# 3.4 曲线拟合的最小二乘法

朱升峰

华东师范大学 数学科学学院

2021 年 4 月

已知有关 $f(x)$ 的离散数据

| $x$ | $x_0$ | $x_1$ | $\dots$ | $x_m$ |
|---|---|---|---|---|
| $f(x)$ | $y_0$ | $y_1$ | $\dots$ | $y_m$ |

要找函数 $y = S^*(x)$ 与 $f(x)$ 最近，几何上看所给数据拟合。记误差

$$\delta = (\delta_0, \delta_1, \cdots, \delta_m)^{\mathrm{T}}$$

其中 $\delta_i = S^*(x_i) - y_i \ (i = 0, 1, \cdots, m)$.
设 $\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)$ 是线性无关函数族，
$\varphi = \mathrm{span}\{\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)\} \subset C[a, b]$

$$\|\delta\|_2^2 = \sum_{i=0}^{m} \delta_i^2 = \sum_{i=0}^{m} [S^*(x_i) - y_i]^2 = \min_{S \in \varphi} \sum_{i=0}^{m} [S(x_i) - y_i]^2,$$

其中

$$S(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x) \quad (n < m).$$

# 一般情形：加权平方和

$$\|\delta\|_2^2 = \sum_{i=0}^{m} \omega(x_i)\left[S(x_i) - f(x_i)\right]^2$$

$\omega(x) \geqslant 0$ 是 $[a, b]$ 上的权函数，表示不同点处的数据比重不同。
求 $S^*(x)$ 的该问题等价于

$$I(a_0, a_1, \cdots, a_n) = \sum_{i=0}^{m} \omega_i \left[\sum_{j=0}^{n} a_j \varphi_j(x_i) - f(x_i)\right]^2$$

$I(a_0, a_1, \cdots, a_n)$ 是关于 $a_0, a_1, \cdots, a_n$ 的多元二次函数。由多元函数求最小值的必要条件得

$$\frac{\partial I}{\partial a_k} = 0,\ \ k = 0, 1, \cdots, n,$$

即

$$\frac{\partial I}{\partial a_k} = 2\sum_{i=0}^{m} \omega_i \left[\sum_{j=0}^{n} a_j \varphi_j(x_i) - f(x_i)\right] \varphi_k(x_i) = 0, \qquad k = 0, 1, \cdots, n$$

记

$$(\varphi_k, \varphi_j) = \sum_{i=0}^{m} \omega(x_i)\varphi_k(x_i)\varphi_j(x_i) \quad k, j = 0, 1, \cdots, n.$$

$$(f, \varphi_k) = \sum_{i=0}^{m} \omega(x_i) f(x_i)\varphi_j(x_i) \equiv d_k, \quad k = 0, 1, \cdots, n.$$

关于 $a_0, a_1, \cdots, a_n$ 的线性方程组

$$\sum_{j=0}^{n} (\varphi_k, \varphi_j) a_j = d_k, \quad k = 0, 1, \cdots, n$$

称之为法方程

$$\begin{bmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) & \cdots & (\varphi_0, \varphi_n) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) & \cdots & (\varphi_1, \varphi_n) \\ \vdots & & \ddots & \vdots \\ (\varphi_n, \varphi_0) & (\varphi_n, \varphi_n) & \cdots & (\varphi_n, \varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_n \end{bmatrix}$$

$$\boxed{Ga = d}$$

$\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)$ 线性无关$\not\Leftrightarrow$系数矩阵行列式不为零?

## 定义 1

设 $\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x) \in C[a, b]$ 的任意线性组合在点集 $\{x_i\}_{i=0}^m (m \geqslant n)$ 上至多只有 $n$ 个不同的零点，则称 $\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)$ 在点集 $\{x_i\}_{i=0}^m$ 满足 Haar 条件.

## 定理 2

$\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x) \in C[a,b]$ 在点集 $\{x_i\}_{i=0}^m$ 满足 Haar 条件，则法方程系数矩阵非奇异，即法方程存在唯一解 $a_k = a_k^*$ $(k = 0, 1, \cdots, n)$。

- $1, x, \cdots, x^n$ 在点集 $\{x_i\}_{i=0}^m$ 上满足 Haar 条件。

令

$$S^*(x) = a_0^*\varphi_0(x) + a_1^*\varphi_1(x) + \cdots + a_n^*\varphi_n(x)$$

$S^*(x)$ 是 $f(x)$ 在 $\Phi$ 中的唯一最佳平方逼近函数，令 $\delta(x) = f(x) - S^*(x)$, 则最佳平方逼近的误差为

$$\|\delta(x)\|_2^2 = \sum_{k=0}^{n} a_k^* \left(\varphi_k(x), f(x)\right).$$

若取 $\varphi_k(x) = x^k, \rho(x) \equiv 1, f(x) \in C[0,1]$，则要在 $H_n$ 中求 $n$ 次最佳平方逼近多项式

$$S^*(x) = a_0^* + a_1^*x + \cdots + a_n^*x^n$$

- 当 $n$ 较大时，Hilbert 矩阵高度病态，选用 $\{1, x, \cdots, x^n\}$ 作基直接求解较困难，通常采用正交基。

**例** 求下列数据的拟合曲线

| $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $f_i$ | 4 | 4.5 | 6 | 8 | 8.5 |
| $\omega_i$ | 2 | 1 | 3 | 1 | 1 |

**解** 画散点图看到各点在一条直线附近，可用线性函数作拟合曲线。
$S_1(x) = a_0 + a_1 x$，$m = 4$，$n = 1$，$\varphi_0(x) = 1$，$\varphi_1(x) = x$。

$$(\varphi_0, \varphi_0) = \sum_{i=0}^{4} \omega_i = 8, \quad (\varphi_0, \varphi_1) = (\varphi_1, \varphi_0) = \sum_{i=0}^{4} \omega_i x_i = 22$$

$$(\varphi_1, \varphi_1) = \sum_{i=0}^{4} \omega_i x_i^2 = 74, \quad (\varphi_0, f) = \sum_{i=0}^{4} \omega_i f_i = 47$$

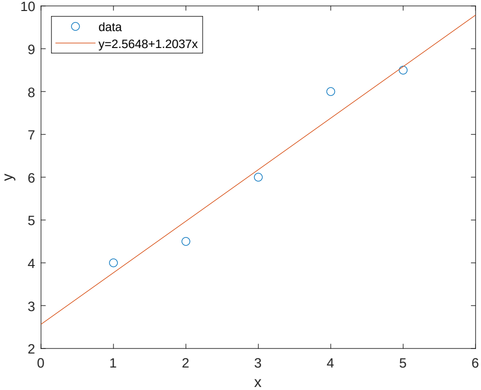
$$(\varphi_1, f) = \sum_{i=0}^{4} \omega_i x_i f_i = 145.5$$

解线性方程组

$$\begin{cases} 8a_0 + 22a_1 = 47 \\ 22a_0 + 74a_1 = 145.5 \end{cases}$$

解出 $a_0 = 0.934, \ a_1 = 0.426$, 故

$$S_1^*(x) = 2.5648 + 1.2037x.$$



# 非线性最小二乘拟合

有时需要非线性函数，如指数函数、对数函数、幂函数等拟合给定的数据，这时建立的法方程是一个非线性方程组，这类拟合问题称为非线性最小二乘拟合。

**思路**：作变量代换，转化为线性拟合

- $S(x) = ae^{bx}$

  令 $\bar{y} = \ln y$, 则 $\bar{y}(x) = bx + A$，其中 $A = \ln a$，即 $a = \mathrm{e}^A$.

- $S(x) = a\ln x + b$

  令 $\bar{x} = \ln x$, 则 $y(x) = a\bar{x} + b$

- $S(x) = ax^b$

  令 $\bar{y} = \ln y$, $\bar{x} = \ln x$, 则 $\bar{y}(x) = b\bar{x} + A$，其中 $A = \ln a$，即 $a = \mathrm{e}^A$.

**例** 设数据 $(x_i, y_i)\,(i=0,1,2,3,4)$ 由表给出. 表第 4 行为 $\ln y_i = \bar{y}_i$, 画散点图确定模型为 $y = a\mathrm{e}^{bx}$，用最小二乘法确定 $a$ 和 $b$。

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y_i$ | 1.00 | 1.25 | 1.50 | 1.75 | 2.00 |
| $y_i$ | 5.10 | 5.79 | 6.53 | 7.45 | 8.46 |
| $\bar{y}$ | 1.629 | 1.756 | 1.876 | 2.008 | 2.135 |

**解** 画散点图确定曲线方程为 $y = a\mathrm{e}^{bx}$，两边取对数得 $\ln y = \ln a + bx$, 若令 $\bar{y} = \ln y$, $A = \ln a$, 则得 $\bar{y} = A + bx$, 确定 $A, b$, 先将 $(x_i, y_i)$ 转化为 $(x_i, \bar{y}_i)$

$$(\varphi_0, \varphi_0) = 5, \quad (\varphi_0, \varphi_1) = \sum_{i=0}^{4} x_i = 7.5, \quad (\varphi_1, \varphi_1) = \sum_{i=0}^{4} x^i = 11.875,$$

$$(\varphi_0, \bar{y}) = \sum_{i=0}^{4} \bar{y}_i = 9.404, \quad (\varphi_1, \bar{y}) = \sum_{i=0}^{4} x_i \bar{y}_i = 14.422.$$

法方程

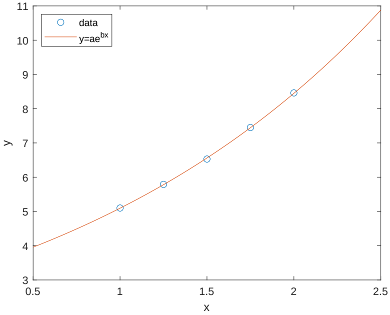
$$\begin{cases} 5A + 7.50b = 9.404 \\ 7.50A + 11.875b = 14.422 \end{cases}$$

解得

$$A = 1.122, b = 0.505, a = \mathrm{e}^A = 3.071$$

最小二乘法拟合曲线为

$$y = 3.071\mathrm{e}^{0.505x}$$



# 用正交多项式作最小二乘拟合

若 $\{\varphi_j(x)\}_{j=0}^n$ 是关于点集带权正交的函数族，即

$$(\varphi_j(x), \varphi_k(x)) = \begin{cases} 0, & j \neq k \\ A_k > 0, & j = k \end{cases}$$

此时法方程的系数矩阵 $G$ 为非奇异对角阵，法方程解为

$$a_k^* = \frac{(f, \varphi_k)}{(\varphi_k, \varphi_k)}, \quad k = 0, 1, \cdots, n.$$

最佳平方逼近函数为

$$S^*(x) = \sum_{k=0}^{n} \frac{(f(x), \varphi_k(x))}{\|\varphi_k(x)\|_2^2} \varphi_k(x).$$

最佳平方逼近的误差为

$$\|\delta(x)\|_2^2 = \|f(x)\|_2^2 - \sum_{k=0}^{n} \frac{(\varphi_k, f)^2}{(\varphi_k, \varphi_k)}$$

# 正交多项式族

根据点集 $\{x_i\}_{i=0}^m$ 及权函数 $\omega(x)$，用递推公式构造带权正交的多项式 $\{p_k(x)\}_{k=0}^n$:

$$\begin{cases} p_0(x) = 1, \quad p_1(x) = x - \alpha_0 \\ p_{k+1}(x) = (x - \alpha_k)p_k(x) - \beta_k p_{k-1}(x), \quad k = 1, 2, \cdots, n-1 \end{cases}$$

$$\begin{cases} \alpha_k = \dfrac{(xp_k, p_k)}{(p_{k-1}, p_{k-1})}, \quad k = 0, 1, 2, \cdots, n-1 \\ \beta_k = \dfrac{(p_k, p_k)}{(p_{k-1}, p_{k-1})}, \quad k = 1, 2, \cdots, n-1 \end{cases}$$

可以证明 $\{p_k(x)\}$ 正交。
由正交多项式作最小二乘拟合，可计算出

$$a_k^* = \frac{(f, p_k)}{(p_k, p_k)} = \frac{\sum_{i=0}^m \omega(x_i) f(x_i) p_k(x_i)}{\sum_{i=0}^m \omega(x_i) p_k^2(x_i)}, \quad k = 0, 1, \cdots, n.$$

代入

$$S^*(x) = a_0^* p_0(x) + a_1^* p_1(x) + \cdots + a_n^* p_n(x).$$

**例** 给定数据表，求二次最小二乘拟合多项式

| $x_i$ | 0 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|
| $y_i$ | 1.00 | 1.75 | 1.96 | 2.19 | 2.44 | 2.71 | 3.00 |

**解**

$$(p_0, p_0) = \sum_{i=0}^{6} 1 = 7, (f, p_0) = \sum_{i=1}^{6} y_i = 15.05, \quad (xp_0, p_0) = \sum_{i=1}^{6} x_i = 4.5$$

$$a_0^* = (f, p_0) \,/\, (p_0, p_0) \approx 2.15, \quad \alpha_0 = (xp_0, p_0) \,/\, (p_0, p_0) \approx 0.64$$

$$p_1(x) = x - 0.64 \quad a_1^* \approx 1.98, \quad \alpha_1 \approx 0.36, \quad \beta_1 \approx 0.094$$

$$p_2(x) = x^2 - 0.98x + 0.12 \quad a_2^* \approx 1.00$$

$$S_2^*(x) = a_0^* p_0(x) + a_1^* p_1(x) + a_2^* p_2(x) = x^2 + x + 1$$

- 正交多项式是目前为止多项式拟合最好的方法：

  不用解线性方程组，根据次数 $n$，只用递推公式方便地计算正交多项式。

- MATLAB 正交多项式最小二乘拟合函数:

  $$\mathrm{p} = \mathrm{polyfit}\ (\mathrm{x,y,n})$$

  输入参数 x:=$[x_0, x_1, \cdots, x_m]$; y:=$[y_0, y_1, \cdots, y_m]$

  返回次数为 $n$ 的多项式 $p(x)$ 的系数。

  $p$ 中的系数按降幂排列，$p$ 的长度为 $n+1$。

- 见exp32.m