最小二乘解

2021 年 10 月 19 日

1　$Ax = b$ 的最小二乘解

本文的目标是求方程 $Ax = b$ 的最小二乘解, 即找到使得 $\|Ax - b\|_2$ 最小的 x.

Definition 1. 设 $A \in M_{N\times p}(\mathbb{C})$. 若 $A^\dagger \in M_{p\times N}(\mathbb{C})$ 满足

(i) $AA^\dagger$ 和 $A^\dagger A$ 均为 Hermite 矩阵(共轭转置等于本身);

(ii) $AA^\dagger A = A$;

(iii) $A^\dagger AA^\dagger = A^\dagger$.

则称 $A^\dagger$ 是 A 的 Moore–Penrose 广义逆.

先交代一些记号. 对任意函数 $f : \mathcal{X} \to \mathbb{R}$, 定义

$$\arg\min_x f(x) := \{x \in \mathcal{X} :\ f(x) = \min_t f(t)\}.$$

对 $\forall\, n \in \mathbb{N}$, $\forall\, x := (x_1, \ldots, x_n) \in \mathbb{C}_{n,1}$, $\|x\|_2 := [\sum_{i=1}^n (x_i)^2]^{1/2}$.

下面是本文的主要结论.

Theorem 2. 设 $A \in M_{N\times p}(\mathbb{C})$, $b \in M_{N\times 1}(\mathbb{C})$. 则

$$\arg\min_x \|Ax - b\|_2 = \{A^\dagger b + (I_p - A^\dagger A)y :\ y \in M_{p\times 1}(\mathbb{C})\},$$

其中 $A^\dagger$ 是 A 的 Moore–Penrose 广义逆. 此外, $A^\dagger b$ 是最小二乘解中 2 范数最小的.

Remark 3. 在 A 比较特殊时, $Ax = b$ 有唯一解.

(i) 当 A 可逆时, $Ax = b$ 有唯一解 $\hat{x} = A^{-1}x$;

(ii) 当 $N > p$ 且 $\operatorname{rank} A = p$ 时, $Ax = b$ 有唯一解 $\hat{x} = (A^*A)^{-1}A^*b$;

(iii) 当 $N < p$ 且 $\operatorname{rank} A = N$ 时, $Ax = b$ 有唯一解 $\hat{x} = A^*(AA^*)^{-1}b$.

2 正交投影算子

Definition 4. 设 $\mathcal{X}$ 是 Banach 空间. 线性映射 $P:\ \mathcal{X} \to \mathcal{X}$ 若满足

$$P^2 = P,$$

则称 P 是 $\mathcal{X}$ 上的投影算子. 记

$$\mathcal{R}(P) := \{Px:\ x \in \mathcal{X}\} \quad 且 \quad \mathcal{N}(P) := \{x \in \mathcal{X}:\ Px = \theta\}.$$

Proposition 5. 设 $\mathcal{H}$ 是 Hilbert 空间, P 是 $\mathcal{H}$ 上的有界投影算子. 则 $\mathcal{R}(P) = [\mathcal{N}(P)]^{\perp}$ 当且仅当 P 是对称算子.

证明. 先证 "$\Rightarrow$". 设 $\mathcal{R}(P) = [\mathcal{N}(P)]^{\perp}$, 则

$$\begin{aligned}(Px, y) &= (Px, Py + (I - P)y) = (Px, Py) \\ &= (Px + (I - P)x, Py) = (x, Py).\end{aligned}$$

再证 "$\Leftarrow$". 设 P 是对称算子, 则对 $\forall x \in \mathcal{R}(P)$, $\forall y \in \mathcal{N}(P)$, 存在 $z \in \mathcal{H}$ 使得 $x = Pz$, 因此

$$(x, y) = (Pz, y) = (z, Py) = 0.$$

故 $\mathcal{R}(P) \subset [\mathcal{N}(P)]^{\perp}$. 又因为由正交分解有

$$\mathcal{N}(P) \oplus [\mathcal{N}(P)]^{\perp} = \mathcal{H} = \mathcal{N}(P) \oplus \mathcal{R}(P),$$

因此 $\mathcal{R}(P) = [\mathcal{N}(P)]^{\perp}$. □

Definition 6. 有界对称投影算子被称为正交投影算子(orthogonal projection).

Lemma 7. 设 $\mathcal{X}$ 是 Banach 空间且 P 是 $\mathcal{X}$ 上的投影算子. 则 $\mathcal{N}(P) = \mathcal{R}(I - P)$.

3 Moore–Penrose 广义逆的一些性质

Lemma 8. Moore–Penrose 广义逆存在且唯一.

证明. 先证存在性. 用 SVD 分解构造. 严格证明待补.

再证唯一性. 设 P, $Q \in M_{p\times N}(\mathbb{C})$ 满足 Definition 1 中条件, 则

$$AP = AQAP = (AQ)^*(AP)^* = Q^*(APA)^* = Q^*A^* = AQ,$$

从而

$$P = APA = AQA = Q.$$

Lemma 8 证毕. $\square$

Proposition 9. 设 $A \in M_{N\times p}(\mathbb{C})$ 且 $A^\dagger \in M_{p\times N}(\mathbb{C})$ 是 A 的 Moore–Penrose 广义逆. 则

(i) $(A^*)^\dagger = (A^\dagger)^*$;

(ii) $\mathcal{N}(A) = \mathcal{N}(A^\dagger A)$;

(iii) $AA^\dagger$ 是正交投影算子且 $\mathcal{R}(AA^\dagger) = \mathcal{R}(A)$;

(iv) $A^\dagger A$ 是正交投影算子且 $\mathcal{R}(A^\dagger A) = \mathcal{R}(A^*)$.

证明. (i) 和 (ii) 是显然的. 下证 (iii). 容易得到 $AA^\dagger$ 是正交投影算子且 $\mathcal{R}(AA^\dagger) \subset \mathcal{R}(A)$, 故只需证 $\mathcal{R}(A) \subset \mathcal{R}(AA^\dagger)$. 事实上, 对任意 $y \in \mathcal{R}(A)$, 存在 $x \in M_{p\times 1}(\mathbb{C})$ 使得 $y = Ax$, 故

$$y = Ax = AA^\dagger Ax = AA^\dagger y \in \mathcal{R}(AA^\dagger).$$

因此 $\mathcal{R}(A) \subset \mathcal{R}(AA^\dagger)$, (iv) 证毕.

再证 (iv). 显然 $A^\dagger A$ 是正交投影算子, 且由 Definition 1(i) 和本命题的 (iii)知

$$\mathcal{R}(A^\dagger A) = \mathcal{R}(A^*(A^\dagger)^*) = \mathcal{R}(A^*(A^*)^\dagger) = \mathcal{R}(A^*),$$

(v) 证毕. $\square$

Lemma 10. 设 $A \in M_{N\times p}(\mathbb{C})$ 且 $A^\dagger \in M_{p\times N}(\mathbb{C})$ 是 A 的 Moore–Penrose 广义逆. 则

(i) 当 A 可逆时, $A^\dagger = A^{-1}$;

(ii) 当 $N > p$ 且 $\operatorname{rank} A = p$ 时, $A^\dagger = (A^*A)^{-1}A^*$;

(iii) 当 $N < p$ 且 $\operatorname{rank} A = N$ 时, $A^\dagger = A^*(AA^*)^{-1}$.

4 Theorem 2 的证明

Theorem 2 的证明. 对任意 $x \in M_{p\times 1}(\mathbb{C})$, 由 Proposition 9(iii) 知

$$\|Ax - b\|_2 = \|Ax - AA^\dagger b\|_2 + \|AA^\dagger b - b\|_2.$$

因此 $Ax = b$ 的最小二乘解是 $Ax = AA^\dagger b$ 的解. 注意到 $A^\dagger b$ 是 $Ax = AA^\dagger b$ 的一个解, 故 $A^\dagger b + \mathcal{N}(A)$ 是 $Ax = AA^\dagger b$ 的全部解. 又由 Proposition 9(ii)(iii) 和 Lemma 7 知,

$$\mathcal{N}(A) = \mathcal{N}(A^\dagger A) = \mathcal{R}(I - A^\dagger A) = \{(I_p - A^\dagger A)y:\ y \in M_{p\times 1}(\mathbb{C})\}.$$

Theorem 2 证毕. □

Remark 3 的证明. 由 Lemma 10 可直接得到. □