Mínimos quadrados

Esmeralda Sousa Dias

É frequente ser necessário determinar uma curva bem ajustada a um conjunto de dados obtidos experimentalmente. Por exemplo, suponha que como resultado de uma certa experiência se obtém a seguinte tabela de dados:

$$\begin{array}{c|c|c|c} x_1 & x_2 & \cdots & x_n \\ \hline y_1 & y_2 & \cdots & y_n \end{array} \tag{1}$$

Os dados (x_i, y_i) podem representar-se por pontos do plano. Pretende-se assim encontrar a curva $y = f(x)$ que melhor se ajusta aos dados (pontos) obtidos. Na figura seguinte indicam-se algumas possibilidades.

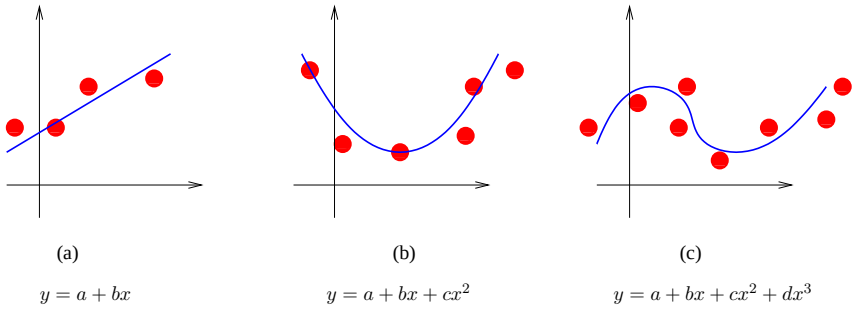

Figura 1: Ajuste de curvas a um conjunto de pontos

Comecemos por analisar o caso em que se pretende determinar uma recta de equação $y = ax + b$ que melhor se ajuste aos dados da Tabela (1). Em Estatística, este problema é conhecido como modelo de regressão linear. Se os pontos (x_i, y_i) são colineares então os coeficientes a e b satisfazem o sistema de equações lineares seguinte

$$\begin{cases} y_1 = a + bx_1 \\ y_2 = a + bx_2 \\ \quad\cdots \\ y_n = a + bx_n \end{cases} \iff \mathbf{y} = A\mathbf{x}, \tag{2}$$

com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Se os pontos (x_i, y_i) não são colineares (como é o caso na Figura 1-(a)) então o sistema (2) é impossível. Neste caso, iremos procurar uma solução de mínimos quadrados

$\hat{\mathbf{x}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$, onde $\hat{a}$ e $\hat{b}$ definem a recta $y = \hat{a} + \hat{b}x$ indicada na Figura 1-(a). Precisemos o que entendemos por uma solução de mínimos quadrados para um sistema de equações lineares $\mathbf{y} = A\mathbf{x}$.

Definição 1.1. Se A é uma matriz $n \times p$ e $\mathbf{y} \in \mathbb{R}^n$, **uma solução de mínimos quadrados** para o sistema $\mathbf{y} = A\mathbf{x}$ é um vector $\hat{\mathbf{x}}$ de $\mathbb{R}^p$ tal que

$$\|\mathbf{y} - A\hat{\mathbf{x}}\| \leq \|\mathbf{y} - A\mathbf{x}\|, \tag{3}$$

para todo o $\mathbf{x}$ de $\mathbb{R}^p$.

A designação "mínimos quadrados" (ou least squares em inglês) está relacionada com o facto de $\|\mathbf{y} - A\mathbf{x}\|$ ser a raiz quadrada de uma soma de quadrados e de se pretender minimizar esta quantidade.

Para encontrar uma solução de mínimos quadrados para o sistema $\mathbf{y} = A\mathbf{x}$ iremos aplicar os conhecimentos adquiridos neste curso de Álgebra Linear. Relembremos portanto alguns resultados essenciais:

(i) $A\mathbf{x}$ é uma combinação linear das colunas de A, ou equivalentemente $A\mathbf{x}$ pertence ao espaço das colunas de A (abreviadamente $EC(A)$).

(ii) A equação (3) é equivalente a dizer-se que $A\hat{\mathbf{x}}$ está mais próximo de $\mathbf{y}$ que qualquer outro vector do espaço das colunas de A. Pelo teorema da melhor aproximação isto quer dizer que:

- Uma solução de mínimos quadrados, $\hat{\mathbf{x}}$, para $\mathbf{y} = A\mathbf{x}$, é tal que $A\hat{\mathbf{x}}$ **é a projecção ortogonal de** $\mathbf{y}$ sobre o espaço das colunas de A:

$$A\hat{\mathbf{x}} = \mathrm{proj}_{EC(A)}\, \mathbf{y} \tag{4}$$

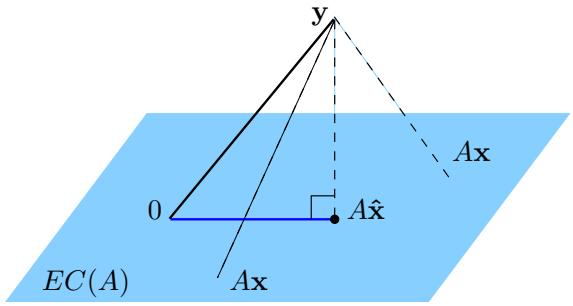

Figura 2: O vector $A\hat{\mathbf{x}}$ está mais próximo de $\mathbf{y}$ do que qualquer outro vector $A\mathbf{x} \in EC(A)$.

Nota: Note que a equação (4) é sempre possível já que ambos os vectores $\mathrm{proj}_{EC(A)}\, \mathbf{y}$ e $A\hat{\mathbf{x}}$ pertencem ao espaço das colunas de A.

O teorema da decomposição ortogonal diz-nos que qualquer vector de um espaço linear pode escrever-se como a soma da projecção ortogonal desse vector sobre um

subespaço com a projecção ortogonal sobre o subespaço complementar. Assim, o vector $\mathbf{y} - A\hat{\mathbf{x}}$ é ortogonal ao espaço das colunas de A, isto é pertence ao complemento ortogonal de $EC(A)$. Como o complemento ortogonal de espaço das colunas de uma matriz é o núcleo da sua transposta temos:

$$\begin{aligned}(\mathbf{y} - A\hat{\mathbf{x}}) \perp EC(A) \iff & \quad (\mathbf{y} - A\hat{\mathbf{x}}) \in N(A^T) \iff A^T(\mathbf{y} - A\hat{\mathbf{x}}) = 0 \\ \iff & \quad A^T\mathbf{y} - A^T A\hat{\mathbf{x}} = 0 \iff A^T A\hat{\mathbf{x}} = A^T\mathbf{y}.\end{aligned} \tag{5}$$

Podemos então concluir que uma solução de mínimos quadrados, $\hat{\mathbf{x}}$, satisfaz a equação

$$A^T A\mathbf{x} = A^T\mathbf{y}. \tag{6}$$

A equação (6) é designada por **equação normal** para o sistema $\mathbf{y} = A\mathbf{x}$.

Acabámos de mostrar que uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ satisfaz a equação normal. Podemos também mostrar o resultado inverso, isto é, que uma qualquer solução da equação normal é uma solução de mínimos quadrados do sistema referido. Este é o resultado enunciado no teorema seguinte.

Teorema 1.1. O conjunto de soluções de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ coincide com o conjunto (não vazio) das soluções da equação normal $A^T A\mathbf{x} = A^T\mathbf{y}$.

Demonstração. Já mostrámos que o conjunto das soluções de mínimos quadrados é não vazio, e que uma solução de mínimos quadrados verifica a equação normal. Basta portanto mostrar que qualquer solução da equação normal é uma solução de mínimos quadrados do sistema. Para tal, suponha-se que $\hat{\mathbf{x}}$ é solução da equação normal e mostre-se que $\hat{\mathbf{x}}$ é uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ (ou seja, que $A\hat{\mathbf{x}}$ é a projecção ortogonal de $\mathbf{y}$ sobre $EC(A)$). Como,

$$\hat{\mathbf{x}} \text{ solução de } A^T A\mathbf{x} = A^T\mathbf{y} \iff A^T A\hat{\mathbf{x}} = A^T\mathbf{y},$$

então por (5) tem-se que $(\mathbf{y} - A\hat{\mathbf{x}}) \perp EC(A)$. Logo,

$$\mathbf{y} = A\hat{\mathbf{x}} + (\mathbf{y} - A\hat{\mathbf{x}}),$$

é a decomposição ortogonal de $\mathbf{y}$ na soma de um vector de $EC(A)$ com um vector de $EC(A)^\perp$. Da unicidade da decomposição ortogonal tem-se que $A\hat{\mathbf{x}}$ tem de ser a projecção ortogonal de $\mathbf{y}$ sobre $EC(A)$. □

Exemplo 1. Suponha que se fizeram as seguintes observações:

x	1	2	3
y	1	2	2

Estas observações correspondem a três pontos no plano xy, representados na Figura 3. Vemos claramente que não existe nenhuma recta $y = a + bx$ que passe por estes três pontos. Esta afirmação é equivalente a dizer que o sistema seguinte é impossível

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \iff A\mathbf{x} = \mathbf{y}. \tag{7}$$

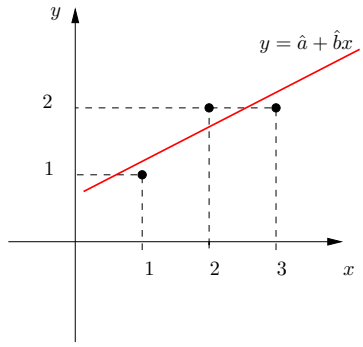

Figura 3: Recta de mínimos quadrados do Exemplo 1.

Pretende-se determinar a recta $y = \hat{a} + \hat{b}x$ que melhor se ajusta aos pontos observados. Isto é, pretende-se determinar um vector $\hat{\mathbf{x}} \in \mathbb{R}^2$ que seja uma solução de mínimos quadrados de $A\mathbf{x} = \mathbf{y}$.

Usando o Teorema 1.1, a equação normal é:

$$A^T A\mathbf{x} = A^T\mathbf{y} \Longleftrightarrow \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}.$$

Como a matriz $A^T A$ é invertível então o sistema de equações normais tem solução única dada por $(A^T A)^{-1} A^T \mathbf{y}$. Ou seja:

$$\begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 11 \\ 15 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}$$

A equação da recta de mínimos quadrados é: $y = \frac{2}{3} + \frac{1}{2}x$.

♦

Sendo $\hat{\mathbf{x}}$ uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$, então $A\hat{\mathbf{x}}$ é uma aproximação de $\mathbf{y}$ (a melhor aproximação). A distância de $A\hat{\mathbf{x}}$ a $\mathbf{y}$ é designada por **erro de mínimos quadrados**, ou seja $\|\mathbf{y} - A\hat{\mathbf{x}}\|$. O vector $\mathbf{d} = \mathbf{y} - A\hat{\mathbf{x}}$ é denominado por **vector dos desvios**.

Para o Exemplo 1 calcule-se o vector dos desvios, bem como o respectivo erro de mínimos quadrados. Recorde-se que a solução de mínimos quadrados obtida é $\hat{\mathbf{x}} = (\frac{2}{3}, \frac{1}{2})$ e $\mathbf{y} = (1, 2, 2)$. Logo, o vector dos desvios é:

$$\mathbf{d} = \mathbf{y} - A\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix} = \begin{bmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix}$$

O erro de mínimos quadrados é $\|\mathbf{d}\| = \frac{\sqrt{6}}{6}$. Na Figura 4 encontram-se representados os desvios, d_1, d_2 e d_3, ou seja as componentes de $\mathbf{d} = (d_1, d_2, d_3)$.

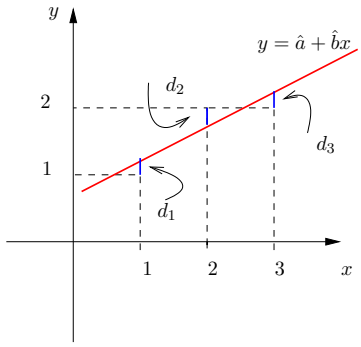

Figura 4: Desvios para o Exemplo 1.

No Exemplo 1 o problema de mínimos quadrados possui solução única. Põe-se naturalmente a questão de saber quando é que para um dado sistema $A\mathbf{x} = \mathbf{y}$ existe solução única de mínimos quadrados. Obviamente que tal se verifica sempre que a matriz quadrada $A^T A$ é invertível. O resultado que se segue dá-nos um critério para a unicidade da solução de mínimos quadrados em termos da matriz A.

Teorema 1.2. A matriz $A^T A$ é invertível se e só se as colunas de A são linearmente independentes. Neste caso, o sistema $A\mathbf{x} = \mathbf{y}$ tem solução de mínimos quadrados única, dada por

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y}.$$

Demonstração. A unicidade de solução de mínimos quadrados é obviamente equivalente à existência de inversa de $A^T A$. Quando esta matriz é invertível então, da equação normal segue que a solução de mínimos quadrados é dada pela expressão no teorema. Falta pois mostrar que a matriz $A^T A$ é invertível se e só se as colunas de A são linearmente independentes. A prova deste resultado será feita em dois passos: (i) mostrar que o núcleo de A é igual ao núcleo de $A^T A$; (ii) usar (i) para mostrar o resultado.

(i) Mostrar que $N(A) = N(A^T A)$:

Se $\mathbf{x}$ pertence ao $N(A)$, isto é, $A\mathbf{x} = 0$, então pertence ao núcleo de $A^T A$, já que

$$A\mathbf{x} = 0 \Longrightarrow A^T A\mathbf{x} = 0.$$

Ou seja $N(A) \subset N(A^T A)$.

Para a implicação inversa, suponha-se agora que $\mathbf{x} \in N(A^T A)$, isto é, $A^T A\mathbf{x} = 0$. Então

$$A^T A\mathbf{x} = 0 \Longrightarrow \mathbf{x}^T A^T A\mathbf{x} = 0.$$

Porém,

$$\mathbf{x}^T A^T A\mathbf{x} = 0 \Longleftrightarrow \|A\mathbf{x}\|^2 = 0 \Longleftrightarrow A\mathbf{x} = 0 \Longleftrightarrow \mathbf{x} \in N(A).$$

(ii) Suponha-se que $A^T A$ é invertível, então o sistema homogéneo $A^T A\mathbf{x} = 0$ tem exclusivamente a solução $\mathbf{x} = 0$, ou seja $N(A^T A) = \{0\}$. Como, por (i), $N(A^T A) = \{0\} = N(A)$, então a equação $A\mathbf{x} = 0$ tem sómente a solução nula.

Por definição de independência linear, dizer-se que $A\mathbf{x} = 0$ tem sómente a solução nula, é equivalente a afirmar que as colunas de A são linearmente independentes.

Para a implicação inversa: Se que as colunas de A são linearmente independentes então $A\mathbf{x} = 0$ se e só se $\mathbf{x} = 0$, isto é, $N(A) = \{0\}$. Logo, $N(A^T A) = \{0\} = N(A)$.

Se $N(A^T A) = \{0\}$ então, pelo teorema da dimensão, a matriz quadrada $A^T A$ tem característica máxima e portanto é invertível. $\square$

Nota: No caso de A ser uma matriz quadrada invertível, a solução de mínimos quadrados para $A\mathbf{x} = \mathbf{y}$, dada pelo Teorema 1.2, reduz-se a $\hat{\mathbf{x}} = A^{-1}A^{-T}A^T\mathbf{y} = A^{-1}\mathbf{y}$. Neste caso, o sistema $A\mathbf{x} = \mathbf{y}$ não é impossível e tem solução única dada por $A^{-1}\mathbf{y}$, coincidente com a solução de mínimos quadrados.

A técnica descrita para ajustar uma recta a um conjunto de dados generaliza-se facilmente quando a curva é definida por um polinómio (como no caso da Figura 1-(b),(c)). Por exemplo, suponha-se que se pretende determinar o polinómio $y = a_0 + a_1 x + \cdots + a_k x^k$ que melhor se ajuste aos dados:

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Então substituindo em $y = a_0 + a_1 x + \cdots + a_k x^k$ os n valores de x e de y da tabela, obtemos n equações lineares

$$\begin{cases} a_0 + a_1 x_1 + \cdots + a_k x_1^k = y_1 \\ a_0 + a_1 x_2 + \cdots + a_k x_2^k = y_2 \\ \cdots\cdots\cdots \\ a_0 + a_1 x_n + \cdots + a_k x_n^k = y_n \end{cases} \iff A\mathbf{x} = \mathbf{y}.$$

Resolvendo, como anteriormente, a equação normal $A^T A\mathbf{x} = A^T\mathbf{y}$ obtemos os coeficientes a_i ($i = 0, \ldots, k$) do polinómio de mínimos quadrados.

Exemplo 2. Nos primeiros 5 meses do ano, as vendas (em milhares de euros) de uma certa empresa foram: 2, 2.2, 2.6, 3.2 e 4. Marcámos estes dados no gráfico abaixo e conjecturámos que as vendas podem ser aproximadas por um polinómio de grau 2.

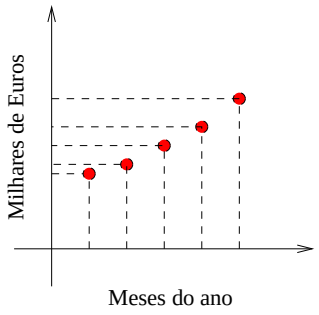

Figura 5: Dados do Exemplo 2.

Pretende-se determinar $y(x) = a_0 + a_1 x + a_2 x^2$ que melhor se ajuste aos dados obtidos, e usar esta função para deduzir qual o volume de vendas que a empresa irá realizar no último mês do ano.

Vamos portanto determinar a aproximação de mínimos quadrados para o sistema:

$$\begin{cases} a_0 + a_1 + a_2 = 2 \\ a_0 + 2a_1 + 4a_2 = 2.2 \\ a_0 + 3a_1 + 9a_2 = 2.6 \\ a_0 + 4a_1 + 16a_2 = 3.2 \\ a_0 + 5a_1 + 25a_2 = 4 \end{cases} \Longleftrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \mathbf{x} = \mathbf{y},$$

com $\mathbf{x} = \begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}^T$ e $\mathbf{y} = \begin{bmatrix} 2 & 2.2 & 2.6 & 3.2 & 4 \end{bmatrix}^T$.

A equação normal é:

$$A^T A\mathbf{x} = A^T\mathbf{y} \Longleftrightarrow \begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 14 \\ 47 \\ 185.4 \end{bmatrix}$$

A solução de mínimos quadrados para o sistema $A\mathbf{x} = \mathbf{y}$ é:

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 2 \\ -0.1 \\ 0.1 \end{bmatrix}$$

O polinómio procurado é $y(x) = 2 - 0.1x + 0.1x^2$. Assim, prevê-se que as vendas da empresa no último mês do ano seja de $y(12) = 15.2$ mil euros.

♦

2 Leitura suplementar: Aproximação de funções; Séries de Fourier.

Esta matéria não será leccionada nas aulas mas é incluída nestas notas a título informativo.

No espaço linear das funções reais contínuas no intervalo $[a, b]$ (abreviadamente $C([a, b])$) pode definir-se o seguinte produto interno:

$$\langle f\,, g\rangle = \int_a^b f(x)g(x)\,dx \tag{8}$$

Pretende-se aproximar uma dada função de $C([a, b])$ usando somente funções de um subespaço de $C([a, b])$. Por exemplo:

(i) Determinar a melhor aproximação de e^x no intervalo $[0, 1]$ pelo polinómio de segundo grau $a_0 + a_1x + a_2x^2$.

(ii) Determinar a melhor aproximação da função $f(x) = x$ no intervalo $[0, 2\pi]$ por uma função da forma

$$a_0 + a_1 \cos x + a_2 \cos(2x) + b_1 \operatorname{sen} x + b_2 \operatorname{sen}(2x).$$

O subespaço do exemplo (i) é o subespaço de $C([0,1])$ gerado por $\{1, x, x^2\}$ e para (ii) é o subespaço de $C([0,2\pi])$ gerado por $\{1, \cos x, \cos(2x), \operatorname{sen} x, \operatorname{sen}(2x)\}$.

A primeira questão que se coloca é a de saber o que se entende por "melhor aproximação". Como é sabido, tendo definido um produto interno num espaço linear, podemos sempre definir uma norma e consequentemente uma distância. No caso do produto interno (8), a norma e a distância entre duas funções são dadas respectivamente por:

$$\|f\|^2 = \int_a^b f^2(x)\,dx \quad \text{e} \quad \operatorname{dist}(f,g) = \|f-g\| = \sqrt{\int_a^b (f(x)-g(x))^2\,dx} \tag{9}$$

Pelo Teorema da melhor aproximação, a função de um dado subespaço S de $C([a,b])$ que melhor aproxima uma dada função f de $C([a,b])$ é a projecção ortogonal de f sobre S. Ou seja, a função $\hat{f} \in S$ que minimiza $\|f-g\|$ para qualquer $g \in S$ (Ver Figura 6).

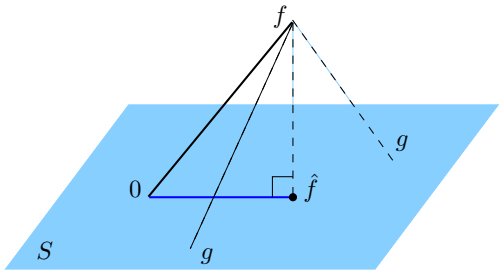

Figura 6: Melhor aproximação de uma função f por uma função do subespaço S.

O erro (global) cometido ao aproximar uma dada função f pela função $\hat{f}$ é dado por

$$e = \|f-\hat{f}\| = \int_a^b \left| f(x) - \hat{f}(x) \right| dx,$$

e é designado habitualmente por **erro médio quadrático**[1]. Geometricamente, o erro médio quadrático é a área da região situada entre os gráficos de f e de $\hat{f}$ (ver Figura 7).

Resumindo:

A solução do problema de aproximação de uma função $f \in C([a,b])$ por funções de um subespaço, de dimensão finita, S de $C([a,b])$, é a projecção ortogonal de f sobre S, isto é,

$$\hat{f} = \operatorname{proj}_S f.$$

Esta função minimiza o erro médio quadrático e é designada por **aproximação de mínimos quadrados** de f por funções de S.

Para determinar a projecção ortogonal de um vector sobre um subespaço de dimensão finita S, devemos encontrar uma base ortogonal para S.

[1] A razão da designação "erro médio quadrático"está relacionada com o Teorema do valor médio para integrais

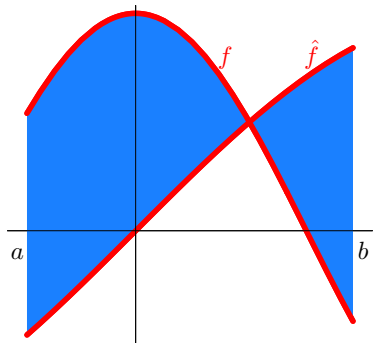

Figura 7: Erro médio quadrático.

No caso de $B = \{g_1, g_2, \ldots, g_k\}$ ser uma base ortogonal de S, então a expressão para a aproximação de mínimos quadrados de f é dada por

$$\begin{aligned} \hat{f} = \operatorname{proj}_S f &= \operatorname{proj}_{g_1} f + \operatorname{proj}_{g_2} f + \cdots + \operatorname{proj}_{g_k} f \\ &= \frac{\langle f, g_1\rangle}{\|g_1\|^2} g_1 + \frac{\langle f, g_2\rangle}{\|g_2\|^2} g_2 + \cdots + \frac{\langle f, g_k\rangle}{\|g_k\|^2} g_k. \end{aligned} \tag{10}$$

Recorde que, conhecida uma base para o subespaço S, o processo de ortogonalização de Gram-Schmidt permite obter uma base ortogonal de S.

Exemplo 3. Vejamos como obter uma base ortonormada para o subespaço de $C([0,1])$, S, gerado por $\{1, x, x^2\}$.

Designe-se as funções $1, x$ e x^2 por $h_1 = 1$, $h_2 = x$ e $h_3 = x^2$. Então,

$$\|h_1\|^2 = \int_0^1 1dx = 1 \qquad \Rightarrow \quad \|h_1\| = 1$$

$$\|h_2\|^2 = \int_0^1 x^2 dx = \tfrac{x^3}{3}\Big|_0^1 = \tfrac{1}{3}$$

$$\|h_3\|^2 = \int_0^1 x^4 dx = \tfrac{x^5}{5}\Big|_0^1 = \tfrac{1}{5}$$

Uma base ortogonal para S, seja $\{g_1, g_2, g_3\}$, obtida pelo processo de ortogonalização de Gram-Schmidt é:

$$\begin{aligned} g_1 &= h_1 = 1 \\ g_2 &= h_2 - \operatorname{proj}_{g_1} h_2 = h_2 - \tfrac{\langle h_2, g_1\rangle}{\|g_1\|^2} g_1 = h_2 - \langle h_2, g_1\rangle \\ &= x - \int_0^1 x\,dx = x - \tfrac{x^2}{2}\Big|_1^0 = x - \tfrac{1}{2} \Rightarrow \|g_2\|^2 = \tfrac{1}{12} \\ g_3 &= h_3 - \operatorname{proj}_{g_1} h_3 - \operatorname{proj}_{g_2} h_3 = h_3 - \tfrac{\langle h_3, g_1\rangle}{\|g_1\|^2} g_1 - \tfrac{\langle h_3, g_2\rangle}{\|g_2\|^2} g_2 \\ &= x^2 - \int_0^1 x\,dx - 12(x - \tfrac{1}{2}) \int_0^1 x^2 (x - \tfrac{1}{2})\,dx \\ &= x^2 - \tfrac{1}{2} - (x - \tfrac{1}{2}) = x^2 - x \Rightarrow \|g_3\|^2 = \tfrac{1}{30}. \end{aligned}$$

A aproximação de mínimos quadrados para a função e^x por funções de S é

$$\begin{aligned}\hat{f} = \operatorname{proj}_S f &= \langle e^x, 1\rangle + \frac{\langle e^x, (x-\frac{1}{2})\rangle}{\|x-\frac{1}{2}\|^2}\left(x-\frac{1}{2}\right) + \cdots + \frac{\langle e^x, x^2-x\rangle}{\|x^2-x\|^2}(x^2-x)\\ &\langle e^x, 1\rangle + 12\langle e^x, (x-\frac{1}{2})\rangle\left(x-\frac{1}{2}\right) + 30\langle e^x, x^2-x\rangle(x^2-x)\\ &= (e-1) + 6(3-e)(x-\frac{1}{2}) + 30(e-3)(x^2-x)\end{aligned}$$

♦

Séries de Fourier

Dada uma função contínua, f, no intervalo $[0, 2\pi]$, pretende-se aproximá-la por uma função t do tipo

$$t(x) = a_0 + a_1\cos x + \cdots + a_n\cos(nx) + b_1 \operatorname{sen} x + \cdots + b_n \operatorname{sen}(nx).$$

Isto é, pretende-se aproximar f por uma função do subespaço gerado por

$$V = \{1, \cos x, \cos(2x), \cdots, \cos(nx), senx, \operatorname{sen}(2x)\cdots, \operatorname{sen}(nx)\}.$$

A função t é conhecida pela designação de **polinómio trigonométrico** (de ordem n, se a_n e b_n são ambos não nulos).

Pode mostrar-se que V é um conjunto linearmente independente e, portanto, uma base para $W = \operatorname{Span} V$.

Para determinar a aproximação de mínimos quadrados de uma função $f \in C([0, 2\pi])$ por uma função de W, é conveniente encontrar uma base ortogonal (ou ortonormada) para W. Se $\{g_0, g_1, \cdots, g_n, h_1, \cdots, h_n\}$ for uma base ortonormada para W então a aproximação de mínimos quadrados para f será:

$$\begin{aligned}\hat{f} = \operatorname{proj}_W f &= \langle f, g_0\rangle g_0 + \langle f, g_1\rangle g_1 + \cdots + \langle f, g_n\rangle g_n + \langle f, h_1\rangle h_1 + \cdots + \langle f, h_n\rangle h_n\\ &= \tilde{a}_0 g_0 + a_1 g_1 + \cdots + a_n g_n + b_1 h_1 + \cdots + b_n h_n.\end{aligned}$$

Relativamente ao produto interno $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\,dx$, podemos verificar que V é uma base ortogonal de W. Normalizando V, isto é dividindo cada função de V pela sua norma, obtemos a base ortonormada

$$\begin{aligned}g_0 = \frac{1}{\sqrt{2\pi}}, g_1 = \frac{1}{\sqrt{\pi}}\cos x, \quad \cdots, g_n = \frac{1}{\sqrt{\pi}}\cos(nx)\\ h_1 = \frac{1}{\sqrt{\pi}}\operatorname{sen} x, \quad \cdots, h_n = \frac{1}{\sqrt{\pi}}\operatorname{sen}(nx).\end{aligned}$$

Logo, a aproximação de mínimos quadrados de f será

$$\begin{aligned}\hat{f} = \operatorname{proj}_W f &= \frac{a_0}{2} + [a_1\cos x + \cdots + a_n\cos(nx)] + [b_1\cos x + \cdots + b_n\cos(nx)]\\ &= \frac{a_0}{2} + \sum_{k=1}^{n}(a_k\cos(kx) + b_k\operatorname{sen}(kx))\end{aligned}$$

onde

$$
\begin{aligned}
a_0 &= \frac{2}{\sqrt{2\pi}}\langle f, g_0\rangle = \frac{2}{\sqrt{2\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}dx = \frac{1}{\pi}\int_0^{2\pi} f(x)dx \\
a_1 &= \frac{1}{\sqrt{\pi}}\langle f, g_1\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos x dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x dx \\
&\vdots \\
a_n &= \frac{1}{\sqrt{\pi}}\langle f, g_n\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos(nx) dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(nx) dx \\
b_1 &= \frac{1}{\sqrt{\pi}}\langle f, h_1\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\operatorname{sen} x dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\operatorname{sen} x dx \\
&\vdots \\
b_n &= \frac{1}{\sqrt{\pi}}\langle f, h_n\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\operatorname{sen}(nx) dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\operatorname{sen}(nx) dx
\end{aligned}
$$

Os coeficientes $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ são conhecidos como os **coeficientes de Fourier** de f.

Fixado n, o polinómio trigonométrico de ordem menor ou igual a n que aproxima a função f é:

$$f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos(kx) + b_k \operatorname{sen}(kx))$$

Sob certas condições sobre a função f, pode mostrar-se que o erro médio quadrático se aproxima de zero à medida que n tende para infinito. Isto significa que f pode identificar-se como sendo o limite da soma finita quando n tende para infinito, isto é,

$$f(x) = \frac{a_0}{2} + \lim_{n\to\infty} \sum_{k=1}^{n} (a_k \cos(kx) + b_k \operatorname{sen}(kx)) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \operatorname{sen}(kx)) .$$

A última igualdade é designada por **série de Fourier** de f no intervalo $[0, 2\pi]$.

Exemplo 4. Determinar a aproximação de mínimos quadrados de $f(x) = x$ em $[0, 2\pi]$, por um polinómio trigonométrico de ordem menor ou igual a 3.

Calculemos os coeficientes de Fourier $a_0, a_1, a_2, b_1, b_2, b_3$:

$$
\begin{aligned}
a_0 &= \frac{1}{\pi}\int_0^{2\pi} x\, dx = \frac{x^2}{2\pi}\Big|_0^{2\pi} = 2\pi & a_1 &= \frac{1}{\pi}\int_0^{2\pi} x\cos(x)\, dx = 0 \\
a_2 &= \frac{1}{\pi}\int_0^{2\pi} x\cos(2x)\, dx = 0 & & \\
b_1 &= \frac{1}{\pi}\int_0^{2\pi} x\operatorname{sen}(x)\, dx = -2 & b_2 &= \frac{1}{\pi}\int_0^{2\pi} x\operatorname{sen}(2x)\, dx = -1 \\
b_3 &= \frac{1}{\pi}\int_0^{2\pi} x\operatorname{sen}(3x)\, dx = -\frac{2}{3} & &
\end{aligned}
$$

Assim, a aproximação de $f(x) = x$ por um polinómio de ordem menor ou igual a três é:

$$x \approx \pi - 2\operatorname{sen} x - \operatorname{sen}(2x) - \frac{2}{3}\operatorname{sen}(3x).$$

Na Figura 8 encontram-se representados os polinómios trigonométricos, respectivamente de ordens menores ou iguais a 1, 2 e 3, que aproximam $f(x) = x$.

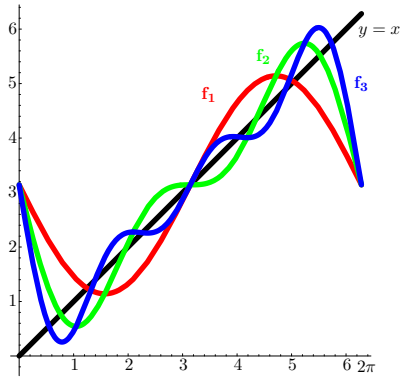

Figura 8: Aproximação de $f(x) = x$ pelos polinómios trigonométricos: $f_1 = \pi - 2\,\mathrm{sen}(x)$, $f_2 = \pi - 2\,\mathrm{sen}(x) - \mathrm{sen}(2x)$ e $f_3 = \pi - 2\,\mathrm{sen}(x) - \mathrm{sen}(2x) - \frac{2}{3}\,\mathrm{sen}(3x)$.

♦