第三讲 线性最小二乘问题

3.5 线性最小二乘问题的求解方法

3.8 最小二乘问题的推广与应用

怎么求解线性最小二乘问题

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2$$

求解线性最小二乘问题的直接法

正规方程法, QR 分解法, 奇异值分解法.

3-5-1 正规方程法

定理 设 $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$, 则 $x_* \in \mathbb{R}^n$ 是线性最小二乘问题的解当且仅当残量 $r = b - Ax_*$ 与 $\mathrm{Ran}(A)$ (值域) 正交, 即 x_* 是下面的**正规方程**的解

$$A^{\intercal}(b - Ax) = 0 \quad \text{或} \quad A^{\intercal}Ax = A^{\intercal}b. \tag{3.1}$$

(板书)

解的存在性与唯一性

由于

$$A^\intercal b \in \mathrm{Ran}(A^\intercal) = \mathrm{Ran}(A^\intercal A),$$

因此正规方程 $A^\intercal A x = A^\intercal b$ 是相容 (consistent) 的, 即 **最小二乘解总是存在的**.

解的存在性与唯一性

由于

$$A^\intercal b \in \operatorname{Ran}(A^\intercal) = \operatorname{Ran}(A^\intercal A),$$

因此正规方程 $A^\intercal Ax = A^\intercal b$ 是相容 (consistent) 的, 即 **最小二乘解总是存在的**.

定理 设 $A \in \mathbb{R}^{m\times n}$ $(m > n)$, 则 $A^\intercal A$ 对称正定当且仅当 A 是列满秩的, 即 $\operatorname{rank}(A) = n$. 此时, 线性最小二乘问题的解是 **唯一的**, 其表达式为

$$x = (A^\intercal A)^{-1}A^\intercal b.$$

运算量

- 计算 $A^{\intercal}A$: $\quad\rightarrow\quad mn^2$ % 只需计算下三角或上三角部分
- 计算 Cholesky 分解: $\quad\rightarrow\quad \frac{1}{3}n^3$
- 回代求解: $\quad\rightarrow\quad \mathcal{O}(n^2)$

总的运算量大约为 $\boxed{mn^2 + \frac{1}{3}n^3 + \mathcal{O}(n^2)}$

运算量

- 计算 $A^\intercal A$: $\quad\rightarrow\quad mn^2$ % 只需计算下三角或上三角部分
- 计算 Cholesky 分解: $\quad\rightarrow\quad \dfrac{1}{3}n^3$
- 回代求解: $\quad\rightarrow\quad \mathcal{O}(n^2)$

总的运算量大约为 $\boxed{mn^2 + \dfrac{1}{3}n^3 + \mathcal{O}(n^2)}$

正规方法的优缺点

- 优点: 运算量最小, 简单直观.
- 缺点: $A^\intercal A$ 的条件数是 A 的平方, 对于病态问题, 不建议使用.

例 下面的例子说明, 计算 $A^{\intercal}A$ 可能会损失计算精度: 设

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & & \\ & \varepsilon & \\ & & \varepsilon \end{bmatrix},$$

则

$$A^{\intercal}A = \begin{bmatrix} 1+\varepsilon^2 & 1 & 1 \\ 1 & 1+\varepsilon^2 & 1 \\ 1 & 1 & 1+\varepsilon^2 \end{bmatrix}.$$

记 ε_u 为机器精度, 则当 $\varepsilon_u < \varepsilon < \sqrt{\varepsilon_u}$ 时有 $\varepsilon^2 < \varepsilon_u$, 由于舍入误差的原因, 通过浮点运算计算得到的 $A^{\intercal}A$ 是 奇异 的.

但我们注意到 A 是满秩的.

3-5-2 QR 分解法

假定 $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ 是满秩的.

设 A 的 QR 分解为 $A = QR$, 我们用三种不同的方法来推导线性最小二乘问题的解.

3-5-2 QR 分解法

假定 $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ 是满秩的.

设 A 的 QR 分解为 $A = QR$, 我们用三种不同的方法来推导线性最小二乘问题的解.

$$\boxed{x_* = R^{-1}Q^{\intercal}b}$$

- 方法一: 将 Q 的扩充成一个正交矩阵. **(板书)**
- 方法二: 利用正交投影, 将 b 写成 $b = (QQ^{\intercal} + I - QQ^{\intercal})b$. **(板书)**
- 方法三: 解正规方程. **(板书)**

关于 QR 分解法的说明

- 用 QR 分解来求最小二乘解的运算量大约为 $2mn^2$ (如果采用 Householder 变换的话, 运算量大约为 $2mn^2 - \frac{2}{3}n^3$). 当 $m \gg n$ 时, 大约为正规方程的两倍. 当 $m = n$ 时, 几乎相同.
- 通常 QR 算法比较稳定, 是求解最小二乘问题的 **首选方法**, 特别是当 A 条件数较大 (病态) 时.

3-5-3 奇异值分解法

设 A 列满秩, 奇异值分解 $A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\intercal$. 令 U_n 为 U 的前 n 列, 即 $U = [U_n, \tilde{U}]$, 则

$$\|Ax - b\|_2^2 = \left\| U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\intercal x - b \right\|_2^2 = \left\| \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\intercal x - [U_n, \tilde{U}]^\intercal b \right\|_2^2$$

$$= \left\| \begin{bmatrix} \Sigma V^\intercal x - U_n^\intercal b \\ -\tilde{U}^\intercal b \end{bmatrix} \right\|_2^2 = \|\Sigma V^\intercal x - U_n^\intercal b\|_2^2 + \|\tilde{U}^\intercal b\|_2^2 \geq \|\tilde{U}^\intercal b\|_2^2,$$

等号当且仅当 $\Sigma V^\intercal x - U_n^\intercal b = 0$ 时成立, 即

$$\boxed{x = (\Sigma V^\intercal)^{-1} U_n^\intercal b = V\Sigma^{-1} U_n^\intercal b}$$

这就是线性最小二乘问题的解.

关于 SVD 法的几点说明

相比于正规方程法和 QR 分解法, SVD 法具有更高的健壮性

但需要计算 A 的 SVD, 运算量 **远超** 正规方程法和 QR 分解法.

所以只有当系数矩阵秩亏或者接近秩亏时才使用 (此时 QR 分解法可能会失效).

例 分别用三种方法求解最小二乘问题, 比较运算时间. (LS_3methods.m)

三种方法的运算时间如下（以秒为单位）：

n	正规方程法	QR 分解法	奇异值分解法
500	0.0050	0.0220	0.0370
1000	0.0160	0.0340	0.1440
1500	0.0490	0.1330	0.5530
2000	0.0870	0.2070	1.4840
2500	0.1910	0.4430	3.1160
3000	0.2500	0.6950	5.9600
3500	0.4640	1.2130	10.0500
4000	0.4940	1.5700	14.8750
4500	0.6690	2.1680	20.6410
5000	1.0720	2.9350	28.6360

最小二乘问题的推广与应用

推广与应用

推广：正则化，加权正则化，约束最小二乘问题，…

应用：多项式数据拟合，应线性预测，信号恢复与图像处理，…

3-8-1 推广：正则化

对于欠定线性方程组, 解不唯一, 问题不适定, 需要 **正则化** (regularization):

$$\boxed{\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 + \frac{\alpha}{2}\|x\|_2^2},$$

其中第二项为 **正则项**, $\alpha > 0$ 是正则化参数.

3-8-1 推广：正则化

- 对于欠定线性方程组, 解不唯一, 问题不适定, 需要 **正则化** (regularization):

$$\boxed{\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 + \frac{\alpha}{2}\|x\|_2^2},$$

其中第二项为 **正则项**, $\alpha > 0$ 是正则化参数.

- 目标函数 $$\boxed{J(x) = \frac{1}{2}\|Ax-b\|_2^2 + \frac{\alpha}{2}\|x\|_2^2}$$ 是严格凸的二次函数,

存在唯一的最小值点, 即稳定点:

$$\boxed{\frac{\partial J}{\partial x} = 0 \iff (A^\intercal A + \alpha I)x = A^\intercal b}$$

由于 $A^\intercal A + \alpha I$ 对称正定, 故非奇异, 所以 **存在唯一解**.

3-8-1 推广：加权正则化

一类应用更广泛的正则化方法是 **加权正则化**:

$$\boxed{\min_{x\in\mathbb{R}^n} J(x) \triangleq \frac{1}{2}\|Ax-b\|_2^2 + \frac{\alpha}{2}\|Wx\|_2^2},$$

其中 $W \in \mathbb{R}^{p\times n}$ 是加权矩阵 (可以是长方阵).

3-8-1 推广：加权正则化

一类应用更广泛的正则化方法是 **加权正则化**:

$$\boxed{\min_{x\in\mathbb{R}^n} J(x) \triangleq \frac{1}{2}\|Ax-b\|_2^2 + \frac{\alpha}{2}\|Wx\|_2^2},$$

其中 $W \in \mathbb{R}^{p\times n}$ 是加权矩阵 (可以是长方阵).

类似地, 问题的解满足

$$\boxed{\frac{\partial J}{\partial x} = 0 \Longleftrightarrow (A^\intercal A + \alpha W^\intercal W)x = A^\intercal b}$$

如果 $A^\intercal A + \alpha W^\intercal W$ 非奇异, 则存在唯一解.

3-8-1 推广：约束最小二乘问题

考虑带有约束的最小二乘问题

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 \\ &\text{s.t.} \quad Bx = f \end{aligned}$$

3-8-1 推广：约束最小二乘问题

考虑带有约束的最小二乘问题

$$
\begin{aligned}
&\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 \\
&\text{s.t.} \quad Bx = f
\end{aligned}
$$

- 对应的 Lagrange 函数为

$$
J(x) = \frac{1}{2}\|Ax-b\|_2^2 + \lambda^\intercal (Bx-f)
$$

， λ 是 Lagrange 乘子

- 问题的解就是 $J(x)$ 的鞍点:

$$
\frac{\partial J}{\partial x} = 0,\ \frac{\partial J}{\partial \lambda} = 0 \Longleftrightarrow \begin{bmatrix} A^\intercal A & B^\intercal \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\intercal b \\ f \end{bmatrix}
$$

3-8-2 应用：多项式数据拟合

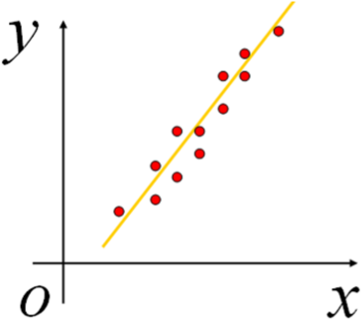
最小二乘的一个重要应用就是低次多项式数据拟合:

已知 n 个点 $\{(t_i, f_i)\}_{i=1}^n$, 寻找一个低次多项式来拟合这些数据.

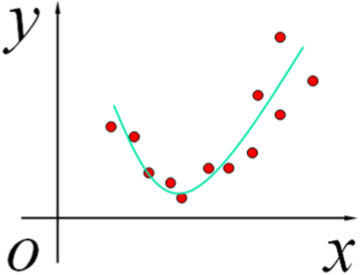
3-8-2 应用：多项式数据拟合

最小二乘的一个重要应用就是低次多项式数据拟合:

已知 n 个点 $\{(t_i, f_i)\}_{i=1}^{n}$，寻找一个低次多项式来拟合这些数据.



$y = ax + b$



$y = ax^2 + bx + c$

设拟合多项式为

$$\boxed{p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m}, \quad (m \ll n)$$

设拟合多项式为

$$\boxed{p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m}, \qquad (m \ll n)$$

- 将上述 n 个点 $\{(t_i, f_i)\}_{i=1}^n$ 代入可得

$$\begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^m \\ 1 & t_2 & t_2^2 & \cdots & t_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^m \end{bmatrix}_{n\times(m+1)} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \quad \text{或} \quad Ax = f,$$

- 超定方程组, 解通常不存在, 转化为求解最小二乘问题

$$\boxed{\min_{x\in\mathbb{R}^{m+1}} \|f - Ax\|_2^2}$$

3-8-2 应用：线性预测

预测一个时间序列的未来走向, 其中一个常用方法就是 **线性预测**.

3-8-2 应用：线性预测

预测一个时间序列的未来走向，其中一个常用方法就是 **线性预测**.

- 假定时间序列在 t_k 时刻的值 f_k 线性依赖于前 m 个时刻的值，即

$$f_k = a_1 f_{k-1} + a_2 f_{k-2} + \cdots + a_m f_{k-m}$$

> 已测得前 n 个值: $f_i,\ i = 0, 1, \ldots, n-1$, 如何预测未来的取值？ $(n \gg m)$

3-8-2 应用：线性预测

预测一个时间序列的未来走向，其中一个常用方法就是 **线性预测**.

- 假定时间序列在 t_k 时刻的值 f_k 线性依赖于前 m 个时刻的值，即

$$f_k = a_1 f_{k-1} + a_2 f_{k-2} + \cdots + a_m f_{k-m}$$

❓ 已测得前 n 个值：$f_i,\ i = 0, 1, \ldots, n-1$，如何预测未来的取值？ $(n \gg m)$

- 将现有的数据代入关系式可得

$$\begin{bmatrix} f_{m-1} & f_{m-2} & f_{m-3} & \cdots & f_0 \\ \vdots & \vdots & \vdots & & \vdots \\ f_{n-2} & f_{n-3} & f_{n-4} & \cdots & f_{n-m-1} \end{bmatrix}_{(n-m)\times m} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} f_m \\ \vdots \\ f_{n-1} \end{bmatrix} \quad \text{或} \quad Ax = f,$$

这也是一个超定问题，其解也是通过求解最小二乘问题来获得.

3-8-2 应用：信号恢复与图像处理

信号去噪

在获取数字信号时，由于各种各样的原因，最后得到的信号总会带有一定的噪声，即

$$\boxed{b = x + \eta}$$

其中 x 是真实的信号，b 是观察到的信号，η 是噪声.

3-8-2 应用：信号恢复与图像处理

信号去噪

在获取数字信号时，由于各种各样的原因，最后得到的信号总会带有一定的噪声，即

$$b = x + \eta$$

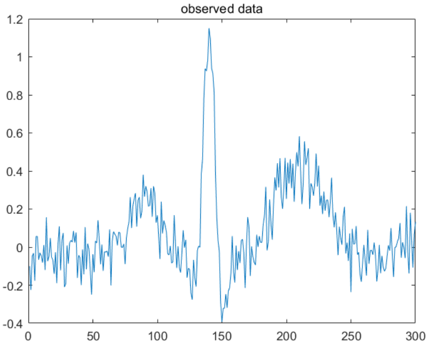
其中 x 是真实的信号，b 是观察到的信号，η 是噪声.

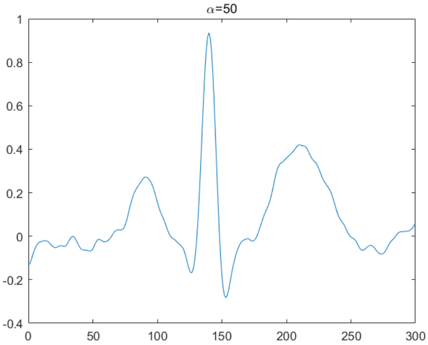
去噪 是数字信号和图像处理的基本问题，有效方法之一是加权最小二乘法

$$\min_x \frac{1}{2}\|x-b\|_2^2 + \frac{1}{2}\alpha\|Dx\|_2^2$$

其中 D 是离散的二阶导算子或 TV 算子.

例 去噪举例: LS_denoise.m





数字图像去噪与去模糊

数字图像的获取模型

$$f = \mathcal{B}(x) + \eta$$

其中 x 是真实图像, f 是观察到的图像, $\mathcal{B}(\cdot)$ 是卷积算子 (模糊机制), η 是噪声.

数字图像去噪与去模糊

数字图像的获取模型

$$f = \mathcal{B}(x) + \eta$$

其中 x 是真实图像, f 是观察到的图像, $\mathcal{B}(\cdot)$ 是卷积算子 (模糊机制), η 是噪声.

Tikhonov regularization model

$$\min \|\mathcal{B}(x) - f\|_2^2 + \mu^2 \|x\|_2^2$$

数字图像去噪与去模糊

数字图像的获取模型

$$f = \mathcal{B}(x) + \eta$$

其中 x 是真实图像, f 是观察到的图像, $\mathcal{B}(\cdot)$ 是卷积算子 (模糊机制), η 是噪声.

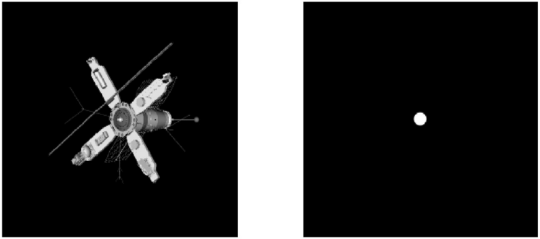
Tikhonov regularization model

$$\min \|\mathcal{B}(x) - f\|_2^2 + \mu^2 \|x\|_2^2$$

Weighted regularization model

$$\min \|\mathcal{B}(x) - f\|_2^2 + \mu^2 \|Dx\|_2^2$$

例 由于对焦不准而造成图像模糊 (此时 $\mathcal{B}$ 是一个线性算子), 噪声为高斯白噪声.



(a) (b)

(a) The original satellite image; (b) the out-of-focus blurring function.

(a) (b) (c)

(a) The blurred and noisy image (b) the restored image using the spatial invariant model; (c) the restored image using the spatial variant model