第三章 最小二乘问题的解法

—理论, 正交变换

杜磊
dulei@dlut.edu.cn

大连理工大学 数学科学学院
创新园大厦 B1207

2019 年 10 月 8 日

内容提要

最小二乘问题

定理 (多项式插值)

假设 $x_1, x_2, \ldots, x_m \in \mathbb{R}$ 互不相等，$y_1, y_2, \ldots, y_m \in \mathbb{R}$ 为对应的函数值. 则存在满足 $p(x_i) = y_i \, (i = 1, \ldots, m)$ 的 $m-1$ 次多项式

$$p(x) = c_{m-1} x^{m-1} + \cdots + c_1 x + c_0.$$

定理 (多项式插值)

假设 $x_1, x_2, \ldots, x_m \in \mathbb{R}$ 互不相等，$y_1, y_2, \ldots, y_m \in \mathbb{R}$ 为对应的函数值. 则存在满足 $p(x_i) = y_i\,(i = 1, \ldots, m)$ 的 $m-1$ 次多项式

$$p(x) = c_{m-1}x^{m-1} + \cdots + c_1 x + c_0.$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{m-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{m-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{m-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}.$$

例 (龙格 (Runge) 函数)

$$f(x) = \frac{1}{1+25x^2},\ x \in [-1, 1].$$

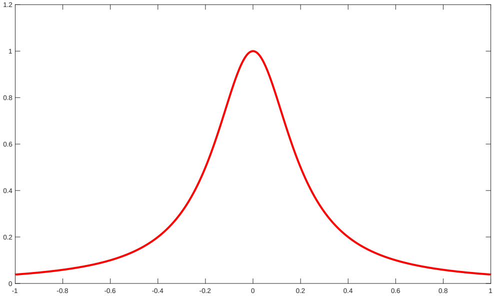
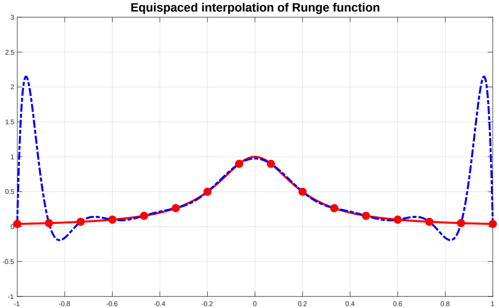


Fig: Runge function

例 (龙格 (Runge) 函数)

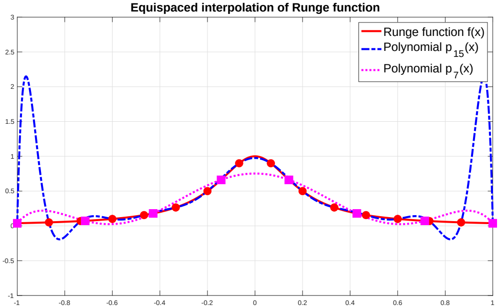
$$f(x) = \frac{1}{1+25x^2},\ x \in [-1, 1].$$



Equispaced interpolation of Runge function with 16 points

例 (龙格 (Runge) 函数)

$$f(x) = \frac{1}{1+25x^2},\ x \in [-1,1].$$



Equispaced interpolation of Runge function with 16 and 8 points

拟合多项式

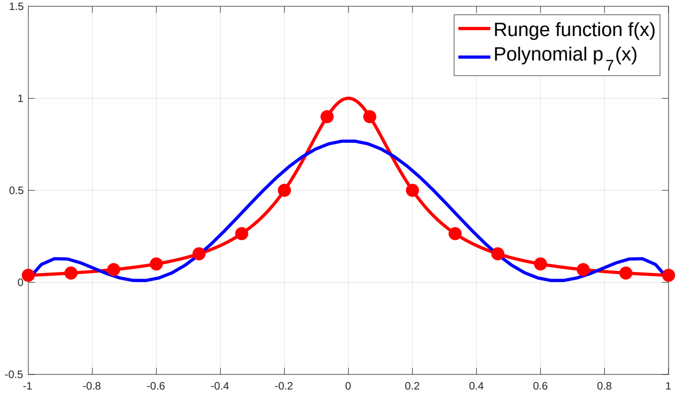


图: 使用 16 个等距点构造 7 次多项式拟合龙格函数.

定义

给定矩阵 $A \in \mathbb{R}^{m\times n}$ 及向量 $b \in \mathbb{R}^m$, 确定 $x \in \mathbb{R}^n$, 使得

$$\|b - Ax\|_2 = \|r(x)\|_2 = \min_{y\in\mathbb{R}^n} \|r(y)\|_2 = \min_{y\in\mathbb{R}^n} \|b - Ay\|_2 .$$

这就是所谓的最小二乘问题, 简称 LS (Least Squares) 问题, 其中的 $r(x)$ 常常被称为残向量.

最小二乘问题

定义

给定矩阵 $A \in \mathbb{R}^{m\times n}$ 及向量 $b \in \mathbb{R}^m$, 确定 $x \in \mathbb{R}^n$, 使得

$$\|b - Ax\|_2 = \|r(x)\|_2 = \min_{y\in\mathbb{R}^n} \|r(y)\|_2 = \min_{y\in\mathbb{R}^n} \|b - Ay\|_2 .$$

这就是所谓的最小二乘问题, 简称 LS (Least Squares) 问题, 其中的 $r(x)$ 常常被称为残向量.

在所讨论的最小二乘问题中,

- 若 r 线性地依赖于 x, 则称其为线性最小二乘问题;
- 若 r 非线性地依赖于 x, 则称其为非线性最小二乘问题.

最小二乘解

最小二乘问题的解 x 又可称做线性方程组

$$Ax = b,\ \ A \in \mathbb{R}^{m\times n} \tag{1}$$

的最小二乘解, 即 x 在残向量 $r(x) = b - Ax$ 的 2-范数最小的意义下满足方程组 (1).

线性方程组 (1) 分类:

- 若 $m > n$, 称 (1) 为超定方程组或矛盾方程组;
- 若 $m < n$, 称 (1) 为欠定方程组.

学习内容

主要讨论情况: $m > n, \mathrm{rank}(A) = n$ 的最小二乘问题的性质与求解.

基本概念

记 $A \in \mathbb{R}^{m\times n}$, A 的值域定义为

$$\mathcal{R}(A) = \{y \in \mathbb{R}^m : \ y = Ax,\ x \in \mathbb{R}^n\}.$$

记 $A \in \mathbb{R}^{m\times n}$, A 的零空间定义为

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : \ Ax = 0\}.$$

一个子空间 $\mathcal{S} \subset \mathbb{R}^n$ 的正交补定义为

$$\mathcal{S}^{\perp} = \{y \in \mathbb{R}^n : \ y^{\mathrm{T}} x = 0, \forall x \in \mathcal{S}\}.$$

线性方程组解的存在性

定理

方程组 $Ax = b,\ A \in \mathbb{R}^{m\times n}$ 的解存在的充分必要条件是

$$\operatorname{rank}(A) = \operatorname{rank}([A, b]).$$

定理

假定方程组 $Ax = b,\ A \in \mathbb{R}^{m\times n}$ 的解存在，并且假定 x 是其任一给定的解, 则方程组的全部解的集合是

$$x + \mathcal{N}(A).$$

推论

方程组 $Ax = b,\ A \in \mathbb{R}^{m\times n}$ 的解唯一的充分必要条件是

$$\mathcal{N}(A) = 0.$$

定理

线性最小二乘问题 $\|b - Ax\|_2 = r(x) = \min_{y \in \mathbb{R}^n} \|b - Ay\|_2$ 的解总是存在的，而且其解唯一的充分必要条件是 $\mathcal{N}(A) = 0$.

记最小二乘问题的解集为 $\mathcal{X}_{\mathrm{LS}}$，即

$$\mathcal{X}_{\mathrm{LS}} = \{x \in \mathbb{R}^n, \|b - Ax\|_2 = \min_{y \in \mathbb{R}^n} \|b - Ay\|_2\}.$$

最小 2-范数解

解集 $\mathcal{X}_{\mathrm{LS}}$ 中有且仅有一个解其 2-范数最小.

定理

$x \in \mathcal{X}_{\mathrm{LS}}$ 的充分必要条件是

$$A^{\mathrm{T}} A x = A^{\mathrm{T}} b.$$

正则化方程组 (法方程组)、Moore-Penrose 广义逆

称 $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$ 为最小二乘问题的正则化方程组或法方程组.

Algorithm 1 正则化方法

1: 计算 $C = A^{\mathrm{T}}A,\ d = A^{\mathrm{T}}b$;
2: Cholesky 分解: $C = L^{\mathrm{T}}L$;
3: 解方程组: $Ly = d, L^{\mathrm{T}}x = y$.

定义

若 $X \in \mathbb{R}^{m\times n}$ 满足 $AXA = A,\ XAX = X,\ (AX)^{\mathrm{T}} = AX,\ (XA)^{\mathrm{T}} = XA$, 则称 X 是 A 的 Moore-Penrose 广义逆, 通常记做 $A^{\dagger}$.

最小二乘问题的解 x 可表示为 $x = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b$, 可验证 $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 为 A 的 Moore-Penrose 广义逆.

增广方程组

求解最小二乘问题等价于求如下增广方程组

$$\begin{bmatrix} I & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

相对误差分析

定理

设 b_1 和 $\tilde{b}_1$ 分别是 b 和 $\tilde{b}$ 在 $\mathcal{R}(A)$ 上的正交投影. 若 $b_1 \neq 0$, 则

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \kappa_2(A) \frac{\left\|b_1 - \tilde{b}_1\right\|_2}{\|b_1\|_2},$$

其中 $\kappa_2(A) = \|A\|_2 \left\|A^\dagger\right\|_2$.

称 $\kappa_2(A)$ 为最小二乘问题的条件数. 若 $\kappa(A)$ 很大, 则称最小二乘问题是病态的; 反之, 若 $\kappa(A)$ 很小, 则称最小二乘问题是良态的.

定理

设 A 的列向量线性无关, 则 $\kappa_2(A)^2 = \kappa_2(A^\mathrm{T} A)$.

例子

考虑解 $Ax = b$, 其中

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & & \\ & \epsilon & \\ & & \epsilon \end{bmatrix}, \ b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

例子

考虑解 $Ax = b$, 其中

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & & \\ & \epsilon & \\ & & \epsilon \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

计算得

$$A^{\mathrm{T}}A = \begin{bmatrix} 1+\epsilon^2 & 1 & 1 \\ 1 & 1+\epsilon^2 & 1 \\ 1 & 1 & 1+\epsilon^2 \end{bmatrix}, \quad A^{\mathrm{T}}b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

例子

考虑解 $Ax = b$, 其中

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & & \\ & \epsilon & \\ & & \epsilon \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

计算得

$$A^{\mathrm{T}}A = \begin{bmatrix} 1+\epsilon^2 & 1 & 1 \\ 1 & 1+\epsilon^2 & 1 \\ 1 & 1 & 1+\epsilon^2 \end{bmatrix}, \quad A^{\mathrm{T}}b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

解得

$$x = \frac{1}{3+\epsilon^2}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad r = \frac{1}{3+\epsilon^2}\begin{bmatrix} \epsilon^2 \\ -1 \\ -1 \\ -1 \end{bmatrix}.$$

Householder Symposium

Previously called the Gatlinburg Symposia, now named in honor of its founder, Alston Householder, a pioneer of numerical linear algebra.

图: Jim Wilkinson, Wallace Givens, George Forsythe, Alston Householder, Peter Henrici, and Fritz Bauer. (Matlab: load gatlin, image(X), colormap(map))

定义

设 $w \in \mathbb{R}^{n\times n}$ 满足 $\|w\|_2 = 1$ 定义 H 为

$$H = I - 2ww^{\mathrm{T}},$$

则称 H 为 Householder 变换 (也称初等反射矩阵或镜像变换).

Householder 变换

定义

设 $w \in \mathbb{R}^{n\times n}$ 满足 $\|w\|_2 = 1$ 定义 H 为

$$H = I - 2ww^{\mathrm{T}},$$

则称 H 为 Householder 变换 (也称初等反射矩阵或镜像变换).

定理

Householder 变换 H 满足以下性质:

1. 对称性: $H^{\mathrm{T}} = H$;
2. 正交性: $H^{\mathrm{T}}H = I$;
3. 对合性: $H^2 = I$;
4. 反射性: 设 $x = u + \alpha w, u \in \mathrm{span}\{w\}^{\perp}, \alpha \in \mathbb{R}$, 则 $Hx = u - \alpha w$;
5. 特征值: $n-1$ 个为 1, 1 个为 -1.

定理

设 $0 \neq x \in \mathbb{R}^n$, 则可构造单位向量 $w \in \mathbb{R}^n$, 使得 Householder 变换 H 满足

$$Hx = \alpha e_1,$$

其中 $\alpha = \pm \|x\|_2$.

Householder 变换

定理

设 $0 \neq x \in \mathbb{R}^n$, 则可构造单位向量 $w \in \mathbb{R}^n$, 使得 Householder 变换 H 满足

$$Hx = \alpha e_1,$$

其中 $\alpha = \pm \|x\|_2$.

证明.

由于

$$Hx = (I - 2ww^{\mathrm{T}})x = x - 2(w^{\mathrm{T}}x)w,$$

故欲使 $Hx = \alpha e_1$, 则 w 应为

$$w = \frac{x - \alpha e_1}{\|x - \alpha e_1\|_2}.$$

□

Householder 变换

实际应用:

- 为避免误差抵消 (cancellation), 可取 $\alpha = -\mathrm{sign}(x_1)\left\|x\right\|_2$.
- Householder 变换可表示为 $H = I - \frac{2}{v^{\mathrm{T}}v}vv^{\mathrm{T}} = I - \beta vv^{\mathrm{T}}$, 使得 v 的第一个分量为 1.
- 为避免计算 $v^{\mathrm{T}}v$ 出现上溢或下溢, 可选取适当正数 α, 令 $v = \alpha v$.
- 通常不显示计算矩阵 H, 而仅存储 β 和 v. 例如计算
  $HA = (I - \beta vv^{\mathrm{T}})A = A - \beta v(A^{\mathrm{T}}v)^{\mathrm{T}}$.

Givens 变换 (平面旋转变换)

$$
\begin{aligned}
G(i,k,\theta) &= I + s(e_i e_k^{\mathrm{T}} - e_k e_i^{\mathrm{T}}) + (c-1)(e_i e_i^{\mathrm{T}} + e_k e_k^{\mathrm{T}}) \\
&= \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & & s & & \\ & & & \ddots & & & \\ & & -s & & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix},
\end{aligned}
$$

其中 $c = \cos(\theta), s = \sin(\theta)$.

Givens 变换 (平面旋转变换)

设 $x \in \mathbb{R}^n$. 令 $y = G(i,k,\theta)x$, 则有

$$
\begin{aligned}
y_i &= cx_i + sx_k, \\
y_k &= -sx_i + cx_k, \\
y_j &= x_j, j \neq i, k.
\end{aligned}
$$

Givens 变换 (平面旋转变换)

设 $x \in \mathbb{R}^n$. 令 $y = G(i,k,\theta)x$, 则有

$$
\begin{aligned}
y_i &= cx_i + sx_k, \\
y_k &= -sx_i + cx_k, \\
y_j &= x_j, j \neq i, k.
\end{aligned}
$$

若要 $y_k = 0$, 可取

$$
c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}}, \quad s = \frac{x_k}{\sqrt{x_i^2 + x_k^2}},
$$

便有

$$
y_i = \sqrt{x_i^2 + x_k^2}, \quad y_k = 0.
$$

Givens 变换 (平面旋转变换)

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$$

Algorithm 2 计算 Givens 变换

function: $[c, s] = \mathbf{Givens}(a, b)$
if $b = 0$ **then**
  $c = 1;\ \ s = 0;$
else
  if |b|>|a| **then**
    $\tau = a/b;\ \ s = 1/\sqrt{1+\tau^2};\ \ c = s\tau;$
  else
    $\tau = b/a;\ \ c = 1/\sqrt{1+\tau^2};\ \ s = c\tau;$
  end if
end if

正交变换法

设 $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. 由于2 范数具有正交不变性, 故对任意的正交矩阵 $Q \in \mathbb{R}^{m\times m}$, 有

$$\|Ax - b\|_2 = \left\| Q^{\mathrm{T}}(Ax - b) \right\|_2 .$$

正交变换法

设 $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. 由于2 范数具有正交不变性, 故对任意的正交矩阵 $Q \in \mathbb{R}^{m\times m}$, 有

$$\|Ax - b\|_2 = \left\|Q^{\mathrm{T}}(Ax - b)\right\|_2.$$

若 A 列满秩, 并假定 A 的 QR 分解为

$$A = [Q_1, Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix}, \text{ 其中 } Q_1 \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},$$

则有

$$\|Ax - b\|_2^2 = \left\|Q^{\mathrm{T}}Ax - Q^{\mathrm{T}}b\right\|_2^2 = \left\|Rx - Q_1^{\mathrm{T}}b\right\|_2^2 + \left\|Q_2^{\mathrm{T}}b\right\|_2^2.$$

正交变换法

设 $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. 由于2 范数具有正交不变性, 故对任意的正交矩阵 $Q \in \mathbb{R}^{m\times m}$, 有

$$\|Ax - b\|_2 = \left\|Q^{\mathrm{T}}(Ax - b)\right\|_2 .$$

若 A 列满秩, 并假定 A 的 QR 分解为

$$A = [Q_1, Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix}, \text{ 其中 } Q_1 \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},$$

则有

$$\|Ax - b\|_2^2 = \left\|Q^{\mathrm{T}}Ax - Q^{\mathrm{T}}b\right\|_2^2 = \left\|Rx - Q_1^{\mathrm{T}}b\right\|_2^2 + \left\|Q_2^{\mathrm{T}}b\right\|_2^2 .$$

因此, x 是最小二乘问题解当且仅当 x 是方程组 $Rx = Q_1^{\mathrm{T}}b$ 的解.

QR 分解定理

定理

设 $A \in \mathbb{C}^{m\times n}(m \geq n)$, 则 A 有 QR 分解

$$A = Q\begin{bmatrix} R \\ 0 \end{bmatrix},$$

其中 $Q \in \mathbb{C}^{m\times m}$ 是正交矩阵, $R \in \mathbb{C}^{n\times n}$ 是具有非负对角元的上三角矩阵; 而且当 $m = n$ 且 A 非奇异时, 上述的分解还是唯一的.

QR 分解定理

定理

设 $A \in \mathbb{C}^{m\times n}(m \geq n)$, 则 A 有 QR 分解

$$A = Q\begin{bmatrix} R \\ 0 \end{bmatrix},$$

其中 $Q \in \mathbb{C}^{m\times m}$ 是正交矩阵, $R \in \mathbb{C}^{n\times n}$ 是具有非负对角元的上三角矩阵; 而且当 $m = n$ 且 A 非奇异时, 上述的分解还是唯一的.

1. Householder 变换
2. Givens 变换
3. (Modified) Gram-Schmidt 正交化过程

```
for j = 1 : n do
    if j < m then
        [v, β] = house(A(j : m, j))
        A(j : m, j : n) = (I_{m-j+1} − β v v^T) A(j : m, j : n)
        d(j) = β
        A(j+1 : m, j) = v(2 : m−j+1)
    end if
end for
```

QR 分解: Householder 法

```
for j = 1 : n do
    if j < m then
        [v, β] = house(A(j : m, j))
        A(j : m, j : n) = (I_{m−j+1} − β v v^T) A(j : m, j : n)
        d(j) = β
        A(j + 1 : m, j) = v(2 : m − j + 1)
    end if
end for
```

注:

- 运算量为 $2n^2(m - n/3)$
- 数值精度高
- 通常 Givens 变换需要更多运算量, 大约是 Householder 法的两倍

作　业

p. 97, 1, 3, 4, 5, 11

QQ 答疑群: 737278793