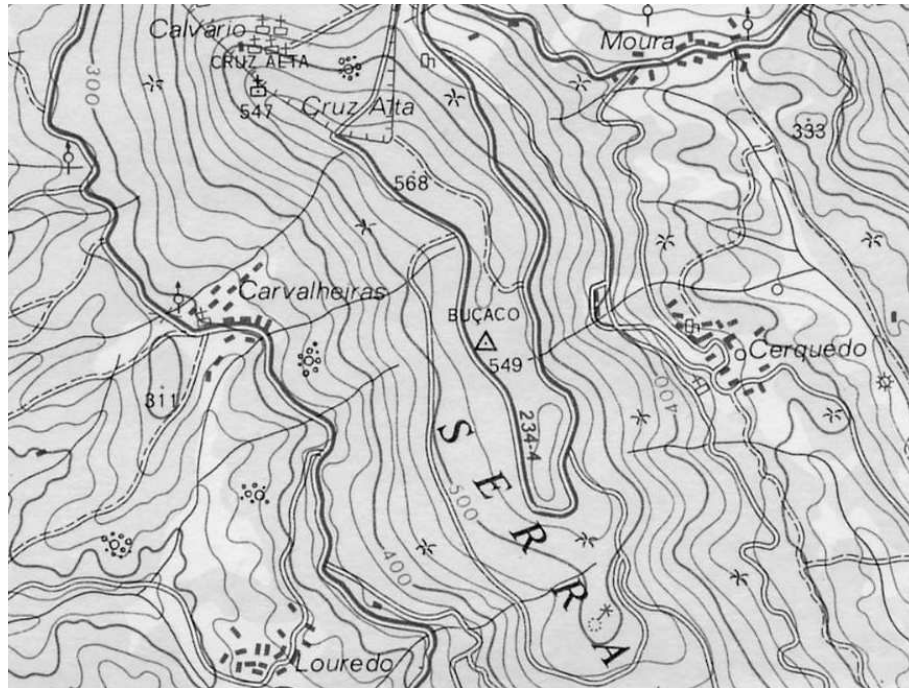


Figura 4.4: Carta topográfica.

1. aproxime a estrada Cruz Alta - Buçaco desde o seu início até ao quarto entroncamento;
2. aproxime a curva de nível de 500 m.

Exercício 4.4.2 Mostre que a curva de Ferguson-Coons se pode escrever na forma

$$\begin{cases} x(t) = \phi_0(t)x_0 + \phi_1(t)x_1 + \phi_2(t)x_2 + \phi_3(t)x_3 \\ y(t) = \phi_0(t)y_0 + \phi_1(t)y_1 + \phi_2(t)y_2 + \phi_3(t)y_3 \end{cases}, \quad t \in [0, 1],$$

com

$$\begin{aligned} \phi_0(t) &= 2t^3 - \lambda t^3 - 3t^2 + 2\lambda t^2 - \lambda t + 1, \\ \phi_1(t) &= -2t^3 + \lambda t^3 - \lambda t^2 + 3t^2, \\ \phi_2(t) &= \lambda t^3 - 2\lambda t^2 + \lambda t, \\ \phi_3(t) &= -\lambda t^3 + \lambda t^2. \end{aligned}$$

Exercício 4.4.3 Mostre que, dada a curva de Ferguson-Coons tal como no exercício anterior, $\phi_i(t) \geq 0$, $i = 0, 1, 2, 3$, $t \in [0, 1]$, se e só se $\lambda \in [0, 3]$.

Capítulo 5

Observações e mínimos quadrados

5.1 Introdução

Consideremos uma determinada entidade que pretendemos medir. Para determinar a medida em questão fazemos, em geral, várias observações ou medições e portanto somos conduzidos a uma diversidade de valores para a medida dessa entidade. A variabilidade dos resultados da medição são originados por vários factores e portanto os valores obtidos, são em geral redundantes. Deveremos possuir um critério que nos permita a partir dos valores observados, determinar um valor único para a medida da entidade. O critério que iremos estudar é o critério dos mínimos quadrados e, ao usar tal critério, diremos que estamos a proceder ao ajustamento dos valores observados.

Neste capítulo os termos “observação” e “medição” irão ser usados com o mesmo significado. As medições são caracterizadas pelo seguinte conjunto de propriedades:

1. o acto de medir/observar, que designamos por medição/observação, significa sempre a realização de uma operação física ou uma série de operações;
2. o resultado numérico obtido por medição é usado para representar a medida do objecto e transporta consigo todas as circunstâncias do processo pelo qual foi obtido;
3. as medições são sempre efectuadas com a ajuda de objectos (excepto a operação de contagem simples);
4. ao medir estamos a estabelecer uma comparação com um padrão estabelecido por convenção vindo, portanto, o resultado da medida expresso em unidades;

5. as medições referem-se a conceitos teóricos ou abstrações geométricas (tais como comprimento, área, amplitude de um ângulo, etc) que são usadas para caracterizar uma determinada entidade mas não têm significado físico;
6. o resultado numérico obtido por medição só tem significado quando está associado ao conceito teórico a que a observação diz respeito.

Consideremos, por exemplo, um triângulo para o qual pretendemos determinar a medida da sua área. Para o efeito deveremos determinar a medida do comprimento da base e da altura. A medição é efectuada com o auxílio de uma régua. Ao dizermos que o lado mede 0.10 m estamos a estabelecer uma comparação do comprimento do lado com o metro padrão. Dizer que o lado mede 0.10 m é um abuso de linguagem; deveríamos ser precisos e afirmar: “a medida do comprimento do lado é 0.10 m”.

5.2 Modelo matemático e modelo estocástico

Consideremos um rectângulo $[ABCD]$ para o qual pretendemos determinar a medida da área. Sabemos que a medida da área é dada por $\ell_A = \ell_{\overline{AB}}\ell_{\overline{BC}}$, sendo $\ell_{\overline{AB}}$ e $\ell_{\overline{BC}}$, respectivamente, as medidas dos lados \overline{AB} e \overline{BC} . Atendendo ao seu carácter aleatório, cada observação realizada é encarada como uma concretização de uma variável aleatória. Assim teremos as variáveis aleatórias (v.a.) X_A , $X_{\overline{AB}}$ e $X_{\overline{BC}}$, que estão definidas no universo $\Omega = \{\text{observações}\}$, também chamado **espaço fundamental** ou **conjunto dos acontecimentos**, e que assumem valores reais. As duas últimas v.a. são concretizadas, por exemplo, n e m vezes respectivamente, obtendo-se duas amostras. Estas amostras permitem definir duas variáveis estocásticas (ou estatísticas)

$$x_{\overline{AB}} : \{O_1, \dots, O_n\} \in \Omega \longrightarrow \{x_{\overline{AB},1}, \dots, x_{\overline{AB},n}\}$$

e

$$x_{\overline{BC}} : \{O_1, \dots, O_m\} \in \Omega \longrightarrow \{x_{\overline{BC},1}, \dots, x_{\overline{BC},m}\}.$$

As observações efectuadas não são mais do que concretizações da mesma v.a.. Atendendo a este facto podemos admitir que:

1. as observações são independentes (umas não influenciam as outras);

2. as observações são efectuadas com igual precisão (com o mesmo instrumento de medida).

Na determinação da medida da área do rectângulo foi definido um **modelo matemático**, isto é, um sistema teórico de conceitos abstractos – medida da área, medida do comprimento dos lados – que descreve uma entidade física – a área de um rectângulo. Neste sistema teórico distinguem-se, claramente, duas partes: a primeira (modelo funcional) é constituída pela expressão matemática envolvendo as variáveis do modelo e descreve as propriedades deterministas da entidade em estudo; a segunda (modelo estocástico) é constituída pelo conjunto de hipóteses que são estabelecidas sobre as observações efectuadas.

O modelo estocástico surge uma vez que as observações são concretizações de variáveis aleatórias. Por definição, diz-se que um fenómeno é estocástico ou conjectural sempre que os seus casos particulares dependem do acaso e a respeito dos quais só é possível enunciar probabilidades.

5.3 Alguns conceitos probabilísticos

5.3.1 Distribuições unidimensionais

As v.a. podem dividir-se em dois grupos: discretas e contínuas. As discretas podem tomar um conjunto discreto de valores $x_1, x_2, \dots, x_n, \dots$, finito ou não. Isto é, se X_d for uma v.a. discreta temos que

$$X_d : \Omega \longrightarrow \{x_1, x_2, \dots, x_n, \dots\}.$$

As v.a. contínuas podem tomar quaisquer valores num intervalo real ou mesmo em toda a recta real. Assim, se X_c for uma v.a. contínua

$$X_c : \Omega \longrightarrow \mathbb{R}.$$

Comecemos por considerar o caso discreto. Seja X_d uma v.a discreta e sejam $x_i, i = 1, 2, \dots$, valores que X_d pode tomar. Seja Ω o universo de X_d . Consideremos os acontecimentos

$$A_i = \{w \in \Omega : X_d(w) = x_i\}, \quad i = 1, 2, \dots,$$

e seja p_i a probabilidade de cada um destes acontecimentos ocorrer, isto é,

$$p_i = P(A_i) = P(X_d = x_i), \quad i = 1, 2, \dots,$$

em que $0 \leq p_i \leq 1$ e $\sum_i p_i = 1$.

Definição 5.1 (Função de probabilidade) *Seja $X_d : \Omega \longrightarrow \{x_1, x_2, \dots, x_n, \dots\}$ uma v.a. com $P(X_d = x_i) = p_i$. A função*

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x \mapsto f(x) = \begin{cases} p_i & x = x_i \\ 0 & x \neq x_i \end{cases}$$

é designada função de probabilidade da v.a. X_d .

No caso contínuo, esta definição não tem significado, sendo substituída pela de função densidade de probabilidade. Porém, antes de definir esse conceito, vamos apresentar a definição de função de repartição (ou de distribuição) que é válida tanto no caso discreto como no caso contínuo.

Definição 5.2 (Função de repartição) *Seja X uma v.a. (discreta ou contínua). A função*

$$F : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x \mapsto F(x) = P(X \leq x)$$

é designada função de distribuição ou função de repartição da v.a. X .

Notamos que, no caso discreto,

$$\begin{aligned} F(x) &= P(X_d \leq x) \\ &= P(\{w \in \Omega : X_d(w) \leq x\}) \\ &= P(\cup_{i:x_i \leq x} \{w \in \Omega : X_d(w) = x_i\}) \\ &= \sum_{i:x_i \leq x} P(\{w \in \Omega : X_d(w) = x_i\}) \\ &= \sum_{i:x_i \leq x} f(x_i) \\ &= \sum_{i:x_i \leq x} p_i \end{aligned}$$

Vamos agora considerar a definição de função densidade de probabilidade. Seja X_c uma v.a. contínua, i.e. $X_c : \Omega \rightarrow \mathbb{R}$ em que Ω denota o espaço fundamental.

Definição 5.3 (Função densidade) Dizemos que $f : \mathbb{R} \rightarrow \mathbb{R}$ é a função densidade (de probabilidade) da v.a. contínua X_c se f é não negativa e $\int_{\mathbb{R}} f(x)dx = 1$. Nesse caso tem-se que

$$F(x) = P(X_c \leq x) = \int_{-\infty}^x f(t)dt.$$

A definição anterior poderia ter sido usada para definir v.a. contínua. Assim, uma v.a. X_c é contínua se a sua função de repartição admitir a representação expressa na definição anterior.

Observação 5.4 Antes de prosseguir, façamos uma pequena reflexão sobre o significado da designação de função densidade introduzida na definição anterior. Considere-se, uma vez que $P(X \in \mathbb{R}) = 1$, uma unidade de probabilidade distribuída (ou repartida) por todo o eixo real de tal forma que a quantidade de probabilidade atribuída ao intervalo $(-\infty, x]$ seja igual a $F(x)$, qualquer que seja x . Nestas condições, a quantidade de probabilidade no intervalo $(x, x+\epsilon]$ é dada pela diferença $F(x+\epsilon) - F(x)$. Logo, o quociente $(F(x+\epsilon) - F(x))/\epsilon$ é a quantidade média de probabilidade nesse intervalo. O limite $\lim_{\epsilon \rightarrow 0} (F(x+\epsilon) - F(x))/\epsilon = F'(x)$, se existir, representa a densidade de probabilidade no ponto x . No entanto, nos pontos onde $F'(x)$ existe tem-se que, pela definição anterior, $F'(x) = f(x)$. Assim, convencionando escrever $f(x) = 0$ nos pontos onde $F'(x)$ não existe tem-se que a função densidade determina univocamente a função de repartição.

O facto de $P(X = x) = F(x) - F(x-0)$, ver Exercício 5.3.1, permite concluir que se a função de repartição F for contínua em todo o $x \in \mathbb{R}$, então $P(X = x) = 0$, i.e., todos os pontos de \mathbb{R} têm probabilidade zero (no entanto $X = x$ não é um acontecimento impossível).

Se a função de repartição tem uma descontinuidade num ponto $x \in \mathbb{R}$, então $P(X = x) = F(x) - F(x-0) > 0$, isto é, o salto de função é a probabilidade da v.a. assumir o valor x .

As funções de densidade e de repartição de uma v.a. são determinadas (e caracterizadas) por um conjunto de parâmetros que são

úteis para a compreensão do seu comportamento. O primeiro desses parâmetros expressa a ideia intuitiva de média e é designado por **esperança** ou **valor esperado** ou ainda **valor mais provável** da v.a.. Este valor será “o mais provável de ocorrer” quando se considera a concretização da v.a. em causa.

Definição 5.5 (Esperança) (i) *Seja X_d uma v.a discreta que pode tomar os valores $x_i, i = 1, 2, \dots$, e seja f a sua função de probabilidade. Chamamos esperança (média, ou valor mais provável) de X_d a*

$$E(X_d) = \sum_i x_i f(x_i).$$

(ii) *Seja X_c uma v.a contínua de função densidade f . Chamamos esperança (média, ou valor mais provável) de X_c a*

$$E(X_c) = \int_{\mathbb{R}} x f(x) dx.$$

Um outro parâmetro importante é a **variância** pois dá-nos informação relativamente à dispersão dos valores da v.a. em torno do seu valor mais provável.

Definição 5.6 (Variância e desvio padrão) *Seja X uma v.a. de esperança $E(X)$. Ao valor de $E[(X - E(X))^2]$ chamamos **variância** da v.a. X que é denotada por σ_X^2 . À raiz quadrada da variância da v.a. X chamamos **desvio padrão** de X e é representado por σ_X .*

Note-se que:

1. se X_d é uma v.a. discreta que pode tomar os valores $x_i, i = 1, 2, \dots$, e f a sua função de probabilidade, então

$$\sigma_{X_d}^2 = \sum_i (x_i - E(X_d))^2 f(x_i);$$

2. se X_c é uma v.a. contínua de função densidade f , então

$$\sigma_{X_c}^2 = \int_{\mathbb{R}} (x - E(X_c))^2 f(x) dx.$$

A variância de uma v.a. também pode ser definida por $\sigma_X^2 = E(X^2) - E(X)^2$. De facto,

$$\sigma_X^2 = E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - E(X)^2.$$

5.3.2 Distribuições bidimensionais

Em inúmeras situações o estudo probabilístico ou estatístico envolve $k > 1$ propriedades ou características qualitativas dos elementos w do espaço fundamental Ω . Nesses casos, faz-se corresponder a cada ponto $w \in \Omega$ um ponto (x_1, \dots, x_k) de \mathbb{R}^k , isto é,

$$w \mapsto (X_1(w), \dots, X_k(w)) := X(w).$$

Quando para todo o ponto (x_1, \dots, x_k) de \mathbb{R}^k o conjunto

$$\{w \in \Omega : X_1(w) = x_i, i = 1, \dots, k\}$$

é um acontecimento, diz-se que $X(w)$ é um **vector aleatório** ou uma **variável aleatória k-dimensional**. Daqui por diante o vector $X(w)$ escreve-se $X = (X_1, \dots, X_k)$.

As v.a. bidimensionais são suficientemente importantes para justificar o seu estudo particular. Neste caso (quando $k = 2$), em vez de (X_1, X_2) escreveremos, na maior parte das vezes, (X, Y) .

Dada uma v.a. bidimensional (X, Y) a probabilidade de se obter um ponto na região do plano \mathbb{R}^2 definido pelas desigualdades $X \leq x$ e $Y \leq y$,

$$P(X \leq x, Y \leq y) = P(\{w \in \Omega : X(w) \leq x, Y(w) \leq y\}),$$

existe sempre (por definição) e permite introduzir uma função real de duas variáveis reais de acordo com a seguinte definição.

Definição 5.7 (Função de repartição) *Seja (X, Y) uma v.a. bidimensional. A função*

$$\begin{aligned} F : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto F(x, y) = P(X \leq x, Y \leq y) \end{aligned}$$

é designada função de repartição (ou distribuição) da v.a. bidimensional (X, Y) .

Sejam $a < b, c < d$ e

$$I = \{w \in \Omega : a < X(w) \leq b, c < Y(w) \leq d\},$$

$$A = \{w \in \Omega : X(w) \leq a, Y(w) \leq c\},$$

$$B = \{w \in \Omega : a < X(w) \leq b, Y(w) \leq c\},$$

$$C = \{w \in \Omega : X(w) \leq a, c < Y(w) \leq d\}.$$

Então

$$F(b, d) = P(I \cup A \cup B \cup C) = P(I) + P(A) + P(B) + P(C).$$

Por outro lado

$$P(A) = F(a, c),$$

$$F(b, c) = P(B) + F(a, c) \implies P(B) = F(b, c) - F(a, c),$$

$$F(a, d) = P(C) + F(a, c) \implies P(C) = F(a, d) - F(a, c).$$

Substituindo obtemos

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(b, c) - F(a, d).$$

Provamos assim o seguinte teorema.

Teorema 5.8 *Seja (X, Y) um v.a. com função de repartição F . Se $a < b$ e $c < d$ então*

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(b, c) - F(a, d).$$

Vamos agora apresentar a noção de função de probabilidade (para v.a. discretas) e a de função densidade (para v.a. contínuas). Seja (X, Y) uma variável que pode tomar os valores $\{(x_i, y_j), i = 1, \dots, j = 1, \dots\}$ e sejam p_{ij} valores não negativos tais que

$$P(X = x_i, Y = y_j) = p_{ij},$$

com $\sum_i \sum_j p_{ij} = 1$.

Definição 5.9 (Função de probabilidade) *Seja (X, Y) uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, \dots, j = 1, \dots\}$ e sejam p_{ij} as probabilidades*

$$P(X = x_i, Y = y_j) = p_{ij}.$$

A função

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$

$$(x, y) \mapsto f(x, y) = \begin{cases} p_{ij} & (x, y) = (x_i, y_j) \\ 0 & (x, y) \neq (x_i, y_j) \end{cases}$$

é designada por função de probabilidade de (X, Y) .

Esta noção permite concluir que, quando se consideram v.a. discretas, a função de repartição é definida por

$$F : \mathbb{R} \longrightarrow \mathbb{R}$$

$$(x, y) \mapsto F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j) \cdot$$

Consideremos agora o caso contínuo e a noção de função densidade.

Definição 5.10 (Função densidade) *Seja (X, Y) um v.a. com função de repartição F . Dizemos que uma função $f(x, y)$ não negativa, integrável e tal que*

$$\int_{\mathbb{R}^2} f(x, y) dx dy = 1$$

é a função densidade da v.a. se

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(t, u) dt du.$$

Se a função de repartição admitir a representação anterior, dizemos que (X, Y) é uma v.a. contínua.

5.3.3 Distribuições marginais

Quando consideramos um vector de duas v.a., (X, Y) , podemos estar interessados em estudar o que se passa com cada uma das variáveis isoladamente. Poderemos pretender

$$P(X \leq x) = P(\{w \in \Omega : X(w) \leq x, Y(w) \text{ qualquer}\}) = P(X \leq x, Y < +\infty).$$

Para isso, vamos definir a função

$$F_1(x) := \lim_{y \rightarrow +\infty} F(x, y)$$

A esta função chamaremos função de repartição marginal da v.a. X . De igual modo, a função

$$F_2(y) := \lim_{x \rightarrow +\infty} F(x, y)$$

é designada por função de repartição marginal de Y .

Consideremos o caso em que (X, Y) é uma v.a. discreta que pode tomar os valores $\{x_{ij}, i = 1, 2, \dots, j = 1, 2, \dots\}$ e seja f a função

de probabilidade desta v.a. bidimensional. Assim, as funções de repartição marginais são dadas por

$$\begin{aligned} F_1(x) = P(X \leq x) &= P(\{w \in \Omega : X(w) \leq x, Y(w) \text{ qualquer}\}) \\ &= P(X \leq x, Y < +\infty) \\ &= \sum_{x_i \leq x} \sum_{y_j} f(x_i, y_j) \end{aligned}$$

e

$$\begin{aligned} F_2(y) = P(Y \leq y) &= P(\{w \in \Omega : X(w) \text{ qualquer}, Y \leq y\}) \\ &= P(X < +\infty, Y \leq y) \\ &= \sum_{x_i} \sum_{y_j \leq y} f(x_i, y_j). \end{aligned}$$

Estas funções podem ainda ser representadas em termos das chamadas funções de probabilidade marginais que são definidas, para o caso da v.a. discreta em causa por

$$f_1(x) = \begin{cases} \sum_{y_j} f(x, y_j) & x = x_i \\ 0 & x \neq x_i \end{cases},$$

designada por função de probabilidade marginal de X , e

$$f_2(y) = \begin{cases} \sum_{x_i} f(x_i, y) & y = y_j \\ 0 & y \neq y_j \end{cases}$$

designada por função de probabilidade marginal de Y . Com estas definições temos que

$$F_1(x) = \sum_{x_i \leq x} f_1(x_i)$$

e

$$F_2(y) = \sum_{y_j \leq y} f_2(y_j).$$

Consideremos agora (X, Y) uma v.a. contínua com f a sua função densidade. Então

$$F_1(x) = \int_{-\infty}^x \int_{\mathbb{R}} f(t, y) dy dt = \lim_{y \rightarrow +\infty} \int_{-\infty}^x \int_{-\infty}^y f(t, u) du dt,$$

e

$$F_2(y) = \int_{-\infty}^y \int_{\mathbb{R}} f(x, t) dt dy = \lim_{x \rightarrow +\infty} \int_{-\infty}^y \int_{-\infty}^x f(t, u) dt du,$$

são as funções de repartição marginal de X e Y respectivamente.

Estas funções podem ainda ser representadas em termos das chamadas funções de densidade marginais que são definidas, para o caso da v.a. contínua em causa por

$$f_1(x) = \frac{d}{dx} F_1(x) = \frac{d}{dx} \int_{-\infty}^x \int_{\mathbb{R}} f(t, y) dy dt = \int_{\mathbb{R}} f(x, y) dy,$$

designada por função de densidade marginal de X , e

$$f_2(y) = \frac{d}{dy} F_2(y) = \int_{\mathbb{R}} f(x, y) dx$$

designada por função de densidade marginal de Y . Com estas definições temos que

$$F_1(x) = \int_{-\infty}^x f_1(t) dt,$$

e

$$F_2(y) = \int_{-\infty}^y f_2(u) du.$$

5.3.4 Independência

Começemos por considerar a seguinte definição.

Definição 5.11 (Independência) *Consideremos o vector aleatório (X, Y) . Dizemos que X é independente de Y se, para todo $(x, y) \in \mathbb{R}^2$, os acontecimentos*

$$A = \{w \in \Omega : X \leq x, Y < +\infty\}, \quad e \quad B = \{w \in \Omega : X < +\infty, Y \leq y\}$$

forem independentes, isto é, se $P(A \cap B) = P(A)P(B)$.

Seja F a função de repartição conjunta da uma v.a. (X, Y) . Então, se X e Y forem independentes, tem-se que, para $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= P(\{X \leq x, Y < +\infty\} \cap \{X < +\infty, Y \leq y\}) \\ &= P(\{X \leq x, Y < +\infty\})P(\{X < +\infty, Y \leq y\}) \\ &= F_1(x)F_2(y) \end{aligned}$$

e portanto a função de distribuição conjunta é o produto das distribuições marginais. Reciprocamente, se a distribuição conjunta é o produto das funções distribuições marginais então as v.a. X e Y são v.a independentes. Provamos assim o seguinte teorema.

Teorema 5.12 *Sejam X e Y duas v.a. e seja F a função de distribuição conjunta do vector aleatório (X, Y) . As v.a. X e Y são independentes se e só se*

$$F(x, y) = F_1(x)F_2(y), \quad (x, y) \in \mathbb{R}^2,$$

em que F_1 e F_2 são as função de distribuição marginal de cada uma das componentes.

Vamos agora demonstrar duas condições necessárias e suficientes para a independência de v.a., uma para o caso discreto e outra para o caso contínuo.

Teorema 5.13 *Seja (X, Y) uma v.a. discreta que podem tomar os valores*

$$\{(x_i, y_j), i = 1, 2, \dots, j = 1, \dots\}.$$

As v.a. X e Y são independentes se e só se

$$f(x_i, y_j) = f_1(x_i)f_2(y_j),$$

em que f é a função de probabilidade conjunta do vector aleatório (X, Y) e f_1, f_2 são as funções de probabilidade marginais de X e Y respectivamente.

Demonstração: Suponhamos que X e Y são independentes. Então

$$P(X = x_i, Y = y_j) = P(X = x_i, Y < +\infty)P(X < \infty, Y = y_j),$$

ou seja

$$f(x_i, y_j) = f_1(x_i)f_2(y_j).$$

Suponhamos agora que $f(x_i, y_j) = f_1(x_i)f_2(y_j)$ e provemos que X e Y são independentes. De facto,

$$\begin{aligned}
 P(X \leq x_i, Y \leq y_j) &= \sum_{\ell=1}^i \sum_{k=1}^j P(X = x_\ell, Y = y_k) \\
 &= \sum_{\ell=1}^i \sum_{k=1}^j f(x_\ell, y_k) \\
 &= \sum_{\ell=1}^i \sum_{k=1}^j f_1(x_\ell) f_2(y_k) \\
 &= \left(\sum_{\ell=1}^i f_1(x_\ell) \right) \left(\sum_{k=1}^j f_2(y_k) \right) \\
 &= P(X \leq x_i, Y < +\infty) P(X < +\infty, Y \leq y_j).
 \end{aligned}$$

o que prova o pretendido. \square

Teorema 5.14 *Seja (X, Y) um vector aleatório contínuo e seja f a sua função densidade. Sejam f_1 e f_2 as funções densidade marginais. As v.a., X e Y são independentes se e só se*

$$f(x, y) = f_1(x)f_2(y), \quad (x, y) \in \mathbb{R}^2.$$

Demonstração: Suponhamos que X e Y são independentes. Então

$$F(x, y) = F_1(x)F_2(y)$$

e portanto

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial x} F_1(x) \frac{\partial}{\partial y} F_2(y)$$

o que permite concluir que

$$f(x, y) = f_1(x)f_2(y).$$

Suponhamos, reciprocamente, que as funções densidade verificam $f(x, y) = f_1(x)f_2(y)$. Então

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(t, u) dt du \\ &= \int_{-\infty}^x \int_{-\infty}^y f_1(t) f_2(u) dt du \\ &= \int_{-\infty}^x f_1(t) dt \int_{-\infty}^y f_2(u) du \\ &= F_1(x) F_2(y) \end{aligned}$$

e portanto as v.a. X e Y são independentes. \square

5.3.5 Funções de variáveis aleatórias

Seja X uma v.a. com função de repartição F_X e consideremos $g : \mathbb{R} \rightarrow \mathbb{R}$. Admita-se que $Y = g(X)$ é a v.a. que assume o valor $y = g(x)$ quando $X = x$. Pretende-se determinar a função de repartição F_Y da v.a. Y . Vamos começar por considerar o seguinte exemplo.

Exemplo 5.15 Consideremos dois casos.

1. Seja $g(x) = ax + b$. Então

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right), \quad \text{se } a > 0. \end{aligned}$$

Se $a < 0$ então

$$F_Y(y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a} - 0\right).$$

2. Seja $g(x) = x^2$. Então

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y} - 0), \quad \text{se } y > 0. \end{aligned}$$

e para $y \leq 0$, tem-se $F_Y(y) = 0$.

No caso geral, se X for uma v.a. contínua temos que

$$F_Y(y) = P(Y \leq y) = P(X \in g^{-1}(] - \infty, y])),$$

onde $g^{-1}(] - \infty, y])$ representa a imagem recíproca por g do intervalo $] - \infty, y]$. Nos casos mais correntes, $g^{-1}(] - \infty, y])$ é um intervalo ou uma união de intervalos donde resulta que a probabilidade do último membro se pode exprimir facilmente em termos de F_X .

Se a v.a. for discreta, a v.a. $Y = g(X)$, com g uma função real de variável real, é também discreta. Neste caso, o nosso objectivo pode ser atingido se determinarmos a função de probabilidade f_Y da v.a. Y a partir da função de probabilidade da v.a. X , que denotamos por f_X . Assim, se X for uma v.a. discreta que pode tomar os valores $D = \{x_i, i = 1, 2, \dots\}$, a v.a. $Y = g(X)$ toma valores $g(D) = \{y_i = g(x_i), i = 1, 2, \dots\}$. Então a função de probabilidade da v.a. Y é dada por

$$f_Y(y_j) = P(Y = y_j) = P(g(X) = y_j) = \sum_{x_i: g(x_i)=y_j} f_X(x_i)$$

e, para $y \neq y_j$, $f_Y(y) = 0$.

O que poderemos dizer quanto à esperança da v.a. $Y = g(X)$? No caso discreto, se X é uma v.a. que pode tomar os valores $\{x_i, i = 1, 2, \dots\}$ e f é a sua função de probabilidade, então a esperança de Y é

$$E(Y) = \sum_{x_i} g(x_i) f(x_i).$$

No caso contínuo, se f for a função densidade de X , a esperança de Y é

$$E(Y) = \int_{\mathbb{R}} g(x) f(x) dx.$$

5.3.6 Esperança matemática de um vector aleatório

Definição 5.16 (Esperança) *Seja (X, Y) uma v.a. bidimensional, $E(X)$ a esperança de X e $E(Y)$ a esperança de Y . Ao vector $(E(X), E(Y))$ chamamos esperança matemática ou valor esperado do vector aleatório (X, Y) e que é denotado por $E(X, Y)$.*

Se (X, Y) uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, 2, \dots, j = 1, 2, \dots\}$ e f é a sua função de probabilidade conjunta então

$$\begin{aligned} E(X, Y) &= \left(\sum_{x_i} x_i f_1(x_i), \sum_{y_j} y_j f_2(y_j) \right) \\ &= \left(\sum_{x_i} \sum_{y_j} x_i f(x_i, y_j), \sum_{x_i} \sum_{y_j} y_j f(x_i, y_j) \right). \end{aligned}$$

Se (X, Y) uma v.a. contínua e f é a sua função densidade conjunta então

$$\begin{aligned} E(X, Y) &= \left(\int_{\mathbb{R}} x f_1(x) dx, \int_{\mathbb{R}} y f_2(y) dy \right) \\ &= \left(\int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) dy dx, \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) dx dy \right). \end{aligned}$$

Consideremos agora a função g , real de duas variáveis reais, e a v.a. $Z = g(X, Y)$ que assume o valor $z = g(x, y)$ quando $(X, Y) = (x, y)$. Temos então que

$$E(Z) = \int_{\mathbb{R}^2} g(x, y) f(x, y) dx dy.$$

Se (X, Y) é uma v.a. discreta que pode tomar os valores $\{(x_i, y_j), i = 1, 2, \dots, j = 1, \dots\}$ e f é a função de probabilidade conjunta então

$$E(Z) = \sum_{x_i} \sum_{y_j} g(x_i, y_j) f(x_i, y_j).$$

No caso contínuo temos que, se f é a função densidade da v.a. X , então

$$E(Z) = \int_{\mathbb{R}} g(x, y) f(x, y) dx dy.$$

Vamos demonstrar alguns resultados importantes.

Teorema 5.17 *Seja (X, Y) uma v.a. e $Z = X + Y$. Então $E(Z) = E(X) + E(Y)$.*

Demonstração: Vamos fazer a demonstração apenas para o caso contínuo. O caso discreto fica como exercício. Temos, sucessivamente, que

$$\begin{aligned} E(X + Y) &= \int_{\mathbb{R}^2} (x + y)f(x, y)dxdy \\ &= \int_{\mathbb{R}^2} xf(x, y)dxdy + \int_{\mathbb{R}^2} yf(x, y)dxdy \\ &= E(X) + E(Y), \end{aligned}$$

o que prova o resultado. \square

Teorema 5.18 *Seja (X, Y) uma v.a. e $Z = XY$. Então, se X e Y forem independentes $E(Z) = E(X)E(Y)$.*

Demonstração: Vamos fazer a demonstração apenas para o caso contínuo. O caso discreto fica como exercício. Temos, sucessivamente, que

$$E(XY) = \int_{\mathbb{R}^2} xyf(x, y)dxdy.$$

Como X e Y são independentes, temos que

$$\int_{\mathbb{R}^2} xyf(x, y)dxdy = \int_{\mathbb{R}^2} xf_1(x)f_2(y)dxdy.$$

Logo $E(X, Y) = E(X)E(Y)$, o que prova o pretendido. \square

5.3.7 Variância, covariância e coeficiente de correlação

Sejam X e Y duas v.a.. Se X é independente de Y então $E(XY) = E(X)E(Y)$. Consideremos a diferença

$$\sigma_{XY} = E(XY) - E(X)E(Y).$$

Esta diferença é nula se as v.a. são independentes (o recíproco pode não ser verdadeiro, i.e. pode suceder que a diferença seja nula sem que haja independência das v.a.). Temos ainda que

$$\sigma_{XY} = E((X - E(X))(Y - E(Y))).$$

Consideremos um referencial com origem no ponto $(E(X), E(Y))$ e a quantidade definida por $(x - E(X))(y - E(Y))$. Esta quantidade

é positiva se x e y estiverem no primeiro e terceiro quadrantes e é negativa se x e y estiverem no segundo e quarto. Assim, se a probabilidade de ocorrerem valores de X e Y acima das respectivas médias ou abaixo das médias é grande então σ_{XY} é positivo e tem um valor elevado. Por outro lado, se a probabilidade de ocorrerem valores de X acima (ou abaixo) da média e de Y abaixo (ou acima) da média é grande então σ_{XY} é negativo mas é, em valor absoluto, grande. Deste modo concluímos que o coeficiente σ_{XY} mede a dependência das v.a. bem como o seu desvio em relação à média.

Definição 5.19 (Covariância) *Sejam X e Y duas v.a. de esperanças $E(X)$ e $E(Y)$. O coeficiente*

$$\sigma_{XY} = E[(X - E(X))(Y - E(Y))]$$

é designado por covariância entre X e Y e também é denotado por $\text{cov}(X, Y)$.

Vamos demonstrar o seguinte resultado.

Teorema 5.20 *Sejam X e Y duas v.a. e $Z = X + Y$. Então*

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}.$$

Demonstração: Temos, sucessivamente, que

$$\begin{aligned} \sigma_Z^2 &= E[Z - E(Z)]^2 \\ &= E[(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))] \\ &= E[(X - E(X))^2] + E[(Y - E(Y))^2] + 2E[(X - E(X))(Y - E(Y))] \\ &= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}, \end{aligned}$$

o que prova o pretendido. \square

À matriz

$$\begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

chamamos **matriz de covariância** da v.a. (X, Y) .

Os conceitos apresentados para v.a. bidimensionais poderiam ser generalizados para v.a. de dimensão $n \in \mathbb{N}$. Vamos apenas, para já, generalizar o conceito de matriz de covariância.

Definição 5.21 Consideremos n v.a. X_i e $i = 1, 2, \dots, n$. A matriz

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \sigma_{X_1X_3} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \sigma_{X_2X_3} & \cdots & \sigma_{X_2X_n} \\ \sigma_{X_3X_1} & \sigma_{X_3X_2} & \sigma_{X_3}^2 & \cdots & \sigma_{X_3X_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \sigma_{X_nX_3} & \cdots & \sigma_{X_n}^2 \end{bmatrix}$$

é designada por matriz de covariância do vector aleatório (X_1, \dots, X_n) .

Para finalizar, vamos definir um conceito de grande interesse prático. Consideremos a quantidade

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

em que σ_{XY} , σ_X e σ_Y são, respectivamente, a covariância de (X, Y) e os desvios padrão de X e Y , respectivamente. A ρ_{XY} chamaremos coeficiente de correlação da v.a. (X, Y) . Como é evidente, se as v.a. X e Y forem independentes $\sigma_{XY} = \rho_{XY} = 0$. Neste caso, e sempre que o coeficiente de correlação for nulo, dizemos que as variáveis X e Y são não correlacionadas. Por outro lado,

$$|\rho_{XY}| \leq 1.$$

De facto, uma vez que, pela desigualdade de Hölder,

$$\sigma_{XY} = E((X - E(X))(Y - E(Y))) \leq [E((X - E(X))^2)]^{1/2} [E((Y - E(Y))^2)]^{1/2} = \sigma_X \sigma_Y.$$

Temos que ρ_{XY} é máximo, em valor absoluto, quando $\sigma_{XY} = \sigma_X \sigma_Y$. Neste caso (quando o coeficiente de correlação $\rho_{XY} = \pm 1$, ou próximo desse valor) dizemos que X e Y são fortemente correlacionadas.

5.3.8 Exercícios

Exercício 5.3.1 Seja F uma função de repartição. Mostre que os seguintes resultados.

1. Para todo o $x \in \mathbb{R}$ tem-se que $0 \leq F(x) \leq 1$.
2. A função F é não decrescente.
3. $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$ e $F(+\infty) := \lim_{x \rightarrow +\infty} F(x) = 1$.

4. Para valores de a e b finitos tais que $a < b$, tem-se que $P(a < X \leq b) = F(b) - F(a)$.
5. A função F é contínua à direita, i.e., $F(x+0) := \lim_{\epsilon \rightarrow 0^+} F(x+\epsilon) = F(x)$.
6. $P(X = x) = F(x) - \lim_{\epsilon \rightarrow 0^-} F(x+\epsilon) := F(x) - F(x-0)$.

Exercício 5.3.2 Seja F uma função de repartição. Mostre que os seguintes resultados.

1. $P(X \leq x) = F(x)$ (por definição).
2. $P(X < x) = F(x-0)$.
3. $P(X > x) = 1 - F(x)$.
4. $P(X \geq x) = 1 - F(x-0)$.
5. $P(a < X \leq b) = F(b) - F(a)$ (exercício anterior).
6. $P(a < X < b) = F(b-0) - F(a)$.
7. $P(a \leq X < b) = F(b-0) - F(a-0)$.
8. $P(a \leq X \leq b) = F(b) - F(a-0)$.

Exercício 5.3.3 Um determinado estabelecimento comercial tem capacidade de vender entre zero e quatro teodolitos num mês. Seja X a v.a. que indica o número de teodolitos vendidos num mês e

k	0	1	2	3	4
$P(X = k)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{8}$

1. Determine a função de distribuição da v.a. X .
2. Qual probabilidade do estabelecimento vender entre 1 e 3 teodolitos num mês?

Exercício 5.3.4 A duração, em milhares de horas, da componente de um tipo de radar, é uma v.a. X com função densidade f tal que $f(x) = 0.1e^{-0.1x}$, se $x > 0$, e $f(x) = 0$, caso contrário. Pretende-se determinar a probabilidade para que uma componente escolhida ao acaso: dure menos de 4×10^3 horas; dure entre 5×10^3 e 10×10^3 horas; dure mais de 15×10^3 horas.

Exercício 5.3.5 Seja (X, Y) uma v.a. cuja função de probabilidade é dada na tabela

x	y	2	4
1		0.2	0.3
2		0.1	0.1
3		0.2	0.1

1. Determine a função de distribuição da v.a. (X, Y) .
2. Calcule $P(X \leq 3, Y \leq 4)$, $P(X \leq 3, 2 < Y \leq 4)$ e $P(1 < X \leq 3, 2 < Y \leq 4)$.

Exercício 5.3.6 Seja (X, Y) uma v.a. que tem por função densidade

$$f(x, y) = \begin{cases} e^{-(x+y)}, & (x, y) \in]0, +\infty[\times]0, +\infty[, \\ 0, & \text{caso contrário.} \end{cases}$$

1. Determine a função de distribuição da v.a. (X, Y) .
2. Calcule $P(X \leq 3, Y \leq 4)$, $P(X \leq 3, 2 < Y \leq 4)$ e $P(1 < X \leq 3, 2 < Y \leq 4)$.

Exercício 5.3.7 Seja (X, Y) uma v.a. cuja função de probabilidade é dada na tabela

x	y	0	1
0		$\frac{1}{3}$	$\frac{4}{15}$
1		$\frac{4}{15}$	$\frac{2}{15}$

1. Determine a função de probabilidade das distribuições marginais de X e Y .
2. Determine a função de distribuição marginal de X e Y .

Exercício 5.3.8 Vai ser atribuído a uma família, através de um sorteio, um apartamento num edifício com três andares, havendo dois tipos de apartamentos em cada andar (tipo A e B). Pretendendo o dono do edifício que a referida família fique, de preferência, com um no primeiro andar e do tipo A , viciou o sorteio de modo que a função probabilidade fosse a seguinte:

	1 ^a andar	2 ^a andar	3 ^a andar
apartamento do tipo A	$\frac{20}{30}$	$\frac{1}{30}$	$\frac{1}{30}$
apartamento do tipo B	$\frac{6}{30}$	$\frac{1}{30}$	$\frac{1}{30}$

1. Determine as distribuições marginais das variáveis em questão.
2. Qual é a probabilidade de sair um apartamento do tipo A? E qual é a probabilidade de sair o primeiro andar?
3. Diga se as variáveis são dependentes ou independentes.

Exercício 5.3.9 Seja (X, Y) uma v.a. e g e h duas funções reais de duas variáveis reais. Mostre que

$$E(\alpha g(X, Y) + \beta h(X, Y)) = \alpha E(g(X, Y)) + \beta E(h(X, Y)), \quad \alpha, \beta \in \mathbb{R}.$$

Exercício 5.3.10 Considere a v.a. (X, Y) que tem por função densidade

$$f(x, y) = \begin{cases} x + y, & (x, y) \in]0, 1[\times]0, 1[, \\ 0, & \text{caso contrário.} \end{cases}$$

1. Determine a esperança do vector aleatório (X, Y) .
2. Determine $E(X)$, $E(X^2)$, σ_X^2 e σ_{XY} .
3. Diga se v.a. X e Y são ou não independentes.

5.4 Alguns conceitos estatísticos

5.4.1 Introdução

A estatística é a ciência que se preocupa com a colecta, análise, interpretação e apresentação dos dados e tem como objectivo fundamental o estudo de uma população. Pode dividir-se em duas disciplinas: a estatística descritiva que se preocupa com a colecta, análise, interpretação dos dados estatísticos; e a estatística indutiva, também chamada amostral ou inferencial, que é aquela que, partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões, fundamentando-se na teoria das probabilidades.

Pode dizer-se que os objectivos da teoria das probabilidades e da estatística indutiva são, de certo modo, inversos: na primeira, parte-se de um determinado esquema ou modelo e procura calcular-se a

probabilidade de certos resultados ou observações; na segunda, parte-se dos resultados ou observações e procura saber-se alguma coisa sobre o esquema ou modelo.

O nosso objectivo nesta secção será, dada uma v.a. X , calcular a sua média e matriz de covariância. Pelo que foi visto, estes parâmetros só podem ser determinados se conhecermos a sua função de probabilidade (no caso discreto) ou a função densidade (no caso contínuo). Não sendo o caso, o processo usado para tal fim consiste em considerar várias concretizações da v.a. X e construir o que chamamos uma amostra de tamanho relevante. Esta amostra permite construir uma variável estatística e o seu estudo permite inferir uma caracterização para a v.a. a que a amostra diz respeito.

5.4.2 Distribuição de frequências

Consideremos uma v.a. X e uma amostra de tamanho n relativamente grande. Sejam x_i , $i = 1, 2, \dots, n$, os valores observados, i.e., seja $A = \{x_i : i = 1, \dots, n\}$ uma amostra. A frequência absoluta de x_i , $F(x_i)$, é o número de vezes que x_i ocorre nos n valores observados na amostra A . A frequência relativa de x_i , $F_r(x_i)$, é igual ao quociente entre a frequência absoluta e o tamanho da amostra, i.e.

$$F_r(x_i) = \frac{F(x_i)}{n}, \quad i = 1, \dots, n.$$

A distribuição das frequências relativas permite determinar aproximações para a probabilidade de um determinado acontecimento. Suponhamos, por exemplo, que pretendemos calcular $P(X = x_i)$. Este valor é aproximado por $F_r(x_i)$. Se pretendemos calcular $P(X \leq x)$ então tomamos para este valor a aproximação $\sum_{x_i \leq x} F_r(x_i)$.

Suponhamos agora que possuímos uma v.a. bidimensional (X, Y) e uma amostra de tamanho $n + m$ $\{(x_i, y_j), i = 1, \dots, n, j = 1, \dots, m\}$. Sejam $F_r(x_i, y_j)$ as frequências relativas de (x_i, y_j) , para todo o i e todo o j . Tal como no caso unidimensional, $P(X = x_i, Y = y_j) \approx F_r(x_i, y_j)$ e $P(X \leq x, Y \leq y) \approx \sum_{x_i \leq x} \sum_{y_j \leq y} F_r(x_i, y_j)$. No entanto, para calcular $P(X = x_i, Y \in \mathbb{R})$ ou $P(X \in \mathbb{R}, Y = y_j)$ necessitamos de determinar as frequências relativas marginais. Assim, para $i =$

$1, \dots, n$, temos que

$$F_{r,x}(x_i) = F_{r,1}(x_i) = \sum_{j=1}^m F_r(x_i, y_j)$$

é a frequência relativa marginal de x_i e, para $j = 1, \dots, m$,

$$F_{r,y}(y_j) = F_{r,2}(y_j) = \sum_{i=1}^n F_r(x_i, y_j)$$

é a frequência relativa marginal de y_j . Então:

- $P(X = x_i, Y = y_j) \approx F_r(x_i, y_j)$;
- $P(X \leq x, Y \leq y) \approx \sum_{x_i \leq x} \sum_{y_j \leq y} F_r(x_i, y_j)$;
- $P(X = x_i, Y \in \mathbb{R}) \approx F_{r,1}(x_i)$;
- $P(X \in \mathbb{R}, Y = y_j) \approx F_{r,2}(y_j)$;
- $P(X \leq x, Y \in \mathbb{R}) \approx \sum_{x_i \leq x} F_{r,1}(x_i)$;
- $P(X_2 \leq y) \approx \sum_{y_j \leq y} F_{r,2}(y_j)$.

Os dados da amostra podem ser classificados de acordo com alguma propriedade dando assim origem a classes ou categorias. Por exemplo, na contagem das alturas da população portuguesa pode ser conveniente agrupar os dados em classes com, por exemplo, cinco centímetros de amplitude.

A frequência absoluta de cada classe é igual ao número de elementos dessa classe. A frequência relativa de uma classe é o quociente entre a frequência absoluta da classe e o número total de observações. A distribuição de frequências relativas para dados classificados pode ser dada graficamente por um gráfico de barras onde no eixo das abscissas se consideram as classes e no eixo das ordenadas as frequências relativas da classe. Este gráfico é designado por **histograma**.

Suponhamos agora que temos uma amostra de uma v.a. bidimensional (X, Y) ,

$$A = \{(x_i, y_j), i = 1, \dots, n, j = 1, \dots, m\}$$

e que os dados estão agrupados por classes

$$[x_{a_1}, x_{a_2}[, [x_{a_2}, x_{a_3}[, \dots , [x_{a_p}, x_{a_{p+1}}[,$$

$$[y_{b_1}, y_{b_2}[, [y_{b_2}, y_{b_3}[, \dots , [y_{b_q}, y_{b_{q+1}}[.$$

Seja

$$P_{a_i, b_j} = F_r([x_{a_i}, x_{a_{i+1}}[\times [y_{b_j}, y_{b_{j+1}}[.$$

Tal como no caso unidimensional, a distribuição de frequências relativas das diversas classes pode ser dada pelo histograma.

Suponhamos que pretendemos determinar a frequência relativa (marginal) da classe $[x_{a_i}, x_{a_{i+1}}[$ ou da classe $[y_{b_j}, y_{b_{j+1}}[$. Temos então que definir frequências relativas marginais. Assim

$$F_{r,1}([x_{a_i}, x_{a_{i+1}}[) := F_r([x_{a_i}, x_{a_{i+1}}[\times \mathbb{R}) = \sum_j p_{a_i, b_j}$$

é designada por frequência relativa marginal da classe $[x_{a_i}, x_{a_{i+1}}[$ e

$$F_{r,2}([y_{b_j}, y_{b_{j+1}}[) := F_r([y_{b_j}, y_{b_{j+1}}[\times \mathbb{R}) = \sum_i p_{a_i, b_j}$$

é designada por frequência relativa marginal da classe $[y_{b_j}, y_{b_{j+1}}[$.

Podemos assim determinar a distribuição de frequências relativas de uma classe $[x_{a_i}, x_{a_{i+1}}[$ sabendo que o valor observado para Y está na classe $[y_{b_j}, y_{b_{j+1}}[$. Esta distribuição de frequências é designada por distribuição de frequência relativa condicionada e pode ser usada para medir o grau de dependência das v.a.. Esta distribuição pode ser determinada considerando apenas a coluna da distribuição de frequências relativa em que j é fixo.

5.4.3 Média, variância, covariância e coeficiente de correlação de uma amostra

Seja X uma v.a. e consideremos a amostra de tamanho n , $\{x_i, i = 1, \dots, n\}$. A

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

chamamos média (aritmética) da amostra.

Vejam se este valor é ou não uma boa aproximação para a esperança da v.a. X . Consideremos as n v.a. $X_i, i = 1, \dots, n$, com a mesma distribuição de X e, com estas v.a., definamos

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Como é evidente, esta nova v.a. pode tomar como valor a média da amostra \bar{x} . Mais ainda,

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

e, portanto, \bar{X} é um estimador cêntrico para $E(X)$ e, como tal, qualquer valor que esta variável assumir pode ser tomado como estimativa para $E(X)$. Por este motivo podemos dizer que \bar{X} é também um estimador consistente para $E(X)$ uma vez que, para $\epsilon > 0$, arbitrariamente pequeno,

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - E(X)| \leq \epsilon) = 1$$

e, portanto, quando $n \rightarrow +\infty$, a probabilidade de \bar{X} tomar um valor na proximidade de $E(X)$ é aproximadamente igual a 1. Como \bar{x} é um valor que \bar{X} pode tomar, para n suficientemente grande, \bar{x} é uma boa aproximação de $E(X)$.

Outro parâmetro importante a calcular é variância de uma amostra. Para isso, vamos definir a variância (corrigida) da amostra na forma

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

onde \bar{x} é a média da amostra e n o seu tamanho.

A razão de usarmos s_x^2 em vez de

$$\bar{s}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

pode ser justificada se mostrarmos que a esperança da v.a.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

com \bar{X} o estimador cêntrico da esperança e X_i , $i = 1, \dots, n$, uma v.a. com a mesma distribuição de X , é igual à variância de X , isto é, $E(S_x^2) = \sigma_X^2$. Neste caso dizemos que S_x^2 é um estimador cêntrico da variância de X e, deste modo, que a variância da amostra s_x^2 é uma estimativa cêntrica σ_X^2 .

Temos sucessivamente

$$\begin{aligned}
 E(S_x^2) &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
 &= \frac{1}{n-1} E\left[\sum_{i=1}^n ((X_i - E(X)) - (\bar{X} - E(X)))^2\right] \\
 &= \frac{1}{n-1} \sum_{i=1}^n E[(X_i - E(X))^2] - 2E[(X_i - E(X))(\bar{X} - E(X))] \\
 &\quad + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{2}{n-1} E\left[\sum_{i=1}^n X_i - nE(X)\right](\bar{X} - E(X)) \\
 &\quad + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{2n}{n-1} E[\bar{X} - E(X)]^2 + \frac{n}{n-1} E[(\bar{X} - E(X))^2] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \sigma_{X_i}^2 - \frac{n}{n-1} E[\bar{X} - E(X)]^2 \\
 &= \frac{n}{n-1} \sigma_X^2 - \frac{n}{n-1} E[(\bar{X} - E(X))^2]
 \end{aligned}$$

e, atendendo a que as v.a. são independentes,

$$\begin{aligned}
 E[(\bar{X} - E(X))^2] &= E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - E(X)\right)^2\right] \\
 &= \frac{1}{n^2} \sum_{i=1}^n E[(X_i - E(X))^2] \\
 &\quad + \frac{1}{n^2} \sum_{i \neq j} E[(X_i - E(X))(X_j - E(X))] \\
 &= \frac{1}{n} \sigma_X^2
 \end{aligned}$$

e portanto

$$E(S_x^2) = \sigma_X^2.$$

Observação 5.22 *A estimativa \bar{s}_x^2 , embora sendo uma estimativa consistente com a variância da v.a. a que a amostra diz respeito, não é uma estimativa cêntrica para este parâmetro.*

Consideremos agora uma v.a. bidimensional (X, Y) e, para esta v.a., uma amostra $A = \{(x_i, y_i) : i = 1, \dots, n\}$. As médias \bar{x} e \bar{y} são estimativas cênticas, respectivamente, para $E(X)$ e $E(Y)$. Mais ainda, as variâncias das v.a. marginais X e Y são estimadas considerando s_x^2 e s_y^2 . Atendendo a que se trata de um vector aleatório poderemos estar interessados estudar a covariância existente entre X e Y devendo-se para este efeito calcular σ_{XY} .

Consideremos a v.a.

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

em que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Exercício 5.4.1 Prove que S_{XY} é um estimador cêntrico para a covariância entre X e Y , isto é, prove que $E(S_{XY}) = \sigma_{XY}$.

Como a v.a. S_{XY} pode tomar o valor

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

concluimos que s_{xy} é uma estimativa cêntrica para a covariância e portanto este valor é uma boa estimativa para aquele parâmetro. A s_{xy} chamamos **covariância da amostra**.

Finalmente, temos que uma estimativa cêntrica para o coeficiente de correlação ρ_{XY} pode ser dado por

$$\frac{s_{xy}}{s_x s_y}$$

sendo s_{xy} a covariância da amostra A e s_x e s_y são, respectivamente, os desvio padrão das amostras $\{x_i : i = 1, \dots, n\}$ e $\{y_i : i = 1, \dots, n\}$.

5.4.4 Exercícios

Exercício 5.4.2 Mediu-se várias vezes a altura de um poste com uma fita de aço obtendo-se os seguintes resultados:

Nº da medição	1	2	3	4	5	6	7	8	9
Valor obtido	31.29	31.24	31.27	31.26	31.36	31.25	31.26	31.27	31.28

Determine a média da amostra recolhida bem como a sua variância e desvio padrão.

Exercício 5.4.3 Determine a capacidade esperada de um depósito cilíndrico, sabendo que se fizeram várias medições, todas nas mesmas condições e independentes, da sua altura h e do raio da base r , tendo-se obtido os seguintes resultados (em metros):

altura h	10.30	10.33	10.31
raio da base r	3.62	3.60	3.64

Exercício 5.4.4 Para determinar o comprimento de uma linha poligonal mediram-se independentemente os seus três segmentos tendo-se obtido os seguinte resultados (em metros):

x_i	1.00	1.12	1.01	1.25	1.10
y_i	3.00	3.36	3.03	3.75	3.50
z_i	4.00	4.36	4.03	4.75	4.50

Determine uma estimativa para o comprimento da linha poligonal e para o seu desvio padrão.

5.5 Propriedades dos erros de observação

5.5.1 Os tipos de erros de observação

Quando se efectua uma medição/observação podemos distinguir 3 fases: calibragem do objecto usado na medição; colocação do objecto em posição dese efectuar a medição; observação do valor marcado no objecto. Associados a este processo aparecem, inevitavelmente, erros. As teoria clássicas apresentam os erros como sendo de três tipos.

1. Erros acidentais ou ocasionais. Têm origem em acidentes ou descuidos aquando da medição. Por exemplo, apontar ao alvo errado,

engano na leitura do resultado, etc. São erros, em geral, de valor significativo e podem ser minimizados atendendo a cuidados simples que podem passar por: planejar as observações com cuidado; efectuar várias medições; usar técnicas simples de validação dos resultados (lógica e bom-senso); testar (e conhecer bem) o material utilizado; repetir a experiência com técnicas diferentes.

2. **Erros sistemáticos.** São os erros que se repetem sempre que a medição é efectuada nas mesmas condições físicas. Não podem, por isso, ser detectados pela repetição sucessiva de experiências. Pode dizer-se que quando existem apenas erros sistemáticos não há nada de errado no processo de observação. Estes erros podem ter origem no observador (problemas de visão, etc), na má calibragem do objecto usado na medição (cada vez menos frequentes), nas condições físicas ou na má escolha do modelo. Neste último caso, o modelo pode ser “corrigido” no sentido de se ter em conta os erros. Para isso, os estudos efectuados na disciplina de Análise Numérica I poderão ser úteis.
3. **Erros aleatórios.** São os erros que permanecem quando são eliminados os erros pertencentes aos dois primeiros grupos. São de pequena amplitude e têm origem desconhecida. Apresentam um comportamento análogo ao das v.a..

Como já foi dito, ao efectuar uma medição, o nosso objectivo é determinar um valor que caracteriza a entidade em estudo. Atendendo ao carácter aleatório da observação, a medição que se realiza é a concretização de uma v.a. que está associada ao conceito teórico que se pretende quantificar. Uma vez que o valor pretendido (valor mais provável da v.a.) é desconhecido realizamos, em geral, várias observações e construímos uma amostra. A média da amostra é uma boa estimativa para a esperança da v.a. e a variância da amostra é um bom indicador do erro cometido ao substituir o valor mais provável da v.a. pela média da amostra.

Como é evidente, a construção da amostra é fundamental para a determinação das estimativas anteriores. A construção dessas amostras é feita realizando um conjunto de observações que podem ser de três tipos.

1. **Directas.** O conceito teórico em estudo a que está associada uma v.a. é quantificado por observação directa da entidade a que o conceito se refere (por exemplo, determinação da amplitude de um ângulo, determinação da medida do comprimento de um lado de um triângulo, etc.).
2. **Indirectas.** O conceito teórico que se pretende quantificar está associada uma v.a. que é função de outras v.a.. Os conceitos teóricos que estão subjacentes a estas últimas v.a. são quantificados por observação directa. Assim obtemos para o primeiro conceito teórico um valor determinado por observação indirecta (por exemplo, mede-se o valor da velocidade e determina-se o deslocamento.).
3. **Condicionadas.** Sobre as observações realizadas existem relações de compatibilidade que são traduzidas, em geral, por relações matemáticas. Assim, estas observações são condicionadas pela referida relação matemática (por exemplo, observação da amplitude dos ângulos internos de um triângulo existe uma relação: a soma das amplitudes é igual a 2π .).

5.6 Precisão: cofactor e peso

Seja $X = (X_1, \dots, X_n)$ uma v.a.. Em geral, a sua média $E(X)$ é tomada em substituição de qualquer valor que a v.a. possa tomar. Ao efectuar essa aproximação, a matriz de covariância (ou a variância, no caso unidimensional) dá-nos um indicador do erro cometido. O conhecimento das estimativas para os erros cometidos é importante para avaliar a precisão do resultado e também para proceder a ajustamentos.

A variância de uma v.a. e a covariância entre duas v.a. são valores que têm significado relativo pois dependem da v.a. a que dizem respeito. Atendendo a este facto, quando se consideram ajustamentos, é geralmente introduzida uma normalização dessas quantidades que é efectuada à custa de uma quantidade denominada **variância de referência**, que será denotada por σ_0^2 (a raiz quadrada da variância de referência é chamada **desvio padrão de referência**). Surge assim a noção cofactor.

Consideremos um vector aleatório (X_1, \dots, X_n) e seja σ_{ij} a covariância entre X_i e X_j (se $i = j$ então consideramos a variância). Seja σ_0^2 a variância de referência. A

$$q_{ij} = \frac{\sigma_{ij}}{\sigma_0^2}$$

chamamos cofactor entre X_i e X_j e à matriz

$$Q = \frac{1}{\sigma_0^2} C,$$

sendo $C = [\sigma_{ij}]$ a matriz de covariância, chamamos matriz cofactor. Se a matriz cofactor for invertível, a sua inversa é designada matriz peso e é denotada por W , isto é,

$$W = [w_{ij}] = Q^{-1} = \sigma_0^2 C^{-1}.$$

Observação 5.23 *Atendendo às definições anteriores, podemos fazer as observações seguintes.*

1. *O termo “peso” deve-se ao facto de, em geral, um peso “grande” significar “grande precisão” e um peso “pequeno” significar “pouca precisão”. Este facto só é totalmente verdade se as matrizes peso e cofactor forem diagonais.*
2. *Se nada é dito em contrário, tomamos a variância de referência igual a 1.*
3. *Atendendo a que a matriz de covariância é uma matriz simétrica então a matriz cofactor e a matriz peso (se existir) são também simétricas.*
4. *Se as v.a. são não correlacionadas então a matriz de covariância é uma matriz diagonal e, como tal, existe a matriz peso, que é também uma matriz diagonal, de elementos da diagonal principal iguais ao produto da variância de referência pelo inverso da variância de cada v.a., isto é,*

$$w_{ii} = \frac{\sigma_0^2}{\sigma_{ii}}.$$

5. *Se as v.a. têm coeficientes de correlação iguais a ± 1 então a covariância é calculada facilmente (basta calcular os desvio padrão) e portanto facilmente podemos determinar a matriz cofactor e matriz peso.*

5.7 Propagação da média, da variância e covariância

Seja $X = (X_1, \dots, X_n)$ uma v.a. para a qual se conhecem as suas propriedades estocásticas (função densidade ou função de probabilidade) e $Y = (Y_1, \dots, Y_m)$ uma v.a. que depende de X pela relação funcional $Y = g(X)$. O nosso objectivo consiste em determinar as propriedades estocásticas da v.a. Y .

Em aplicações práticas este problema surge, geralmente, de forma mais simplificada. Nesses casos, consideram-se três situações: propagação das médias, isto é, determinar $E(Y)$ conhecendo $E(X)$; propagação dos erros aleatórios, isto é, determinar a matriz de covariância de Y sabendo a matriz de covariância de X ; propagação dos erros sistemáticos, situação estudada em Análise Numérica I com recurso à fórmula de propagação do erro. Vamos apenas estudar os dois primeiros casos.

5.7.1 Propagação das médias

Sabemos que, se $Y = g(X)$ então

$$E(y) = E(g(Y)) \Leftrightarrow E(Y_i) = E(g_i(X_1, \dots, X_n)), \quad i = 1, \dots, m.$$

Temos então (no caso contínuo)

$$E(Y_i) = \int_{\mathbb{R}^n} g_i(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

sendo f a função densidade de $X = (X_1, \dots, X_m)$.

No caso de g ser uma função linear a esperança de Y resulta imediatamente. De facto, consideremos $Y = AX + b$, com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Assim sendo, como a esperança é uma funcional linear,

$$Y = AX + b \Leftrightarrow E(Y) = AE(X) + b$$

e então

$$E(Y_i) = \sum_{j=1}^n a_{ij} E(X_j) + b_i, \quad i = 1, \dots, m.$$

Consideremos agora o caso não linear. Iremos particularizar a apresentação para o caso em que $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$, isto é, $m = n = 2$. A generalização para o caso geral é imediata.

Quando a função $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ é não linear, procede-se à sua “linearização” usando polinómios de Taylor. Como se sabe, dada uma função $y = g(x)$ o seu polinómio de Taylor de grau um em torno de $\bar{x} = (\bar{x}_1, \bar{x}_2)$ é dado por

$$\begin{cases} y_1 \approx g_1(\bar{x}) + \frac{\partial g_1}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \frac{\partial g_1}{\partial x_2}(\bar{x})(x_2 - \bar{x}_2) \\ y_2 \approx g_2(\bar{x}) + \frac{\partial g_2}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \frac{\partial g_2}{\partial x_2}(\bar{x})(x_2 - \bar{x}_2) \end{cases}.$$

Temos então, no caso de considerarmos v.a.

$$\begin{cases} Y_1 \approx g_1(E(X)) + \frac{\partial g_1}{\partial x_1}(E(X))(X_1 - E(X_1)) + \frac{\partial g_1}{\partial x_2}(E(X))(X_2 - E(X_2)) \\ Y_2 \approx g_2(E(X)) + \frac{\partial g_2}{\partial x_1}(E(X))(X_1 - E(X_1)) + \frac{\partial g_2}{\partial x_2}(E(X))(X_2 - E(X_2)) \end{cases}.$$

Temos então que

$$Y \approx g(E(X)) + J(X - E(X)),$$

onde J é a matriz de Jacobi de g calculada no ponto \bar{x} , isto é,

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(E(X)) & \frac{\partial g_1}{\partial x_2}(E(X)) \\ \frac{\partial g_2}{\partial x_1}(E(X)) & \frac{\partial g_2}{\partial x_2}(E(X)) \end{bmatrix}. \quad (5.1)$$

Podemos então concluir que, aplicando a função esperança,

$$E(Y) \approx g(E(X)). \quad (5.2)$$

Observação 5.24 *Notemos que o erro cometido na aproximação (5.2) é da ordem da variância da v.a. X .*

5.7.2 Propagação das variâncias e covariâncias

Iremos considerar o caso em que $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$, isto é, $m = n = 2$. A generalização para o caso geral é imediata.

Começemos por considerar o caso linear, isto é,

$$Y = AX + b \Leftrightarrow \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Então,

$$\begin{aligned}
 \sigma_{Y_1}^2 &= E[(Y_1 - E(Y_1))^2] \\
 &= E[(a_{11}(X_1 - E(X_1)) + a_{12}(X_2 - E(X_2)))^2] \\
 &= a_{11}^2 E[(X_1 - E(X_1))^2] + 2a_{11}a_{12}E[(X_1 - E(X_1))(X_2 - E(X_2))] + a_{12}^2 E[(X_2 - E(X_2))^2] \\
 &= a_{11}^2 \sigma_{X_1}^2 + 2a_{11}a_{12} \sigma_{X_1 X_2} + a_{12}^2 \sigma_{X_2}^2.
 \end{aligned}$$

De igual modo podemos concluir que

$$\sigma_{Y_2}^2 = a_{21}^2 \sigma_{X_1}^2 + 2a_{21}a_{22} \sigma_{X_1 X_2} + a_{22}^2 \sigma_{X_2}^2.$$

Relativamente à covariância $\sigma_{X_1 X_2}$ temos que

$$\sigma_{Y_1 Y_2}^2 = E[(Y_1 - E(Y_1))(Y_2 - E(Y_2))],$$

e, por um raciocínio análogo, temos que (prove)

$$\sigma_{Y_1 Y_2}^2 = a_{11}a_{21} \sigma_{X_1}^2 + (a_{11}a_{22} + a_{21}a_{12}) \sigma_{X_1 X_2} + a_{12}a_{22} \sigma_{X_2}^2.$$

Atendendo aos resultados anteriores não é difícil concluir que, sendo

$$C_x = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_1 X_2} & \sigma_{X_2}^2 \end{bmatrix}, \quad C_Y = \begin{bmatrix} \sigma_{Y_1}^2 & \sigma_{Y_1 Y_2} \\ \sigma_{Y_1 Y_2} & \sigma_{Y_2}^2 \end{bmatrix}$$

as matrizes de covariância de X e Y respectivamente, se tem

$$C_Y = AC_X A^T.$$

Para o caso geral temos o seguinte teorema.

Teorema 5.25 *Seja X uma v.a. de dimensão n da qual se conhecem as suas propriedades estocásticas e Y a v.a. de dimensão m dada pela relação $Y = AX + b$, com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Sejam C_X e C_Y , respectivamente, as matrizes de covariância de X e Y . Então*

$$C_Y = AC_X A^T.$$

Por outro lado, sendo Q_X e Q_Y as matrizes cofactor das v.a. X e Y , respectivamente,

$$Q_Y = AQ_X A^T.$$

Observação 5.26 *Notemos que:*

1. mesmo que C_X seja diagonal, isto é, mesmo que as v.a. X_i , $i = 1, \dots, n$, sejam não correlacionadas, a matriz C_Y é, em geral, uma matriz densa, ou seja, as v.a. Y_i , $i = 1, \dots, n$, estão, geralmente, correlacionadas;
2. a determinação da propagação das matrizes peso é deduzida imediatamente do teorema anterior considerando as inversas (caso existam).

Consideremos agora o caso não linear. Tal como no caso linear, comecemos por considerar a particularização $n = m = 2$. Linearizando a função g temos que

$$Y \approx g(E(X)) + J(X - E(X)),$$

com J a matriz de Jacobi de g calculada em $E(X)$, ou seja, a matriz (5.1). Podemos então aplicar o teorema anterior e demonstrar o seguinte resultado.

Teorema 5.27 *Seja X uma v.a. de dimensão n da qual se conhecem as suas propriedades estocásticas e Y a v.a. de dimensão m dada pela relação $Y = g(X)$, com $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sejam C_X e C_Y , respectivamente, as matrizes de covariância de X e Y . Então*

$$C_Y = JC_X J^T,$$

onde J é a matriz de Jacobi de g calculada em $E(X)$. Por outro lado, sendo Q_X e Q_Y as matrizes cofactor das v.a. X e Y , respectivamente,

$$Q_Y = JQ_X J^T.$$

5.7.3 Exercícios

Exercício 5.7.1 Para a estrela de Rigel (β da constelação de Oriente) efectuaram-se as seguintes observações relativas à sua magnitude aparente m e à sua distância à Terra r (em parsec):

magnitude aparente m	2.00	2.01	2.09	1.99	1.97
distância à Terra r	279.0	279.5	278.9	278.8	279.1

Supondo que as observações são todas independentes, determine:

1. a magnitude absoluta M de Rigel sabendo que

$$2.512^{(M-m)} = \left(\frac{10}{r}\right)^2;$$

2. uma estimativa para o erro com que vem afectado o valor determinado na alínea anterior.

Exercício 5.7.2 Para a execução de uma carta usou-se a seguinte projecção cartográfica (ϕ e λ em radianos; x e y em quilómetros):

$$\begin{cases} x = (\cos \phi_0)\lambda, \\ y = (\cos \phi_0) \ln \tan \left(\frac{\pi}{4} + \frac{\phi}{2}\right), \end{cases}$$

sendo (ϕ, λ) as coordenadas geográficas – latitude e longitude – e ϕ_0 a latitude do paralelo central. Para um paralelo central a 40° , fizeram-se as seguintes determinações (considerando-se independentes), com um receptor GPS, para um mesmo vértice geodésico:

ϕ	$40^\circ 11' 52''.57$	$40^\circ 11' 52''.54$	$40^\circ 11' 52''.54$	$40^\circ 11' 52''.61$
λ	$-8^\circ 26' 37''.11$	$-8^\circ 26' 36''.99$	$-8^\circ 26' 37''.20$	$-8^\circ 26' 37''.15$

Determine as coordenadas rectangulares x e y mais prováveis para esse vértices e os respectivos desvios padrão.

Exercício 5.7.3 Para determinar a diferença de nível h entre os pontos A e B (ver figura), efectuaram-se as seguintes observações do comprimento \overline{BC} (em metros) e dos ângulos α , β e γ (em graus)

\overline{BC}	100.0	100.1	100.1	100.3	100.2
α	25.1	25.1	25.0	25.0	24.9
β	58.5	58.3	58.3	58.4	58.3
γ	74.3	74.5	74.4	74.3	74.4

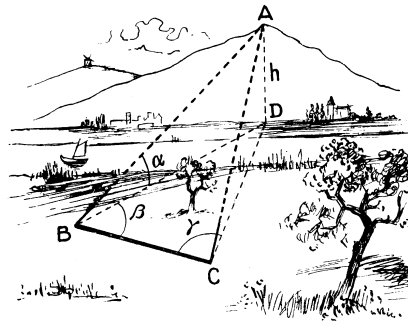
Supondo que as observações são não correlacionadas, determine:

1. a diferença de nível pretendida sabendo que

$$h = \overline{BD} \tan \alpha;$$

com

$$\overline{BD} = \frac{\overline{BC} \sin \gamma}{\sin (180 - \beta - \gamma)}.$$



2. uma estimativa para o erro com que vem afectado o valor determinado na alínea anterior.

Exercício 5.7.4 Para se determinar uma distância D (metros) em Topografia usa-se muitas vezes a relação

$$D = G \sin^2 z, \quad \text{com } G = 100(l_s - l_i),$$

sendo z (grados) o ângulo zenital e l_s e l_i (metros) leituras numa régua graduada (mira). Suponhamos que foram efectuadas as seguintes medições:

z (grados)	101.2733	101.2736	101.2732
l_s (metros)	1.756	1.754	1.757
l_i (metros)	1.000	0.997	1.001

Determine o valor mais provável para a distância D bem como o respectivo desvio padrão.

Exercício 5.7.5 Observando-se o cometa Zork obtiveram-se, num sistema de coordenadas polares, os seguintes valores para o raio (em U.A.) e o ângulo polar (em graus)

raio ρ	2.7	2.0	1.61	1.2	1.02
ângulo θ	48°	67°	83°	108°	126°

A primeira lei de Kepler estabelece uma relação entre o raio ρ , o ângulo θ e a excentricidade e da elipse através da relação

$$\rho = \frac{p}{1 - e \cos \theta},$$

em que p é um parâmetro real. Detremine uma estimativa para o valor mais provável para a excentricidade bem como para o desvio padrão desse valor considerando $p = 1$.

Exercício 5.7.6 Para determinar a medida do perímetro da Terra (p), Eratóstenes (276-195 a.C.) usou a relação

$$p = \frac{2\pi}{\alpha}s,$$

sendo α o ângulo (em radianos) que os raios solares faziam com a vertical do lugar em Alexandria, no solstício de Verão, ao meio-dia, e s a distância de Alexandria a Siena (local onde se sabia que os raios solares, nessa data e nessa hora, incidiam verticalmente).

Suponhamos, por hipótese, que Eratóstenes efectuou as seguintes observações (na realidade ele só efectuou a primeira)

distância s (em estádios)	5000	5020	4980	5007	4991
ângulo α (em graus)	7.2°	7.1°	7.0°	7.3°	7.2°

Considerando as observações não correlacionadas, determine uma estimativa para o valor mais provável do perímetro da Terra obtido por Eratóstenes bem como uma estimativa para o seu desvio padrão (1 estádio = 157.5 m = 0.1575 km).

5.8 O princípio dos mínimos quadrados

5.8.1 Introdução

Na primeira parte deste capítulo introduzimos os conceitos básicos que estão associados ao processo de medição/observação bem como a noção de modelo matemático e alguns tópicos sobre ajustamento. Em particular, foi dito que o ajustamento só tem sentido quando se efectuam observações redundantes.

No sentido estatístico, o ajustamento é um método para determinar estimativas para as variáveis estocásticas e os seus parâmetros de distribuição a partir de uma amostra obtida por observação. De entre os métodos de ajustamento, o dos mínimos quadrados é o mais divulgado.

Desde a sua primeira aplicação a um problema de astronomia por Carl F. Gauss, o método dos mínimos quadrados tem vindo a ser aplicado num vasto conjunto de situações tanto no campo da ciência como no da engenharia. A sua importância prática foi potenciada com o aparecimento dos computadores.

Antes de planear as observações, há que especificar o o modelo funcional que descreve o fenómeno em estudo. Esse modelo é determinado

por um certo número de variáveis (parâmetros ou observações) e por relações entre elas. É claro que existe sempre um número mínimo de variáveis que são estritamente necessárias para descrever uma situação física, um acontecimento ou um conjunto de acontecimentos. Para ilustrar este facto consideremos o seguinte exemplo.

Exemplo 5.28 Suponhamos um triângulo que pretendemos identificar através dos seus ângulos. Atendendo a que a soma dos seus ângulos internos é 2π , conhecendo apenas dois ângulos ficamos com a caracterização estabelecida. No entanto podemos considerar como variáveis do modelo funcional os três ângulos internos.

Denotemos por n_0 o número mínimo de variáveis (independentes) estritamente necessárias ao modelo e por n o número total de observações (encaradas como variáveis do modelo). É manifesto que cada variável pode ser concretizada mais do que uma vez, no entanto, iremos supor que cada variável do modelo é apenas observada uma vez. As observações devem ser funcionalmente independentes, isto é, nenhuma das n observações pode ser obtida das restantes $n - 1$. Em geral temos $n > n_0$ e então $r = n - n_0$ é designado por **redundância** ou **número de graus de liberdade**. Atendendo a que $r > 0$ é necessário utilizar algum processo que permita determinar um valor ajustado para cada variável.

5.8.2 O ajustamento dos mínimos quadrados

Devido às propriedades estocásticas inerentes às observações, as observações redundantes não são, usualmente, compatíveis com o modelo funcional. Assim, cada conjunto mínimo de variáveis conduz a um resultado diferente. Temos assim que introduzir um critério adicional para obter um resultado único (o “melhor resultado”) que se ajuste ao modelo.

Denotemos por ℓ o vector dos valores observados (que inclui as observações redundantes e inconsistentes com o modelo) e por $\hat{\ell}$ o vector dos valores ajustados e que satisfazem o modelo. Temos que $\dim \ell = \dim \hat{\ell} = n$ mas, em geral, $\ell \neq \hat{\ell}$. À diferença

$$v = \hat{\ell} - \ell$$

chamamos **vector dos resíduos** ou **vector das correcções**.

Devido à redundância, pode existir uma infinidade de estimativas para $\hat{\ell}$ (e, conseqüentemente, para v) que satisfaçam o modelo. De entre estas estimativas existe uma que, para além de ser consistente com o modelo, satisfaz o princípio dos mínimos quadrados. Este princípio assegura que os valores de $\hat{\ell}$ estão “o mais próximo possível”, no sentido da distância euclidiana, dos valores da amostra observada ℓ tomando as suas propriedades estocásticas em consideração.

Exemplo 5.29 Consideremos as observações $\ell = (\ell_1, \dots, \ell_n)$ que correspondem aos pontos do plano (x_i, ℓ_i) , $i = 1, \dots, n$, e que pretendemos ajustar a uma recta. A questão que se coloca é a de determinar os valores $\hat{\ell}_i = a + bx_i$, $i = 1, \dots, n$, por forma a que

$$\sum_{i=1}^n (\hat{\ell}_i - \ell_i)^2$$

seja mínimo. Se considerarmos $v = \hat{\ell} - \ell$, com $\hat{\ell} = (\hat{\ell}_1, \dots, \hat{\ell}_n)$ temos que a função a minimizar se pode escrever na forma $v^T v$.

O princípio dos mínimos quadrados consiste em determinar $\hat{\ell}$ (ou v , uma vez que $\hat{\ell} = \ell + v$) por forma a que

$$\phi(v) = v^T W v$$

seja mínima, sendo $W := W_\ell$ a matriz peso das observações. Uma vez que $W = Q^{-1}$ temos que a matriz peso é quadrada de ordem n e os seus elementos traduzem as propriedades estocásticas (tais como a correlação) de todas as observações.

Observação 5.30 *Note-se que a aplicação dos mínimos quadrados não necessita do conhecimento das distribuições associadas às observações. Temos apenas que conhecer a matriz peso W ou a matriz cofactor Q .*

Podemos considerar os seguintes casos particulares:

1. se as observações são não correlacionadas então W é uma matriz diagonal e portanto

$$\phi(v) = \sum_{i=1}^n w_{ii} v_i^2;$$

2. se as observações são não correlacionadas e consideradas com igual precisão (ou igual peso) então $W = I$ e

$$\phi(v) = \sum_{i=1}^n v_i^2 = v^T v.$$

As variáveis do modelo podem não ser (e geralmente não são) caracterizadas nas observações realizadas. Atendendo a este facto, as variáveis que não são observadas são designadas **parâmetros** para as distinguir das observações para as quais sabemos os valores dados pela amostra. Por exemplo, no Exemplo 5.29 as variáveis a e b são parâmetros.

Os valores dos parâmetros são geralmente desconhecidos no início e as suas estimativas são determinadas pelo processo de ajustamento. O vector dos parâmetros é denotado por Δ e $\dim \Delta = \mu$ denota o número de parâmetros.

Relativamente às equações matemáticas entre as variáveis do modelo estas podem ser de dois tipos.

1. **Equações de condição:** são designadas equações de observação as equações envolvendo pelo menos uma observação.
2. **Equações de restrição:** são designadas equações de condição as equações envolvendo apenas parâmetros e constantes.

Neste curso iremos apenas considerar modelos onde aparecem apenas equações de condição. Iremos ainda supor que os parâmetros, tal como as observações, são funcionalmente independentes, isto é, nenhum parâmetro pode ser obtido dos restantes $\mu - 1$.

Observação 5.31 *O vector Δ pode também aparecer nas equações de condição. Assim, o número de condições é maior ou igual a μ .*

As equações de condição ou de restrição podem ser lineares ou não lineares. Iremos apenas estudar o caso linear pois quando as equações são não lineares podemos reduzir este caso ao caso linear, como foi visto na secção anterior.

5.8.3 Ajustamento com equações de condição

Caso geral

Ao iniciar o processo de ajustamento temos que ter em conta os seguintes passos.

1. Especificar número mínimo de observações n_0 .
2. Considerar as n observações no sentido de assegurar que não ocorrerão deficiências.

Neste passo calculamos imediatamente a redundância $r = n - n_0$. Este valor de r significa que entre as n observações deverão existir r condições que deverão ser verificadas.

3. Considerar os parâmetros que se pretendem estimar.

Para cada parâmetro introduzido devemos considerar mais uma condição. Assim sendo, considerando μ parâmetros devemos considerar um total de $c = r + \mu$ condições. Notemos que $0 \leq \mu \leq n_0$ pois, caso contrário, o número de condições excedia o número mínimo de observações o que contradiz o facto dos parâmetros serem funcionalmente independentes. Assim sendo, $r \leq c \leq n$ e $0 \leq \mu \leq n_0$.

4. Especificar as equações de condição que definem o modelo.

Seja Δ o vector dos parâmetros e ℓ o vector das observações. O modelo funcional que iremos considerar é constituído por $c = r + \mu$ condições independentes podem ser lineares ou “linearizadas”. O vector das observações deveria verificar estas equações mas, de facto, o sistema é verificado pelo vector das observações ajustadas $\hat{\ell}$, isto é,

$$A\hat{\ell} + B\Delta = d,$$

ou ainda, sendo v o vector dos resíduos,

$$Av + B\Delta = f$$

com $f = d - A\hat{\ell}$. Notemos que, relativamente ao sistema estabelecido, $A \in \mathbb{R}^{c \times n}$ e $B \in \mathbb{R}^{c \times \mu}$ sendo as suas características iguais a c e μ , respectivamente. Assim sendo, o sistema de equações

$$[A \ B] \begin{bmatrix} v \\ \Delta \end{bmatrix}$$

é indeterminado (tem uma infinidade de soluções) uma vez que a matriz $[A \ B]$ tem característica c .

5. Determinar a melhor solução do sistema $Av + B\Delta = f$ no sentido dos mínimos quadrados.

Atendendo aos passos descritos, o problema que temos que resolver é

$$\begin{cases} \min v^T W v \\ \text{s.a } Av + B\Delta = f \end{cases} .$$

Este problema vai ser resolvido utilizando o método dos multiplicadores de Lagrange. Assim, sendo λ o vector dos multiplicadores de Lagrange ($\lambda \in \mathbb{R}^c$), o problema a minimizar é equivalente a

$$\min \psi(v, \lambda, \Delta) = v^T W v - 2\lambda^T (Av + B\Delta - f).$$

Para resolver este último problema há que determinar as derivadas parciais da função a minimizar em ordem a cada uma das variáveis. Para simplificar a notação, consideremos

$$\frac{\partial \psi}{\partial v}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial v_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial v_n}(v, \lambda, \Delta) \right]^T ,$$

$$\frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial \lambda_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial \lambda_c}(v, \lambda, \Delta) \right]^T ,$$

$$\frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) := \left[\frac{\partial \psi}{\partial \Delta_1}(v, \lambda, \Delta), \dots, \frac{\partial \psi}{\partial \Delta_\mu}(v, \lambda, \Delta) \right]^T .$$

Assim, a função $\psi : \mathbb{R}^{n+c+\mu} \rightarrow \mathbb{R}$ tem um extremo (prova-se que é mínimo) no ponto (v, λ, Δ) que satisfaz as chamadas equações normais

$$\begin{cases} \frac{\partial \psi}{\partial v}(v, \lambda, \Delta) = 0 \\ \frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) = 0 \\ \frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) = 0 \end{cases} .$$

Uma vez que W é uma matriz simétrica, tem-se que (prove)

$$\frac{\partial \psi}{\partial v}(v, \lambda, \Delta) = 2v^T W - 2\lambda^T A.$$

Além disso, também se pode concluir que (prove)

$$\frac{\partial \psi}{\partial \lambda}(v, \lambda, \Delta) = -2(Av + B\Delta - f)$$

e que

$$\frac{\partial \psi}{\partial \Delta}(v, \lambda, \Delta) = -2\lambda^T B.$$

Atendendo a estes resultados temos que as equações normais são da forma

$$\begin{cases} -Wv + A^T \lambda = 0 \\ Av + B\Delta = f \\ B^T \lambda = 0 \end{cases}$$

ou, na forma matricial,

$$\begin{bmatrix} -W & A^T & 0 \\ A & 0 & B \\ 0 & B^T & 0 \end{bmatrix} \begin{bmatrix} v \\ \lambda \\ \Delta \end{bmatrix} = \begin{bmatrix} 0 \\ f \\ 0 \end{bmatrix}.$$

Note-se que a matriz do sistema de equações normais é quadrada de ordem $n + c + \mu$. Além disso é uma matriz simétrica e não singular. De facto,

$$\left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & 0 & B & f \\ 0 & B^T & 0 & 0 \end{array} \right] \xrightarrow{L_2 := AW^{-1}L_1 + L_2} \left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & B^T & 0 & 0 \end{array} \right].$$

A matriz AQA^T é quadrada de ordem c e pode ser vista como uma matriz cofactor resultante da propagação de Q pelas equações de condição. Como, além disso, é invertível (W é invertível e a característica de A é c), temos que, denotando a sua inversa por W_c ($W_c := (AQA^T)^{-1}$),

$$\left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & B^T & 0 & 0 \end{array} \right] \xrightarrow{L_3 := B^T W_c L_2 - L_3} \left[\begin{array}{ccc|c} -W & A^T & 0 & 0 \\ A & AQA^T & B & f \\ 0 & 0 & N & B^T W_c f \end{array} \right]$$

sendo $N = B^T W_c B$. A matriz N é uma matriz quadrada de ordem μ invertível pois W_c é invertível e a matriz B tem característica igual a μ . Como tal, o sistema de equações normais é possível e determinado e a sua solução pode ser obtida do seguinte modo:

1. Resolver $N\Delta = B^T W_c f$;
2. Determinar $\lambda = W_c[f - B\Delta]$;
3. Determinar $v = QA^T \lambda$.

O valor das observações ajustadas é então dado por $\hat{\ell} = v + \ell$. Temos então o seguinte algoritmo para resolver o problema dos mínimos quadrados.

Algoritmo 4.1 Mínimos quadrados: equações do tipo $A\hat{\ell} + B\Delta = d$.

1. $f = d - A\ell$
2. $W_c = (AQAT)^{-1}$
 $N = B^T W_c B$
3. $\Delta = N^{-1} B^T W_c f$
4. $\lambda = W_c[f - B\Delta]$
5. $v = QA^T \lambda$
6. $\hat{\ell} = \ell + v$

Condições apenas com observações

Suponhamos que o número de parâmetros é $\mu = 0$. Neste caso, o número de condições c é igual à redundância r e as equações (lineares) de condição são da forma

$$A\hat{\ell} = d \Leftrightarrow Av = f,$$

com $f = d - A\ell$. Uma vez que, neste caso, $B = 0$, as equações normais são da forma

$$\begin{bmatrix} -W & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

É fácil concluir, atendendo ao que foi feito no caso geral, que a solução deste problema pode ser dada pelo seguinte algoritmo.

Algoritmo 4.2 Mínimos quadrados: equações do tipo $A\hat{\ell} = d$.

1. $f = d - A\ell$
2. $W_c = (AQAT)^{-1}$
3. $\lambda = W_c f$
4. $v = QA^T \lambda$
5. $\hat{\ell} = \ell + v$

Observações indirectas

Vamos agora considerar o caso em que o ajustamento é efectuado com observações e parâmetros mas com a restrição de que cada condição contém apenas uma observação. Neste caso, como n é o número de observações, temos que $c = n$. A forma geral das equações de condição é

$$\hat{\ell} + B\Delta = d,$$

com $B \in \mathbb{R}^{n \times \mu}$. Uma vez que $\hat{\ell} = v + \ell$, as equações de condição podem ser escritas na forma

$$v + B\Delta = f,$$

com $f = d - \ell$.

As equações normais podem ser obtidas como no caso geral considerando $A = I$. No entanto, neste caso, temos que $v = f - B\Delta$ e assim o problema de minimização fica reduzido a

$$\min \psi(\Delta) = (f - B\Delta)^T W (f - B\Delta).$$

Como

$$\psi(\Delta) = f^T W f - 2f^T W B \Delta + \Delta^T B^T W B \Delta,$$

temos que

$$\frac{\partial \psi}{\partial \Delta}(\Delta) = -2f^T W B + 2\Delta^T B^T W B \Delta.$$

Assim sendo, as equações normais

$$\frac{\partial \psi}{\partial \Delta}(\Delta) = 0$$

são da forma

$$B^T W B \Delta = B^T W f$$

que é um sistema linear possível e determinado de ordem μ . Logo, considerando $N = B^T W B$, temos que $\Delta = N^{-1} B^T W f$. O algoritmo para obter a solução do problema dos mínimos quadrados é, neste caso, o seguinte.

Algoritmo 4.3 Mínimos quadrados: equações do tipo $\hat{\ell} + B\Delta = d$.

1. $f = d - \ell$

2. $N = B^T W_c B$
3. $\Delta = N^{-1} B^T W_c f$
4. $v = f - B\Delta$
5. $\hat{\ell} = l + v$

5.8.4 Estimativas para a precisão

Determinação da variância de referência

No final do processo de ajustamento, a variância de referência pode ser dada pela estimativa

$$\sigma_0^2 \approx s_0^2 = \frac{v^T W v}{r},$$

sendo v o vector dos resíduos, W a matriz peso das observações e r a redundância. Note-se que quando considerámos m observações independentes da mesma quantidade, a primeira observação estabelece um valor para a incógnia e as restantes observações são redundantes. Nesse caso, como vimos,

$$s_x^2 = \frac{v^T v}{m - 1}.$$

Determinação da matriz cofactor

Vamos determinar estimativas para as matrizes cofactor tanto das observações ajustadas $\hat{\ell}$ como dos valores dos parâmetros Δ .

Vamos considerar o caso geral, isto é, o caso em que se consideram equações do tipo $A\hat{\ell} + B\Delta = f$. Atendendo ao Algoritmo 4.1 temos o seguinte processo de cálculo.

1. Determinar Q_f . Pelo Teorema 5.25 temos que

$$Q_f = (-A)Q(-A)^T = AQA^T.$$

2. Determinar Q_Δ . Temos, sucessivamente,

$$Q_\Delta = (N^{-1}B^T W_c)Q_f(N^{-1}B^T W_c)^T = N^{-1}B^T W_c Q_f W_c B N^{-1} = N^{-1}B^T W_c B N^{-1}$$

Atendendo à definição da matriz N , temos que $Q_\Delta = N^{-1}$.

3. Determinar Q_λ . Temos que

$$\lambda = W_c[f - B\Delta] = W_c(I - BN^{-1}B^TW_c)f$$

Assim

$$\begin{aligned} Q_\lambda &= W_c(I - BN^{-1}B^TW_c)Q_f(I - BN^{-1}B^TW_c)^TW_c \\ &= (I - W_cBN^{-1}B^T - W_cBN^{-1}B^T + W_cBN^{-1}B^TW_cBN^{-1}B^T)W_c \\ &= (I - W_cBN^{-1}B^T)W_c, \end{aligned}$$

pois $Q_fW_c = I$ e $N^{-1}B^TW_cB = I$. Como tal,

$$Q_v = W_c(I - BN^{-1}B^TW_c).$$

4. Determinar Q_v . Como se pode ver facilmente

$$Q_v = QA^TQ_\lambda(QA^T)^T = QA^TQ_\lambda AQ,$$

ou, o que é equivalente,

$$Q_v = QA^TW_cAQ - QA^TW_cBN^{-1}B^TW_cAQ.$$

5. Determinar $Q_{\hat{\ell}}$. Temos que

$$\begin{aligned} \hat{\ell} &= \ell + v \\ &= \ell + QA^T\lambda \\ &= \ell + QA^TW_c(f - B\Delta) \\ &= \ell + QA^TW_c(I - BN^{-1}B^TW_c)f \\ &= (I - Q_vW)\ell + QA^TQ_\ell d. \end{aligned}$$

Assim,

$$Q_{\hat{\ell}} = (I - Q_vW)Q_\ell(I - Q_vW)^T = Q - 2Q_v + Q_vWQ_v.$$

Ora, como (prove)

$$Q_vWQ_v = Q_v$$

sai que

$$Q_{\hat{\ell}} = Q - Q_v.$$

Podemos resumir o efectuado no seguinte algoritmo.

Algoritmo 4.4 Propagação da matriz cofactor: equações do tipo $A\hat{\ell} + B\Delta = d$.

1. $Q_{\Delta} = N^{-1}$
2. $Q_v = QA^TW_cAQ - QA^TW_cBN^{-1}B^TW_cAQ$
3. $Q_{\hat{\ell}} = Q - Q_v$

Para os casos particulares, conclui-se, de forma similar, que os algoritmos de propagação da matriz cofactor são os seguintes.

Algoritmo 4.5 Propagação da matriz cofactor: equações do tipo $A\hat{\ell} = d$.

1. $Q_v = QA^TW_cAQ$
2. $Q_{\hat{\ell}} = Q - Q_v$

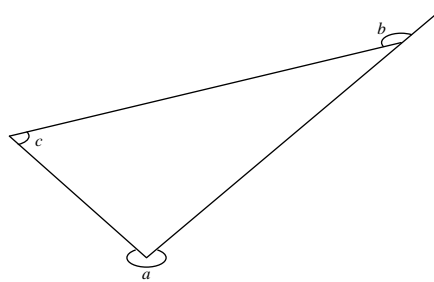
Algoritmo 4.6 Propagação da matriz cofactor: equações do tipo $\hat{\ell} + B\Delta = d$.

1. $Q_{\Delta} = N^{-1}$
2. $Q_v = Q - BN^{-1}B^T$
3. $Q_{\hat{\ell}} = Q - Q_v$

5.8.5 Exercícios

Exercício 5.8.1 Mediram-se, com a mesma precisão, os três ângulos internos de um triângulo tendo-se obtido $\alpha = 40^{\circ} 19' 02''$, $\beta = 70^{\circ} 30' 0''$ e $\gamma = 69^{\circ} 11' 01''$. Supondo que as medições são não correlacionadas, calcule:

1. o valor ajustado dos três ângulos;
2. a matriz cofactor;
3. uma estimativa para a variância de referência.

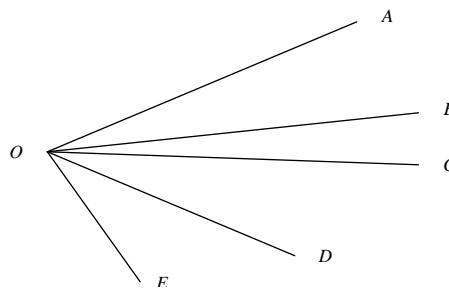


Exercício 5.8.2 Observaram-se os seguintes valores para os ângulos a , b e c do triângulo representados na figura: $a = 320^\circ 19' 40''$, $b = 129^\circ 14' 37''$ e $c = 89^\circ 34' 20''$, sendo dado às observações os pesos $p_a = 3$, $p_b = 4$ e $p_c = 2$. Supondo que as observações são não correlacionadas determine o valor ajustado dos três ângulos internos do triângulo e a matriz cofactor.

Exercício 5.8.3 Mediram-se os ângulos $\angle AOB$, $\angle BOC$, $\angle AOC$, $\angle COD$ e $\angle COE$ indicados na figura obtendo-se os valores

$\angle AOB$	$\angle BOC$	$\angle AOC$	$\angle COD$	$\angle COE$
$30^\circ 15' 1''$	$20^\circ 00' 00''$	$50^\circ 15' 18''$	$30^\circ 00' 00''$	$70^\circ 00' 01''$

não correlacionados e com igual peso.



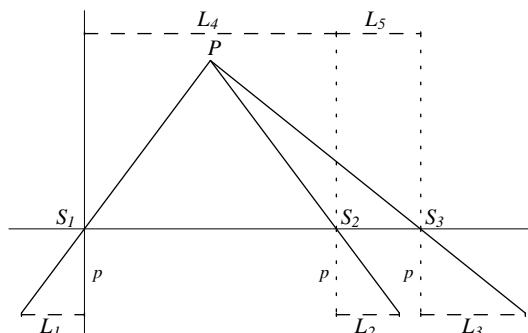
Determine os valores ajustados dos ângulos referidos, a matriz cofactor e uma estimativa para a variância de referência.

Exercício 5.8.4 Para determinar a distância vertical ajustada dos pontos B, C e D a uma plataforma horizontal, situada abaixo destes, mediram-se as seguintes distâncias (em metros):

	$A - B$	$A - C$	$A - D$	$B - C$	$C - D$
Distância vertical	10.216	12.384	15.869	2.138	3.453
Distância horizontal	16	30	18	14	12

Sabendo que a distância vertical do ponto A (ponto mais baixo) a essa plataforma é de 15.000 metros e que o peso de cada distância vertical medida é inversamente proporcional à distância horizontal entre os pontos, determine as distâncias pretendidas.

Exercício 5.8.5 Considere o seguinte esquema que pode representar três câmeras fotográficas colocadas nas posições S_1 , S_1 e S_3 alinhadas sobre o eixo das abscissas e que fotografam um ponto P de coordenadas (x_1, x_2) . Determine os valores ajustados de x_1 e x_2 supondo que é co-



nhecido $p = 100$ mm (sem erro) e ainda as observações L_i , $i = 1, \dots, 5$, supostas não correlacionadas, e os correspondentes desvios padrão e cujos valores são dados na seguinte tabela:

Observação	Valor	Desvio padrão
L_1	16.5 mm	0.10 mm
L_2	3.8 mm	0.10 mm
L_3	20.4 mm	0.10 mm
L_4	10.0 m	0.05 m
L_5	8.0 m	0.05 m

Exercício 5.8.6 Num nivelamento trigonométrico de alta precisão entre quatro marcos, foram feitas as seguintes medições:

lado	dist. (km)	dN (m)	lado	dN (m)
AB	2.25	+39.274	BA	-39.243
BC	4.00	-94.848	CB	+94.892
CD	1.75	+20.052	DC	-20.032
DA	3.00	+35.619	AD	-35.587

Determine a cota ajustada de cada marco, conhecendo $N_A = 87.631m$.

Exercício 5.8.7 Por forma a estabelecer a cota de 3 marcos, B , C e D , fizeram-se dois nivelamentos, $ABCD$ e $ABDA$; foram registados os seguintes resultados (em metros) :

nivelamento 1: $dN_{AB} = +3.753$, $dN_{BC} = +5.548$, $dN_{CD} = +10.427$,
 $dN_{DA} = -19.721$;

nivelamento 2: $dN_{AC} = +9.280$, $dN_{CB} = -5.540$, $dN_{BA} = -3.755$;

Determine os valores mais prováveis para as cotas dos marcos sabendo que $N_A = 169.721$ m.

Exercício 5.8.8 Num nivelamento fechado, $ABCD$, mediram-se as seguintes diferenças de nível :

lado	AB	BC	CD	DA
dN (m)	+5.216	+2.394	+1.055	-8.690

Calcule as cotas mais prováveis para B , C e D , sabendo que $N_A = 100.000$ m e considerando as seguintes condições alternativas :

1. Todas as observações têm igual precisão;
2. Os troços BC e CD têm o dobro do comprimento dos troços AB e DA ;
3. Os troços BC e CD foram medidos duas vezes, sendo os valores dados as médias das diferenças de nível obtidas.

Exercício 5.8.9 Numa poligonal foram medidas as seguintes coordenadas relativas :

lado	ΔM (m)	ΔP (m)
AB	1105.362	1346.542
BC	964.547	-965.426
CD	-892.513	-882.492
DA	-1177.341	501.334
BD	72.084	-1847.982

Considerando que todas as medições foram feitas com igual precisão, calcule as coordenadas ajustadas para as estações B , C e D , sabendo que as coordenadas de A são $M_A = 1000.000$ m, $P_A = 1000.000$ m.

Exercício 5.8.10 Mediram-se com a mesma precisão os três lados de um triângulo tendo-se obtido $a = 324.53$ m, $b = 541.41$ m, $c = 483.22$ m, todos com um desvio padrão de 5 cm, e o ângulo oposto ao lado a , $\hat{a} = 40^{\circ}4'25''$ com um desvio padrão de $5''$.

Determine:

1. o valor ajustado dos ângulos internos;
2. a matriz cofactor;
3. uma estimativa para a variância de referência.

Exercício 5.8.11 A lei de Hooke estabelece que a força F aplicada a uma mola é directamente proporcional ao deslocamento provocado de acordo com a seguinte relação

$$F = k(e - e_0),$$

onde k é a constante da mola, e o comprimento da mola quando sujeita à força F e e_0 o comprimento inicial da mola.

No sentido de determinar a constante da mola usaram-se diferentes forças (conhecidas) tendo sido observados os comprimentos resultantes, dados na seguinte tabela

força F (em gramas)	3	5	8	10
comprimento e (em milímetros)	13.3	16.3	19.4	20.9

Sabendo que o comprimento inicial da mola é $e_0 = 10$ mm e considerando as medições (não correlacionadas) com precisão inversamente proporcional ao comprimento observado, determine a melhor estimativa para a constante da mola, usando o algoritmo dos mínimos quadrados.

Exercício 5.8.12 Suponhamos que l_1 , l_2 e l_3 são três observações independentes de uma distância Δ . Se as observações tiverem pesos w_1 , w_2 e w_3 , respectivamente, mostre que a melhor estimativa para a distância pretendida, segundo o critério dos mínimos quadrados, é igual à média pesada das observações, i.e.,

$$\Delta = \frac{w_1 l_1 + w_2 l_2 + w_3 l_3}{w_1 + w_2 + w_3}.$$

Exercício 5.8.13 As coordenadas de um ponto y do plano foram medidas por dois métodos distintos tendo sido obtidos os seguintes resultados:

- método 1: $y_1 = \begin{bmatrix} 1.1 \\ 2.0 \end{bmatrix}$, com a matriz peso $W_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$;
- método 2: $y_2 = \begin{bmatrix} 1.0 \\ 2.1 \end{bmatrix}$, com a matriz peso $W_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

Supondo que não existe correlação entre y_1 e y_2 , determine:

1. a melhor estimativa para as coordenadas do ponto, usando o algoritmo dos mínimos quadrados;
2. a matriz cofactor dos valores obtidos na alínea anterior e uma estimativa para a variância de referência.

Referências bibliográficas

- R.L. Burden e J.D. Faires (1988), Numerical Analysis, 4th ed., PWS-Kent, Boston.
- B. Fornberg (1988), *Generation of finite difference formulas on arbitrarily spaced grids*, Math. Comp., 51, 699-706.
- E.M. Mikhail e F.A. Ackerman, (1976), Observations and Least Squares, Sohn Wiley & Sons, Washington DC.
- B. Murteira (1979), Probabilidades e Estatística, Mc-Graw Hill, Lisboa.
- H. Pina (1995), Métodos Numéricos, Mc Graw-Hill, Lisboa.
- P.M. Prenter (1976), Splines and Variational Methods, John Wiley & Sons, New York.
- M. Rosa (1992), Tópicos de Análise Numérica, Dep. Matemática, Univ. Coimbra.

Conteúdo

1	Preliminares	3
1.1	Introdução	3
1.2	Erro absoluto e erro relativo	4
1.3	Condicionamento de matrizes	5
1.4	O polinómio de Taylor	8
1.5	Exercícios	10
2	Interpolação polinomial de funções de uma variável	14
2.1	Introdução	14
2.2	Interpolação polinomial de Lagrange	15
2.2.1	Fórmula de Lagrange	16
2.2.2	Fórmula de Newton	18
2.2.3	Erro de interpolação	20
2.2.4	Zeros dos polinómios de Chebyshev	25
2.2.5	Exercícios	27
2.3	O polinómio interpolador de Lagrange segmentado	31
2.3.1	Caso linear	31
2.3.2	Caso quadrático	34
2.3.3	Caso geral	37
2.3.4	Exercícios	39
2.4	Interpolação de Hermite	41
2.4.1	Existência e unicidade	41
2.4.2	Erro de interpolação	44
2.4.3	Interpolação de Hermite segmentada	46
2.4.4	Polinómios osculadores	48
2.4.5	Exercícios	52
2.5	Interpolação com funções <i>spline</i>	54
2.5.1	Abordagem clássica	54
2.5.2	Funções de base	56

2.5.3	Estudo do erro	62
2.5.4	Exercícios	62
2.6	Derivação numérica	64
2.6.1	Aproximação da primeira derivada	64
2.6.2	Aproximação da segunda derivada. Algumas fórmulas	68
2.6.3	Aproximação de derivadas de ordem superior	69
2.6.4	Exercícios	69
3	Interpolação de funções definidas em \mathbb{R}^2	71
3.1	Introdução	71
3.2	Polinómio interpolador de Lagrange	72
3.3	Interpolação de Lagrange segmentada	76
3.4	O polinómio interpolador de Hermite	79
4	Interpolação de curvas paramétricas	82
4.1	Interpolação paramétrica	82
4.2	Interpolação cúbica de Hermite segmentada	84
4.3	Curvas de Bézier	86
4.4	Exercícios	88
5	Observações e mínimos quadrados	89
5.1	Introdução	89
5.2	Modelo matemático e modelo estocástico	90
5.3	Alguns conceitos probabilísticos	91
5.3.1	Distribuições unidimensionais	91
5.3.2	Distribuições bidimensionais	95
5.3.3	Distribuições marginais	97
5.3.4	Independência	99
5.3.5	Funções de variáveis aleatórias	102
5.3.6	Esperança matemática de um vector aleatório	103
5.3.7	Variância, covariância e coeficiente de correlação	105
5.3.8	Exercícios	107
5.4	Alguns conceitos estatísticos	110
5.4.1	Introdução	110
5.4.2	Distribuição de frequências	111
5.4.3	Média, variância, covariância e coeficiente de correlação de uma amostra	113

5.4.4	Exercícios	117
5.5	Propriedades dos erros de observação	117
5.5.1	Os tipos de erros de observação	117
5.6	Precisão: cofactor e peso	119
5.7	Propagação da média, da variância e covariância	121
5.7.1	Propagação das médias	121
5.7.2	Propagação das variâncias e covariâncias	122
5.7.3	Exercícios	124
5.8	O princípio dos mínimos quadrados	127
5.8.1	Introdução	127
5.8.2	O ajustamento dos mínimos quadrados	128
5.8.3	Ajustamento com equações de condição	131
5.8.4	Estimativas para a precisão	136
5.8.5	Exercícios	138