

MAP3122: Métodos numéricos e aplicações
Quadrimestral 2020

Antoine Laurain

Método dos Mínimos Quadrados (MMQ)

Introdução

Encontramos dois tipos de problemas na teoria da aproximação:

1. Ajustar aos dados uma função pertencendo a uma classe determinada de funções, da “melhor” maneira possível.
2. Dada uma função de forma explícita, queremos encontrar uma função “mais simples” (por exemplo um polinômio) que aproxima essa função.

Vamos começar com o primeiro caso.

Aproximação discreta dos mínimos quadrados

- Sejam $\{(x_i, y_i)\}_{i=1}^n$ um conjunto de dados. Os x_i representam *experimentos* enquanto os y_i representam os resultados destes experimentos.
- Queremos achar uma função $x \mapsto g(x)$ que fornece a “melhor aproximação” do conjunto $\{(x_i, y_i)\}_{i=1}^n$. A noção de “melhor aproximação” pode ser interpretada de várias maneiras. Neste curso estudaremos dois métodos clássicos de aproximação:
 - ▶ **Interpolação:** neste abordagem, o gráfico da função g tem que passar por todos os pontos $\{(x_i, y_i)\}_{i=1}^n$, isto quer dizer que a condição $g(x_i) = y_i$ tem que ser satisfeito para todos os $i = 1, 2, \dots, n$.
 - ▶ **Método dos mínimos quadrados (MMQ):** com este método não exigimos que a condição $g(x_i) = y_i$, para todo $i = 1, 2, \dots, n$, esteja satisfeito, mas queremos minimizar as distâncias entre $g(x_i)$ e y_i . Isso deixa mais liberdade de escolher g em comparação com a interpolação.
- A classe mais básica de funções para fazer a aproximação usando MMQ é a classe de funções afins, da forma $g(x) = a_1 x + a_0$. O objetivo da aproximação então é de achar os coeficientes a_0, a_1 de maneira a minimizar as distâncias entre $g(x_i)$ e y_i .

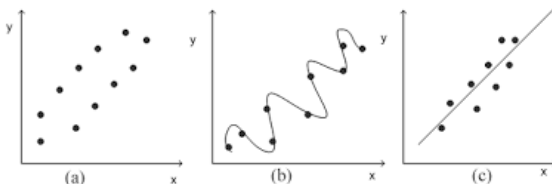


Figura: (a) nuvem de pontos (b) interpolação (c) MMQ

Aproximação discreta dos mínimos quadrados

- O problema de aproximação consiste então em minimizar um conjunto de distâncias $|g(x_i) - y_i|$. Para isso, existem varias possibilidades. Uma escolha natural é de minimizar o erro seguinte:

$$E_{\infty}(a_0, a_1) = \max_{1 \leq i \leq n} |y_i - g(x_i)| = \max_{1 \leq i \leq n} |y_i - (a_1 x_i + a_0)|.$$

Infelizmente este problema é difícil do ponto de vista da matemática e não pode ser tratado neste curso.

- Em vez de usar o máximo das distancias, podemos usar uma soma:

$$E_p(a_0, a_1) = \sum_{i=1}^n |y_i - g(x_i)|^p = \sum_{i=1}^n |y_i - (a_1 x_i + a_0)|^p.$$

- É natural considerar o caso particular $p = 2$, pois $E_2(a_0, a_1)$ tem boas propriedades de diferenciabilidade. Neste curso trabalharemos apenas com o caso $p = 2$, então o MMQ refere-se à minimização do erro:

$$E_2(a_0, a_1) = \sum_{i=1}^n (y_i - g(x_i))^2 = \sum_{i=1}^n (y_i - (a_1 x_i + a_0))^2.$$

- O caso $p = 1$ também é bastante usado, mas ele é difícil pois a função $z \mapsto |z|$ não é diferenciavel em $z = 0$.

Método dos Mínimos Quadrados (MMQ)

- Então o Método dos Mínimos Quadrados (MMQ) consiste em:

$$\begin{array}{ll} \text{minimize} & E_2(a_0, a_1) = \sum_{i=1}^n (y_i - \overbrace{(a_1 x_i + a_0)}^{g(x_i)})^2 \\ \text{com respeito a} & (a_0, a_1) \end{array}$$

- Ulteriormente veremos versões mais gerais do MMQ onde a função g pode depender de um número arbitrário de parâmetros $(a_0, a_1, \dots, a_m)$ e não apenas de dois parâmetros (a_0, a_1) .
- Primeiro observamos que o mínimo de uma função quadrática de uma variável está atingido quando a derivada da função vale zero. Isto é, o mínimo da função $g(a) = a^2$ está atingido quando $g'(a) = 2a = 0$.
- De maneira semelhante, podemos mostrar que o mínimo de $E_2(a_0, a_1)$ com respeito a (a_0, a_1) é atingido quando $\nabla E_2(a_0, a_1) = 0$, onde

$$\nabla E_2(a_0, a_1) = \begin{pmatrix} \frac{\partial E}{\partial a_0}(a_0, a_1) \\ \frac{\partial E}{\partial a_1}(a_0, a_1) \end{pmatrix}$$

e $\frac{\partial E}{\partial a_0}$, $\frac{\partial E}{\partial a_1}$ são as derivadas parciais de E com respeito a a_0 e a_1 , respectivamente.

Condição de otimalidade - sistema normal

- A condição $\nabla E_2(a_0, a_1) = 0$ chama-se **condição de otimalidade**. Vamos escrever essa condição de maneira mais explícita:

$$\frac{\partial E}{\partial a_0} = \frac{\partial}{\partial a_0} \sum_{i=1}^n (y_i - (a_1 x_i + a_0))^2 = -2 \sum_{i=1}^n y_i - (a_1 x_i + a_0) = 0$$

$$\frac{\partial E}{\partial a_1} = \frac{\partial}{\partial a_1} \sum_{i=1}^n (y_i - (a_1 x_i + a_0))^2 = -2 \sum_{i=1}^n x_i (y_i - (a_1 x_i + a_0)) = 0$$

- Obtemos

$$\frac{\partial E}{\partial a_0} = 0 \quad \Leftrightarrow \quad a_0 n + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\frac{\partial E}{\partial a_1} = 0 \quad \Leftrightarrow \quad a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

- Este sistema linear (para as incógnitas (a_0, a_1)) chama-se **sistema normal** e pode ser escrito na forma matricial seguinte:

Sistema Normal:

$$\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

Condição de otimalidade - sistema normal

- Agora vamos reescrever o sistema normal usando uma notação mais eficiente. Sejam

$$u = (1, 1, \dots, 1, 1) \in \mathbb{R}^n, \quad x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, \quad y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n.$$

- Definimos o produto escalar de x e y como: $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.
- Assim obtemos:

$$\begin{aligned} n &= \sum_{i=1}^n 1 = \langle u, u \rangle & \sum_{i=1}^n x_i &= \langle x, u \rangle \\ \sum_{i=1}^n x_i^2 &= \langle x, x \rangle & \sum_{i=1}^n x_i y_i &= \langle x, y \rangle \end{aligned}$$

- O sistema normal escreve-se então:

$$\begin{pmatrix} \langle u, u \rangle & \langle x, u \rangle \\ \langle x, u \rangle & \langle x, x \rangle \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \langle u, y \rangle \\ \langle x, y \rangle \end{pmatrix}$$

- Observamos que a matriz do sistema normal é simétrica, e que ela não depende de y .
- Este sistema tem a solução explícita seguinte:

$$a_0 = \frac{\langle y, u \rangle \langle x, x \rangle - \langle x, u \rangle \langle x, y \rangle}{\langle u, u \rangle \langle x, x \rangle - \langle x, u \rangle^2} \quad a_1 = \frac{\langle u, u \rangle \langle x, y \rangle - \langle u, y \rangle \langle u, x \rangle}{\langle u, u \rangle \langle x, x \rangle - \langle x, u \rangle^2}$$

Sistema normal - exemplo

Consideramos o resultado de um experimento:

x_i	0	1	2	3	4
y_i	0	1	1	4	4

Definimos os vetores correspondentes:

$$u = (1, 1, 1, 1, 1) \in \mathbb{R}^5, \quad x = (0, 1, 2, 3, 4) \in \mathbb{R}^5, \quad y = (0, 1, 1, 4, 4) \in \mathbb{R}^5.$$

O sistema normal é então:

$$\begin{pmatrix} \langle u, u \rangle & \langle x, u \rangle \\ \langle x, u \rangle & \langle x, x \rangle \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \langle u, y \rangle \\ \langle x, y \rangle \end{pmatrix} \Leftrightarrow \begin{pmatrix} 5 & 10 \\ 10 & 30 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 10 \\ 31 \end{pmatrix}$$

cuja solução é

$$a_0 = -\frac{1}{5}, \quad a_1 = \frac{11}{10}.$$

Assim

$$g(x) = \frac{11}{10}x - \frac{1}{5}$$

é a equação da reta que aproxima os pontos tabelados usando MMQ.

Método dos mínimos quadrados - caso discreto

- Seja $\{g_k\}_{k=0}^m$ uma família de funções dada. Queremos aproximar uma função f por $g = \sum_{k=0}^m a_k g_k$.
- O problema é de achar os coeficientes a_k que fornecem a “melhor” aproximação de f no sentido dos mínimos quadrados.
- Nesse caso a função f é tabelada em n pontos distintos $\{x_i\}_{i=1}^n$.
- Assim queremos minimizar a função

$$E(a_0, \dots, a_m) = \sum_{i=1}^n (f(x_i) - g(x_i))^2$$

- Já vimos o caso $m = 1$. No caso geral, precisamos também da condição de otimalidade

$$\frac{\partial E}{\partial a_i}(a_0^*, \dots, a_m^*) = 0 \quad \forall i = 0, \dots, m$$

onde $(a_0^*, \dots, a_m^*)$ é o ponto de mínimo.

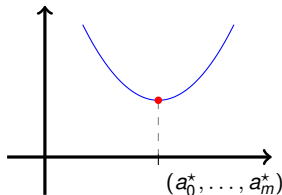


Figura: Mínimo local

Método dos mínimos quadrados - caso discreto

O cálculo fornece

$$\begin{aligned}\frac{\partial E}{\partial a_j}(a_0, \dots, a_m) &= \frac{\partial}{\partial a_j} \sum_{i=1}^n \left[f(x_i) - \sum_{k=0}^m a_k g_k(x_i) \right]^2 \\ &= 2 \sum_{i=1}^n \left[f(x_i) - \sum_{k=0}^m a_k g_k(x_i) \right] (-g_j(x_i)) = 0, \quad \forall j = 0, 1, \dots, m.\end{aligned}$$

ou seja

$$\sum_{i=1}^n \sum_{k=0}^m a_k g_k(x_i) g_j(x_i) = \sum_{i=1}^n f(x_i) g_j(x_i), \quad \forall j = 0, 1, \dots, m.$$

Usamos a notação vetorial

$$g_j = (g_j(x_1), g_j(x_2), \dots, g_j(x_n)) \quad \text{e} \quad f = (f(x_1), f(x_2), \dots, f(x_n))$$

e como consequência dessa notação temos $\sum_{i=1}^n f(x_i) g_j(x_i) = \langle f, g_j \rangle$ e portanto, podemos escrever

$$\sum_{i=1}^n \sum_{k=0}^m a_k g_k(x_i) g_j(x_i) = \sum_{k=0}^m a_k \underbrace{\sum_{i=1}^n g_k(x_i) g_j(x_i)}_{= \langle g_k, g_j \rangle}.$$

Método dos mínimos quadrados - caso discreto

Assim

$$\sum_{i=1}^n \sum_{k=0}^m a_k g_k(x_i) g_j(x_i) = \sum_{i=1}^n f(x_i) g_j(x_i), \quad \forall j = 0, 1, \dots, m$$

é equivalente ao sistema normal (a matriz é simétrica)

$$\begin{bmatrix} \langle g_0, g_0 \rangle & \cdots & \langle g_0, g_m \rangle \\ \vdots & \ddots & \vdots \\ \langle g_m, g_0 \rangle & \cdots & \langle g_m, g_m \rangle \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \langle g_0, f \rangle \\ \vdots \\ \langle g_m, f \rangle \end{bmatrix}.$$

A solução do sistema normal, se existe e é única, fornece o vetor $(a_0^*, \dots, a_m^*)$ que melhor aproxima a função f no sentido dos mínimos quadrados, isto é $f \approx \sum_{k=0}^m a_k^* g_k$.

Exemplo: Observamos um sinal osciloscópio “ f ”. Ele corresponde a superposição de efeitos, um oscilatório, e o outro crescente. Vamos aproximar f por uma função $g = a_0 g_0 + a_1 g_1$, onde $g_0(x) = x$, $g_1(x) = \cos(x)$. Os valores de f observados são

x	x_1	x_2	x_3	x_4	x_5
	0	1.5	3.0	4.5	6.0
$f(x)$	1.0	1.57	2.0	4.3	7.0

Vamos trabalhar com 4 algarismos significativos. Calculamos os coeficientes da matriz do sistema normal na forma

$$\langle g_0, g_0 \rangle = \sum_{i=1}^n x_i^2 = 0 + (1.5)^2 + (3)^2 + (4.5)^2 + (6)^2 = 67.5$$

$$\langle g_0, g_1 \rangle = \langle g_1, g_0 \rangle = \sum_{i=1}^n x_i \cos x_i = 1.948.$$

Calculamos também $\langle g_1, g_1 \rangle$, $\langle g_0, f \rangle$ e $\langle g_1, f \rangle$ e obtemos o sistema normal

$$\begin{bmatrix} 67.5 & 1.948 \\ 1.948 & 2.951 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 69.71 \\ 6.76 \end{bmatrix} \implies \begin{cases} a_0^* = 0.9855, \\ a_1^* = 1.64, \end{cases}$$

resultando em $g(x) = 0.9855x + 1.64 \cos(x)$.

Sistemas sobredeterminados

Vamos considerar uma aplicação do MMQ. Considere o sistema linear $Ax = y$, onde $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ e $A \in \mathbb{R}^{n \times m}$ (n linhas, m colunas) onde $n > m$.

Exemplo:

$$\begin{matrix} & \overset{m=2}{\longleftrightarrow} \\ \updownarrow n=3 & \left(\begin{pmatrix} 1 & 1 \\ 3 & 4 \\ 7 & 5 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 \\ 8 \\ 12 \end{pmatrix} \right) \iff \begin{cases} x_1 + x_2 = -2 \\ 3x_1 + 4x_2 = 8 \\ 7x_1 + 5x_2 = 12 \end{cases} \end{matrix}$$

Em geral, um sistema $Ax = y$ com $n > m$ não tem solução, porque tem demais equações a satisfazer para as m incógnitas x_i .

- Nesse caso, podemos procurar um ponto x^* tal que Ax^* seja “o mais perto possível” de y .
- Matematicamente, isto significa minimizar

$$J(x) = \|Ax - y\|_2^2 := \sum_{i=1}^n \left(\underbrace{\sum_{j=1}^m a_{ij}x_j}_{\text{coeficientes de } Ax-y} - y_i \right)^2.$$

Sistemas sobredeterminados

Para encontrar o mínimo de J , procuramos a solução de

$$\frac{\partial J}{\partial x_k}(x) = 0 \quad \text{para } k = 1, \dots, m.$$

Calculamos

$$\frac{\partial J}{\partial x_k}(x) = 2 \sum_{i=1}^n a_{ik} \underbrace{\left(\sum_{j=1}^n a_{ij} x_j - y_i \right)}_{= (Ax - y)_i} = 2 \langle A_k^T, Ax - y \rangle \quad \forall k = 1, \dots, m,$$

onde A_k é a k -ésima coluna de A . O gradiente $\nabla J(x)$ é o vetor das derivadas parciais de J . Assim temos

$$\begin{aligned} \frac{\partial J}{\partial x_k}(x) = 0 \quad \forall k = 1, \dots, m &\iff \nabla J(x) = 0 \\ &\iff \langle A_k^T, Ax - y \rangle = 0 \quad \forall k = 1, \dots, m \\ &\iff A^T(Ax - y) = 0 \\ &\iff \underbrace{A^T Ax = A^T y}_{\text{sistema normal}}. \end{aligned}$$

Sistemas sobredeterminados

Se $A^T A$ é invertível, então existe uma solução

$$x^* = (A^T A)^{-1} A^T y.$$

Observação: Note que, se $A \in \mathbb{R}^{n \times m}$ e $A^T \in \mathbb{R}^{m \times n}$, então $A^T A \in \mathbb{R}^{m \times m}$. Desta maneira, faz sentido considerar a existência da matriz inversa de $A^T A$.

Exemplo: Sejam $A = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, $x = (x_1)$, $y = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$ e $n = 2$, $m = 1$. Assim $A^T = (2 \ 3)$ e calculamos o valor de x^* na forma

$$\begin{aligned} x^* &= (A^T A)^{-1} A^T y = \left([2 \ 3] \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right)^{-1} [2 \ 3] \begin{bmatrix} 6 \\ 6 \end{bmatrix} \\ &= [13]^{-1} [30] = \frac{30}{13}. \end{aligned}$$

Calculamos $Ax^* = \begin{pmatrix} 60/13 \\ 90/13 \end{pmatrix} \approx \begin{pmatrix} 4,615 \\ 6,923 \end{pmatrix}$ que aproxima $y = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$.

Método dos mínimos quadrados - caso contínuo

- A diferença com o caso discreto é que conhecemos a forma analítica de f e não só alguns valores $f(x_1), f(x_2), \dots, f(x_n)$.
- Existem motivações para aproximar f com $\sum_{k=0}^m a_k g_k$, mesmo se f é conhecida analiticamente. Por exemplo, talvez f tem descontinuidades e queremos uma aproximação contínua.
- Queremos aproximar f no intervalo $[x_I, x_F]$. Precisamos minimizar o resíduo para todos os pontos $x \in [x_I, x_F]$. Assim, no caso discreto, usamos uma soma e no caso contínuo usamos uma integral:

$$E(a_0, \dots, a_m) = \int_{x_I}^{x_F} \underbrace{r(x)^2}_{\text{resíduo}} dx = \int_{x_I}^{x_F} (f(x) - g(x))^2 dx = \int_{x_I}^{x_F} \left(f(x) - \sum_{k=0}^m a_k g_k(x) \right)^2 dx.$$

Como nas seções anteriores, o ponto de mínimo satisfaz

$$\begin{aligned} \frac{\partial E}{\partial a_i}(a_0^*, \dots, a_m^*) &= 0 \quad \text{para todo } i = 0, \dots, m \\ \iff -2 \int_{x_I}^{x_F} \left(f(x) - \sum_{k=0}^m a_k g_k(x) \right) g_i(x) &= 0 \quad \text{para todo } i = 0, \dots, m. \end{aligned}$$

Método dos mínimos quadrados - caso contínuo

Vamos denotar $\langle f, g \rangle$ o produto escalar

$$\langle f, g \rangle = \int_{x_I}^{x_F} f(x)g(x) dx.$$

Obtemos assim

$$\frac{\partial E}{\partial a_i}(a_0^*, \dots, a_m^*) = 0 \iff \int_{x_I}^{x_F} f(x)g_i(x) dx = \sum_{k=0}^m a_k \int_{x_I}^{x_F} g_k(x)g_i(x) dx \quad \text{para todo } i = 0, \dots, m$$

$$\iff \langle f, g_i \rangle = \sum_{k=0}^m a_k \langle g_k, g_i \rangle \quad \text{para todo } i = 0, \dots, m$$

$$\iff \begin{bmatrix} \langle g_0, g_0 \rangle & \cdots & \langle g_0, g_m \rangle \\ \vdots & \ddots & \vdots \\ \langle g_m, g_0 \rangle & \cdots & \langle g_m, g_m \rangle \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \langle f, g_0 \rangle \\ \vdots \\ \langle f, g_m \rangle \end{bmatrix} \quad \text{sistema normal}.$$

Portanto, se a solução for única então $g(x) = \sum_{k=0}^m a_k^* g_k(x)$ será a melhor aproximação de $f(x)$ no sentido dos mínimos quadrados (onde $(a_0^*, \dots, a_m^*)$ é a solução do sistema normal).

Exemplo: Sejam $f(x) = e^x$ e $g(x) = a_0 g_0(x) + a_1 g_1(x)$ com $g_0(x) = 1$, $g_1(x) = x$. Para $x_I = 0$ e $x_F = 1$ definimos o produto escalar

$$\langle f, g \rangle = \int_0^1 f(x)g(x) dx.$$

Nestas condições, o sistema normal será dado por

$$\begin{bmatrix} \langle g_0, g_0 \rangle & \langle g_0, g_1 \rangle \\ \langle g_0, g_1 \rangle & \langle g_1, g_1 \rangle \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \langle f, g_0 \rangle \\ \langle f, g_1 \rangle \end{bmatrix}.$$

Calculamos os valores de cada uma das entradas do sistema.

$$\langle g_0, g_0 \rangle = \int_0^1 g_0^2 = \int_0^1 1 = 1,$$

$$\langle g_1, g_1 \rangle = \int_0^1 g_1^2 = \int_0^1 x^2 = \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3},$$

$$\langle g_0, g_1 \rangle = \int_0^1 g_0 g_1 = \int_0^1 x = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2},$$

$$\langle f, g_0 \rangle = \int_0^1 e^x = e - 1,$$

$$\langle f, g_1 \rangle = \int_0^1 x e^x = e^x (x - 1) \Big|_0^1 = 1$$

e substituímos estes valores no sistema normal

$$\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} e - 1 \\ 1 \end{bmatrix} \implies g(x) = \underbrace{(4e - 10)}_{a_0^*} + \underbrace{(18 - 6e)}_{a_1^*} x.$$

Famílias de funções não lineares nos parâmetros

No capítulo anterior aproximamos f por uma função $g(x) = \sum_{k=0}^m a_k g_k(x)$ que é linear com respeito aos parâmetros $(a_0, a_1, \dots, a_m)$.

- Podemos também procurar uma aproximação que não é linear nos parâmetros, como por exemplo

$$g(x) = \sum_{k=0}^m e^{a_k g_k(x)}.$$

- Esse caso é mais difícil, pois a condição de otimalidade

$$\frac{\partial E}{\partial a_i}(a_0^*, \dots, a_m^*) = 0 \quad \text{para todo } i = 0, \dots, m$$

não é um sistema linear.

- Um remédio é fazer a linearização nos parâmetros.

Exemplo: Seja f dada. Aproximamos f por uma família $a_0 e^{a_1 x}$. O termo $a_0 e^{a_1 x}$ não é linear nos parâmetros (a_0, a_1) . Vamos linearizar usando o logaritmo natural. Aproximamos $F(x) = \ln f(x)$ por uma função

$$\ln(a_0 e^{a_1 x}) = \ln(a_0) + a_1 x = b_0 + b_1 x \quad \text{com} \quad \begin{cases} b_0 = \ln(a_0) \\ b_1 = a_1 \end{cases}.$$

Famílias de funções não lineares nos parâmetros

Assim a função $b_0 + b_1 x$ é linear nos parâmetros, isto é, podemos resolver usando MMQ. Depois achamos a_0 fazendo $a_0 = e^{b_0}$. Isto nos dá a aproximação $g(x) = a_0 e^{a_1 x}$ de $f(x)$.

Exemplo: Queremos aproximar $f(x)$ por uma função

$$g(x) = \frac{a_1 + a_2 x + a_3 x^2}{1 + a_4 x + a_5 x^2}.$$

Aqui g não é linear nos parâmetros $(a_1, a_2, a_3, a_4, a_5)$. Para fazer a linearização nos parâmetros, primeiro podemos substituir a aproximação $f \approx g$ por a aproximação

$$(1 + a_4 x + a_5 x^2)f \approx a_1 + a_2 x + a_3 x^2,$$

e segundo substituir esta aproximação por

$$f(x) \approx a_1 + a_2 x + a_3 x^2 - a_4 x f(x) - a_5 x^2 f(x)$$

que agora é linear nos parâmetros.

- uma vez que achamos $(a_1, a_2, \dots, a_5)$ obtemos $g(x) = \frac{a_1 + a_2 x + a_3 x^2}{1 + a_4 x + a_5 x^2} \approx f(x)$.
- **Observação:** A linearização nos parâmetros é mais simples, mas não fornece a mesma aproximação que se fossemos resolver o problema não linear com mínimos quadrados. As vezes, essa solução pode divergir significativamente da aproximação por mínimos quadrados. Nesse caso, podemos achar uma solução melhor usando métodos mais avançados para resolver sistemas não lineares de equações $\frac{\partial E}{\partial a_i}(a_1^*, \dots, a_n^*) = 0$.

MMQ e polinômios ortogonais

Seja $f \in C^0([x_I, x_F])$. Vamos aproximar f usando um polinômio $g = \sum_{k=0}^m a_k g_k$, onde $g_k(x) = x^k$. O sistema normal é

$$\begin{bmatrix} \langle g_0, g_0 \rangle & \dots & \langle g_0, g_m \rangle \\ \vdots & & \vdots \\ \langle g_0, g_m \rangle & \dots & \langle g_m, g_m \rangle \end{bmatrix} \cdot \begin{bmatrix} a_0^* \\ \vdots \\ a_m^* \end{bmatrix} = \begin{bmatrix} \langle f, g_0 \rangle \\ \vdots \\ \langle f, g_m \rangle \end{bmatrix}$$

com

$$\begin{aligned} \langle g_i, g_j \rangle &= \int_{x_I}^{x_F} g_i(x) g_j(x) dx = \int_{x_I}^{x_F} x^{i+j} dx \\ \langle f, g_j \rangle &= \int_{x_I}^{x_F} f(x) g_j(x) dx \end{aligned}$$

Exemplo: Aproximar $f(x) = \sin(\pi x)$ com $g(x) = a_2 x^2 + a_1 x + a_0$. O sistema normal é

$$\begin{bmatrix} \int_0^1 1 & \int_0^1 x & \int_0^1 x^2 \\ \int_0^1 x & \int_0^1 x^2 & \int_0^1 x^3 \\ \int_0^1 x^2 & \int_0^1 x^3 & \int_0^1 x^4 \end{bmatrix} \cdot \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} = \begin{bmatrix} \int_0^1 \sin(\pi x) \\ \int_0^1 x \sin(\pi x) \\ \int_0^1 x^2 \sin(\pi x) \end{bmatrix}.$$

Resolvendo este sistema linear, obtemos

$$g(x) = -4.12251x^2 + 4.12251x - 0.050465$$

Enfrentamos dois problemas quando m é grande:

1. O custo computacional da resolução do sistema normal de dimensão $(m+1) \times (m+1)$ é alto quando m é grande.
2. Os coeficientes da matriz do sistema normal são

$$\langle g_i, g_j \rangle = \int_{x_l}^{x_F} x^{i+j} dx = \frac{x_F^{i+j+1} - x_l^{i+j+1}}{i+j+1}.$$

Este tipo de matriz chama-se **Matriz de Hilbert**. Matrizes de Hilbert são **mal condicionadas** quando m é grande. Consequentemente, vamos enfrentar erros de arredondamento altos ao resolver este sistema.

Como podemos mitigar estes problemas?

- Não precisamos escolher $g_k(x) = x^k$ em $g = \sum_{k=0}^m a_k g_k$.
- É possível achar uma família $\{g_k\}_{k=0}^m$ de polinômios tal que a matriz do sistema normal é diagonal, o que facilita bastante a resolução do problema.
- Para isso, precisamos achar uma família $\{g_k\}_{k=0}^m$ satisfazendo

$$\int_{x_l}^{x_F} g_i(x) g_j(x) dx = 0 \text{ quando } i \neq j.$$

Este tipo de polinômios chama-se **polinômios ortogonais**.

Definição: As funções $\{\phi_k\}_{k=0}^m$ são **linearmente independente (L.I.)** em $[x_I, x_F]$ se

$$\sum_{k=0}^m c_k \phi_k(x) = 0 \text{ para todos } x \in [x_I, x_F] \implies c_0 = c_1 = \dots = c_m = 0.$$

Se não for o caso, o conjunto $\{\phi_k\}_{k=0}^m$ é **linearmente dependente**.

Teorema: Se ϕ_j é um polinômio de grau j para cada $j = 0, \dots, m$, então $\{\phi_k\}_{k=0}^m$ é L.I. em $[x_I, x_F]$.

Teorema: Seja Π_m o conjunto de todos polinômios de grau $\leq m$. Se o conjunto $\{\phi_k\}_{k=0}^m$, com $\phi_k \in \Pi_m$, for L.I., então todo polinômio de Π_m escreve-se de maneira única como uma combinação linear de polinômios de $\{\phi_k\}_{k=0}^m$.

Definição: Uma função integrável ω chama-se **função peso** no intervalo $[x_I, x_F]$ se $\omega(x) \geq 0$ para todos $x \in [x_I, x_F]$, e se para qualquer subintervalo I de $[x_I, x_F]$, ω não for identicamente zero em I .

- Vamos considerar o problema de mínimos quadrados seguinte:

$$\min_{(a_0, \dots, a_m) \in \mathbb{R}^{m+1}} E(a_0, \dots, a_m) := \int_{x_I}^{x_F} \omega(x) (f(x) - g(x))^2 dx$$

com $f \in \mathcal{C}^0([x_I, x_F])$, $g(x) := \sum_{k=0}^m a_k g_k(x)$, onde as funções $\{g_k\}_{k=0}^m$ são L.I., e ω é uma função peso em $[x_I, x_F]$.

- Nós já calculamos no capítulo anterior o sistema normal no caso particular onde $\omega(x) = 1$ para todos $x \in [x_I, x_F]$.
- No caso geral, o sistema normal tem exatamente a mesma forma, mas o produto scalar é

$$\langle g_i, g_j \rangle = \int_{x_I}^{x_F} \omega(x) g_i(x) g_j(x) dx.$$

- Agora queremos escolher as funções $\{g_k\}_{k=0}^m$ de modo que a matriz do sistema normal seja diagonal, isto é

$$\langle g_i, g_j \rangle = \int_{x_I}^{x_F} \omega(x) g_i(x) g_j(x) dx = 0 \text{ para } i \neq j.$$

- Queremos também $\langle g_i, g_i \rangle > 0$.
- Se adicionalmente temos $\langle g_i, g_i \rangle = 1$ para todos $i = 0, \dots, m$, dizemos que o sistema é **ortonormal**.

- Se conseguirmos achar uma família $\{g_k\}_{k=0}^m$ que satisfaz estas propriedades, então a matriz do sistema normal é diagonal, e obtemos imediatamente a solução:

$$a_i^* = \frac{\langle g_i, f \rangle}{\langle g_i, g_i \rangle} = \frac{\int_{x_I}^{x_F} \omega(x) f(x) g_i(x) dx}{\int_{x_I}^{x_F} \omega(x) g_i(x)^2 dx} \quad \forall i = 0, \dots, m.$$

- Como achar uma família $\{g_k\}_{k=0}^m$ de polinômios ortogonais?
- O teorema seguinte fornece um procedimento recursivo para construir uma tal família. Este teorema pode ser provado por indução.

Teorema: O conjunto de polinômios $\{g_k\}_{k=0}^m$ seguinte é ortogonal em $[x_I, x_F]$, com respeito à função peso ω .

$$g_0(x) = 1, \quad \forall x \in [x_I, x_F],$$

$$g_1(x) = x - B_1, \quad \forall x \in [x_I, x_F], \text{ com } B_1 = \frac{\int_{x_I}^{x_F} x \omega(x) g_0(x)^2 dx}{\int_{x_I}^{x_F} \omega(x) g_0(x)^2 dx},$$

$$g_k(x) = (x - B_k)g_{k-1}(x) - C_k g_{k-2}(x), \quad \forall x \in [x_I, x_F], \text{ para } k \geq 2,$$

$$\text{com } B_k = \frac{\int_{x_I}^{x_F} x \omega(x) g_{k-1}(x)^2 dx}{\int_{x_I}^{x_F} \omega(x) g_{k-1}(x)^2 dx}, \quad C_k = \frac{\int_{x_I}^{x_F} x \omega(x) g_{k-1}(x) g_{k-2}(x) dx}{\int_{x_I}^{x_F} \omega(x) g_{k-2}(x)^2 dx}.$$

Observação: O conjunto de polinômios $\{g_k\}_{k=0}^m$ fornecido por este teorema é L.I. em $[x_I, x_F]$.

Exemplo: Vamos calcular polinômios ortogonais $\{g_k\}_{k=0}^m$ no intervalo $[-1, 1]$, com respeito à função peso $\omega \equiv 1$, usando este teorema. Calculamos

$$g_0(x) = 1, \quad \forall x \in [-1, 1],$$

$$B_1 = \frac{\int_{-1}^1 x\omega(x)g_0(x)^2 dx}{\int_{-1}^1 \omega(x)g_0(x)^2 dx} = \frac{\int_{-1}^1 x dx}{\int_{-1}^1 dx} = 0,$$

$$g_1(x) = x - B_1 = x,$$

$$B_2 = \frac{\int_{-1}^1 x\omega(x)g_1(x)^2 dx}{\int_{-1}^1 \omega(x)g_1(x)^2 dx} = \frac{\int_{-1}^1 x^3 dx}{\int_{-1}^1 x^2 dx} = 0,$$

$$C_2 = \frac{\int_{-1}^1 x\omega(x)g_1(x)g_0(x) dx}{\int_{-1}^1 \omega(x)g_0(x)^2 dx} = \frac{\int_{-1}^1 x^2 dx}{\int_{-1}^1 dx} = \frac{1}{3},$$

$$g_2(x) = (x - B_2)g_1(x) - C_2g_0(x) = x^2 - \frac{1}{3},$$

Observação: Os **polinômios de Legendre** $g_0(x) = 1$, $g_1(x) = x$, $g_2(x) = \frac{3x^2-1}{2}$ são também uma família de polinômios ortogonais no intervalo $[-1, 1]$, com respeito à função peso $\omega \equiv 1$. Cada polinômio de Legendre difere do polinômio g_k calculado neste exemplo apenas por uma constante multiplicativa.

Observação: Cada intervalo $[x_I, x_F]$ e peso ω fornecerá uma nova família de polinômios ortogonais $\{g_k\}_{k=0}^m$. Algumas famílias ficaram famosas e têm nomes, como os polinômios de Legendre, Hermite, Laguerre, ou Chebychev.

Exemplo: Vamos calcular polinômios ortogonais $\{g_k\}_{k=0}^m$ no intervalo $[0, 1]$, com respeito à função peso $\omega \equiv 1$, usando este teorema. Calculamos

$$g_0(x) = 1, \quad \forall x \in [0, 1],$$

$$B_1 = \frac{\int_0^1 x \omega(x) g_0(x)^2 dx}{\int_0^1 \omega(x) g_0(x)^2 dx} = \frac{\int_0^1 x dx}{\int_0^1 dx} = \frac{1}{2},$$

$$g_1(x) = x - B_1 = x - \frac{1}{2},$$

$$B_2 = \frac{\int_0^1 x \omega(x) g_1(x)^2 dx}{\int_0^1 \omega(x) g_1(x)^2 dx} = \frac{\int_0^1 x(x - 1/2)^2 dx}{\int_0^1 (x - 1/2)^2 dx} = \frac{1}{2},$$

$$C_2 = \frac{\int_0^1 x \omega(x) g_1(x) g_0(x) dx}{\int_0^1 \omega(x) g_0(x)^2 dx} = \frac{\int_0^1 x(x - 1/2) dx}{\int_0^1 dx} = \frac{1}{12},$$

$$g_2(x) = (x - B_2)g_1(x) - C_2 g_0(x) = x^2 - x + \frac{1}{6}.$$

Observamos que os polinômios g_1 e g_2 no caso $[x_I, x_F] = [0, 1]$ são diferentes do caso $[x_I, x_F] = [-1, 1]$ (compare com o exemplo anterior), mesmo que o peso ω seja o mesmo.

Exemplo: Vamos calcular polinômios ortogonais $\{g_k\}_{k=0}^m$ no intervalo $[0, \infty)$, com respeito à função peso $\omega(x) = e^{-x}$, usando o teorema. Neste caso o produto escalar é

$$\langle f, g \rangle = \int_0^{\infty} e^{-x} f(x) g(x) dx.$$

Esta integral sempre existe quando f e g são polinômios. Calculamos

$$g_0(x) = 1, \quad \forall x \in [0, \infty),$$

$$B_1 = \frac{\int_0^{\infty} x \omega(x) g_0(x)^2 dx}{\int_0^{\infty} \omega(x) g_0(x)^2 dx} = \frac{\int_0^{\infty} x e^{-x} dx}{\int_0^{\infty} e^{-x} dx} = \frac{[-x e^{-x}]_0^{\infty} - \int_0^{\infty} -e^{-x} dx}{[-e^{-x}]_0^{\infty}} = 1,$$

$$g_1(x) = x - B_1 = x - 1,$$

$$B_2 = \frac{\int_0^{\infty} x \omega(x) g_1(x)^2 dx}{\int_0^{\infty} \omega(x) g_1(x)^2 dx} = \frac{\int_0^{\infty} x e^{-x} (x - 1)^2 dx}{\int_0^{\infty} e^{-x} (x - 1)^2 dx} = 3,$$

$$C_2 = \frac{\int_0^{\infty} x \omega(x) g_1(x) g_0(x) dx}{\int_0^{\infty} \omega(x) g_0(x)^2 dx} = \frac{\int_0^{\infty} x e^{-x} (x - 1) dx}{\int_0^{\infty} e^{-x} dx} = 1,$$

$$g_2(x) = (x - B_2) g_1(x) - C_2 g_0(x) = x^2 - 4x + 2.$$

Observação: Os **polinômios de Laguerre** $g_0(x) = 1$, $g_1(x) = 1 - x$, $g_2(x) = \frac{x^2 - 4x + 2}{2}$ são também uma família de polinômios ortogonais no intervalo $[0, \infty)$, com respeito à função peso $\omega(x) = e^{-x}$. Cada polinômio de Laguerre difere do polinômio g_k calculado neste exemplo apenas por uma constante multiplicativa.

Polinômios ortogonais e mudança de variáveis

- As famílias de polinômios ortogonais conhecidas tais que os polinômios de Legendre, Hermite, Laguerre, ou Chebychev podem ser encontradas facilmente de forma tabelada, por exemplo no Wikipedia.
- Cada uma dessas famílias é definida para um peso ω e um intervalo $[x_I, x_F]$ particular, por exemplo $\omega(x) = e^{-x}$ e $[x_I, x_F] = [-1, 1]$ no caso de polinômios de Legendre.
- Em princípio, deveríamos recalcular uma família de polinômios ortogonais para cada novo intervalo $[x_I, x_F]$, mas podemos evitar esse cálculo usando uma mudança de variável, que permite usar famílias conhecidas de polinômios ortogonais.
- Vamos supor que a função f é definida em $[x_I, x_F]$, mas que dispomos de uma tabela de polinômios ortogonais apenas num outro intervalo $[t_I, t_F]$. Consideramos o problema de mínimos quadrados seguinte:

$$\min_{(a_0, \dots, a_m) \in \mathbb{R}^{m+1}} E(a_0, \dots, a_m) := \int_{x_I}^{x_F} (f(x) - g(x))^2 dx$$

with $g(x) = \sum_{k=0}^m a_k g_k(x)$.

- Vamos definir novas funções $F(t) := f(x(t))$ e $G(t) := g(x(t))$ definidas em $[t_I, t_F]$, onde $x(t) = \alpha t + \beta$ é uma mudança de variáveis satisfazendo (precisamos de $x([t_I, t_F]) = [x_I, x_F]$)

$$\begin{cases} x(t_I) = x_I, \\ x(t_F) = x_F, \end{cases} \iff \begin{cases} \alpha t_I + \beta = x_I, \\ \alpha t_F + \beta = x_F, \end{cases}$$

Assim, podemos calcular α e β resolvendo um sistema linear 2×2 .

- Usando a mudança de variáveis $x(t) = \alpha t + \beta$ na integral, obtemos, usando $dx = |x'(t)|dt$,

$$E(a_0, \dots, a_m) := \int_{t_l}^{t_F} (f(x(t)) - g(x(t)))^2 |x'(t)| dt = \alpha \int_{t_l}^{t_F} (F(t) - G(t))^2 dt$$

- Então vamos resolver o problema de minimização seguinte:

$$\min_{(a_0, \dots, a_m) \in \mathbb{R}^{m+1}} E(a_0, \dots, a_m) = \alpha \int_{t_l}^{t_F} (F(t) - G(t))^2 dt.$$

cuja solução é $G(t) = \sum_{k=0}^m a_k^* G_k(t)$, onde $G_k(t)$ são polinômios ortogonais conhecidos em $[t_l, t_F]$.

- Uma vez calculado $G(t)$, precisamos calcular $g(x) = G(t(x))$, onde $t(x)$ é a transformação inversa de $x(t)$, isto é $t(x)$ satisfaz $x = \alpha t(x) + \beta$, o que fornece $t(x) = \alpha^{-1}(x - \beta)$. Então obtemos a função g que aproxima f na forma seguinte:

$$g(x) = \sum_{k=0}^m a_k^* g_k(x) = G(t(x)) = \sum_{k=0}^m a_k^* G_k(\alpha^{-1}(x - \beta)).$$

Esse mostra também que $g_k(x) = G_k(\alpha^{-1}(x - \beta))$.

Exemplo (mudança de variáveis): Vamos aproximar $f(x) = \sin x$ em $[x_I, x_F] = [0, \pi]$ usando polinômios ortogonais de Legendre. Estes polinômios são dados em $[t_I, t_F] = [-1, 1]$ por

$$G_0(t) = 1, \quad G_1(t) = t, \quad G_2(t) = \frac{3t^2 - 1}{2}.$$

Calculamos $x(t) = \alpha t + \beta = \frac{\pi}{2}(t + 1)$ e verificamos que $x([t_I, t_F]) = [x_I, x_F] = [0, \pi]$. Então temos $F(t) = f(x(t)) = \sin(\frac{\pi}{2}(t + 1))$. Como os polinômios G_1, G_2, G_3 são ortogonais, o sistema normal é diagonal, e podemos calcular diretamente a solução (usando integração por partes)

$$a_0^* = \frac{\langle F, G_0 \rangle}{\langle G_0, G_0 \rangle} = \frac{\int_{-1}^1 \sin(\frac{\pi}{2}(t + 1)) dt}{\int_{-1}^1 dt} = \frac{2}{\pi}, \quad a_1^* = \frac{\langle F, G_1 \rangle}{\langle G_1, G_1 \rangle} = \frac{\int_{-1}^1 t \sin(\frac{\pi}{2}(t + 1)) dt}{\int_{-1}^1 t^2 dt} = 0,$$

$$a_2^* = \frac{\langle F, G_2 \rangle}{\langle G_2, G_2 \rangle} = \frac{10}{\pi} \left(1 - \frac{12}{\pi^2} \right).$$

Assim, obtemos a solução

$$G(t) = a_0^* G_0(t) + a_1^* G_1(t) + a_2^* G_2(t) = \frac{2}{\pi} + \frac{10}{\pi} \left(1 - \frac{12}{\pi^2} \right) \frac{3t^2 - 1}{2}.$$

Usando a transformação inversa $t(x) = \frac{2}{\pi}x - 1$, obtemos a função g que aproxima $f(x) = \sin x$ em $[x_I, x_F] = [0, \pi]$ na forma seguinte:

$$g(x) = \frac{2}{\pi} + \frac{10}{\pi} \left(1 - \frac{12}{\pi^2} \right) \frac{3(\frac{2}{\pi}x - 1)^2 - 1}{2}.$$

Análise harmônica e MMQ

- Neste capítulo vamos supor f periódica, e procuramos uma aproximação $g(x) = \sum_{k=0}^m a_k g_k(x)$ com funções g_k também periódicas, com o mesmo período que f .
- Uma classe bastante usada de funções periódicas, de período 2π , é $\{\cos kx, \sin kx\}_{k=0}^{\infty}$.
- Então seja f uma função periódica de período 2π , vamos aproximar $f(x)$, usando MMQ, por

$$g(x) = a_0 + \sum_{k=1}^m a_k \cos(kx) + b_k \sin(kx).$$

- Se o período de f for diferente de 2π , podemos usar uma mudança de variável para aproximar com $\{\cos kx, \sin kx\}_{k=0}^{\infty}$.
- Este tipo de aproximação chama-se **análise harmônica**.

Análise harmônica e MMQ: domínio contínuo

- O MMQ no contexto de funções periódicas de período 2π é:

$$\min_{(a_0, \dots, a_m, b_1, \dots, b_m) \in \mathbb{R}^{2m+1}} E(a_0, \dots, a_m, b_1, \dots, b_m) := \int_c^{c+2\pi} (f(x) - g(x))^2 dx$$

para qualquer constante c . É comum escolher $c = 0$ ou $c = -\pi$.

- Definimos também o produto escalar

$$\langle f, g \rangle = \int_c^{c+2\pi} f(x)g(x)dx.$$

- As funções $\{\cos kx, \sin kx\}_{k=0}^{\infty}$ são ortogonais com respeito a este produto escalar.
- Por exemplo para $k \neq l$ temos

$$\begin{aligned}\langle \cos(kx), \cos(lx) \rangle &= \int_0^{2\pi} \cos(kx) \cos(lx) dx \\ &= \int_0^{2\pi} \frac{\cos((k-l)x) + \cos((k+l)x)}{2} dx = 0.\end{aligned}$$

- Consequentemente, a matriz do sistema normal do MMQ usando as funções $\{\cos kx, \sin kx\}_{k=0}^{\infty}$ é diagonal.

- Então obtemos imediatamente a solução do MMQ como

$$a_0^* = \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx,$$

$$a_k^* = \frac{\langle f, \cos kx \rangle}{\langle \cos kx, \cos kx \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx,$$

$$b_k^* = \frac{\langle f, \sin kx \rangle}{\langle \sin kx, \sin kx \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx.$$

- Para obter estas formulas, usamos:

$$\langle \cos kx, \cos kx \rangle = \int_0^{2\pi} \cos(kx)^2 dx = \pi,$$

$$\langle \sin kx, \sin kx \rangle = \int_0^{2\pi} \sin(kx)^2 dx = \pi$$

Exemplo (análise harmônica e mudança de variáveis): Vamos fazer a análise harmônica da função F de período T definida por:

$$F(t) = \begin{cases} +1 & \text{em } [0, T/2], \\ -1 & \text{em } [T/2, T]. \end{cases}$$

Vamos usar a mudança de variáveis $t(x) = \frac{T}{2\pi}x$ e definir $f(x) := F(t(x))$. Então temos

$$f(x) = \begin{cases} +1 & \text{em } [0, \pi], \\ -1 & \text{em } [\pi, 2\pi]. \end{cases}$$

Aproximaremos f por

$$g(x) = a_0 + \sum_{k=1}^3 a_k \cos(kx) + b_k \sin(kx).$$

A solução do MMQ é

$$a_0^* = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx = \frac{1}{2\pi} \int_0^{\pi} dx - \frac{1}{2\pi} \int_{\pi}^{2\pi} dx = 0,$$

$$a_1^* = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(x) dx = \frac{1}{\pi} \int_0^{\pi} \cos(x) dx - \frac{1}{\pi} \int_{\pi}^{2\pi} \cos(x) dx = 0,$$

$$a_2^* = \frac{1}{\pi} \int_0^{\pi} \cos(2x) dx - \frac{1}{\pi} \int_{\pi}^{2\pi} \cos(2x) dx = 0,$$

$$a_3^* = \frac{1}{\pi} \int_0^{\pi} \cos(3x) dx - \frac{1}{\pi} \int_{\pi}^{2\pi} \cos(3x) dx = 0,$$

e para os parâmetros b_k^* :

$$b_1^* = \frac{1}{\pi} \int_0^\pi \sin(x) dx - \frac{1}{\pi} \int_\pi^{2\pi} \sin(x) dx = \frac{4}{\pi},$$

$$b_2^* = \frac{1}{\pi} \int_0^\pi \sin(2x) dx - \frac{1}{\pi} \int_\pi^{2\pi} \sin(2x) dx = 0,$$

$$b_3^* = \frac{1}{\pi} \int_0^\pi \sin(3x) dx - \frac{1}{\pi} \int_\pi^{2\pi} \sin(3x) dx = \frac{4}{3\pi}.$$

Então obtemos

$$g(x) = \frac{4}{\pi} \sin(x) + \frac{4}{3\pi} \sin(3x).$$

Usando a transformação inversa $x(t) = \frac{2\pi}{T} t$ obtemos a função $G(t)$ que aproxima $F(t)$ em $[0, T]$:

$$G(t) = g(x(t)) = \frac{4}{\pi} \sin\left(\frac{2\pi}{T} t\right) + \frac{4}{3\pi} \sin\left(\frac{6\pi}{T} t\right).$$

Análise harmônica e MMQ: domínio discreto

- No caso discreto da análise harmônica, consideramos $2n$ pontos $\{x_j\}_{j=0}^{2n-1}$ e uma função f tabelada com valores $\{f(x_j)\}_{j=0}^{2n-1}$ nestes pontos.
- Para facilitar a apresentação, vamos considerar uma discretização do intervalo $[-\pi, \pi]$ com pontos equidistantes e passo de discretização π/n , isto é $x_j = -\pi + \frac{j}{n}\pi$, para $j = 0, 1, \dots, 2n-1$, assim teremos $x_0 = -\pi$ e $x_{2n-1} = \pi - \frac{\pi}{n}$.
- **Hipótese importante:** suponhamos $m < n$.
- Vamos considerar o problema de mínimos quadrados seguinte:

$$\min_{(a_0, \dots, a_m, b_1, \dots, b_{m-1}) \in \mathbb{R}^{2m}} E(a_0, \dots, a_m, b_1, \dots, b_{m-1}) = \sum_{j=0}^{2n-1} (f(x_j) - g(x_j))^2$$

onde a função aproximadora tem a forma seguinte:

$$g(x_j) = \frac{a_0}{2} + a_m \cos(mx_j) + \sum_{k=1}^{m-1} a_k \cos(kx_j) + b_k \sin(kx_j).$$

- Observação: o fato de não colocar um termo $b_m \sin(mx_j)$ na função g é arbitrário. Alguns autores colocam este termo.

- Vamos definir $\phi_k(x) := \cos(kx)$ para $k = 1, \dots, m$, e $\phi_{m+\ell}(x) := \sin(\ell x)$ para $\ell = 1, \dots, m-1$. Definimos também o produto escalar de duas funções g e h definidas nos pontos $\{x_j\}_{j=0}^{2n-1}$ como:

$$\langle g, h \rangle := \sum_{j=0}^{2n-1} g(x_j)h(x_j).$$

- Podemos mostrar que as funções $\{\phi_i\}_{i=0}^{2m-1}$ são **ortogonais** com respeito a este produto escalar. Esta propriedade simplifica bastante a determinação da solução $(a_0^*, \dots, a_m^*, b_1^*, \dots, b_{m-1}^*)$.
- Para mostrar que $\{\phi_i\}_{i=0}^{2m-1}$ é uma família de funções ortogonais, usamos o resultado seguinte.

Lema: Se $r \in \mathbb{N}$ não for um múltiplo de $2n$, então

$$\sum_{j=0}^{2n-1} \cos(rx_j) = 0, \quad \text{e} \quad \sum_{j=0}^{2n-1} \sin(rx_j) = 0.$$

Se $r \in \mathbb{N}$ não for um múltiplo de n , então

$$\sum_{j=0}^{2n-1} (\cos(rx_j))^2 = n, \quad \text{e} \quad \sum_{j=0}^{2n-1} (\sin(rx_j))^2 = n.$$

Idéia da demonstração: escrever $e^{iz} = \cos(z) + i \sin(z)$ e calcular $\sum_{j=0}^{2n-1} e^{irx_j}$.

Usando este lema com $\phi_k(x) := \cos(kx)$ para $k = 1, \dots, m$, e $\phi_{m+\ell}(x) := \sin(\ell x)$ para $\ell = 1, \dots, m-1$, obtemos

$$\langle \phi_k, \phi_{m+\ell} \rangle = \sum_{j=0}^{2n-1} \cos(kx_j) \sin(\ell x_j) = \frac{1}{2} \sum_{j=0}^{2n-1} \sin((k+\ell)x_j) + \sin((k-\ell)x_j) = 0,$$

pois ambos $(k+\ell)$ e $(k-\ell)$ não são múltiplos de $2n$. De fato, usando à hipótese $m < n$, temos

$$\begin{aligned} -2n < 1 \leq k + \ell \leq 2m - 1 < 2n, \\ -2n < 2 - m \leq k - \ell \leq m - 1 < 2n. \end{aligned}$$

Teorema: Consideramos o problema de mínimos quadrados seguinte:

$$\min_{(a_0, \dots, a_m, b_1, \dots, b_{m-1}) \in \mathbb{R}^{2m}} E(a_0, \dots, a_m, b_1, \dots, b_{m-1}) = \sum_{j=0}^{2n-1} (f(x_j) - g(x_j))^2 dx$$

onde a função aproximadora tem a forma seguinte:

$$g(x_j) = \frac{a_0}{2} + a_m \cos(mx_j) + \sum_{k=1}^{m-1} a_k \cos(kx_j) + b_k \sin(kx_j).$$

A solução do problema de mínimos quadrados é:

$$a_k^* = \frac{1}{n} \sum_{j=0}^{2n-1} f(x_j) \cos(kx_j), \quad \forall k = 0, 1, \dots, m,$$

$$b_k^* = \frac{1}{n} \sum_{j=0}^{2n-1} f(x_j) \sin(kx_j), \quad \forall k = 1, 2, \dots, m-1.$$

Idéia da demonstração (a fazer em exercício): Escrever as condições de otimalidade

$$\frac{\partial E}{\partial a_k}(a_0^*, \dots, a_m^*, b_1^*, \dots, b_{m-1}^*) = 0, \quad \frac{\partial E}{\partial b_k}(a_0^*, \dots, a_m^*, b_1^*, \dots, b_{m-1}^*) = 0.$$

e depois usar que $\{\phi_i\}_{i=0}^{2m-1}$ é uma família de funções ortogonais.