最小二乘理论与矩阵分解专题讲义

从经典理论到现代应用的系统性阐述

专题讲义

2025 年 5 月 24 日

摘要

本讲义系统地介绍最小二乘理论及其相关的矩阵分解方法，从经典的向量最小二乘问题出发，逐步扩展到矩阵最小二乘、约束优化、以及在概率论中的应用。重点阐述矩阵分解（SVD、QR等）在解决这类问题中的核心作用，并展示这些工具如何应用于现代统计和信号处理问题中。

目录

1 引言

现代的最小二乘理论不仅包括经典的线性回归问题，还涵盖了矩阵优化、约束优化、正则化方法等多个方面。

矩阵分解技术，特别是奇异值分解(SVD)和QR分解，在解决最小二乘问题中发挥着核心作用。这些分解不仅提供了数值稳定的计算方法，更重要的是，它们揭示了问题的几何结构，使我们能够深入理解解的性质和存在性。

本讲义将系统地构建这一理论体系，从基础概念出发，逐步建立完整的理论框架。

2 经典最小二乘理论

2.1 问题的提出

最小二乘问题

给定矩阵$A \in \mathbb{R}^{m\times n}$和向量$\boldsymbol{b} \in \mathbb{R}^m$，最小二乘问题是寻找向量$\boldsymbol{x} \in \mathbb{R}^n$使得

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

其中$\|\cdot\|_2$表示欧几里得范数。

这个问题的几何意义是：在A的列空间$\mathrm{Col}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\}$中寻找最接近向量$\boldsymbol{b}$的点。

2.2 几何解释

最小二乘问题本质上是一个投影问题。设$\boldsymbol{x}^*$是最小二乘解，则$A\boldsymbol{x}^*$是$\boldsymbol{b}$在$\mathrm{Col}(A)$上的正交投影。

定理 2.1 (最小二乘解的几何特征)**.** 向量$\boldsymbol{x}^*$是最小二乘问题的解当且仅当残差向量$\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}^*$正交于$A$的列空间，即

$$A^T(\boldsymbol{b} - A\boldsymbol{x}^*) = \boldsymbol{0}$$

证明. 设$\boldsymbol{x}^*$为最小二乘解，对于任意$\boldsymbol{h} \in \mathbb{R}^n$，定义函数

$$f(t) = \|A(\boldsymbol{x}^* + t\boldsymbol{h}) - \boldsymbol{b}\|_2^2$$

由于$\boldsymbol{x}^*$是最优解，$f(t)$在$t = 0$处取得最小值，因此$f'(0) = 0$。

计算导数：

$$\begin{aligned} f(t) &= \|A\boldsymbol{x}^* + tA\boldsymbol{h} - \boldsymbol{b}\|_2^2 && (1)\\ &= (\boldsymbol{r} + tA\boldsymbol{h})^T(\boldsymbol{r} + tA\boldsymbol{h}) && (2)\\ &= \|\boldsymbol{r}\|_2^2 + 2t\boldsymbol{r}^T A\boldsymbol{h} + t^2\|A\boldsymbol{h}\|_2^2 && (3) \end{aligned}$$

其中$\boldsymbol{r} = A\boldsymbol{x}^* - \boldsymbol{b}$。因此

$$f'(0) = 2\boldsymbol{r}^T A\boldsymbol{h} = 0$$

由于这对任意$\boldsymbol{h}$成立，得到$A^T\boldsymbol{r} = \boldsymbol{0}$。 □

2.3 正规方程

从几何特征可以直接得到著名的正规方程：

正规方程

最小二乘问题的解满足正规方程：

$$A^T A\boldsymbol{x} = A^T \boldsymbol{b}$$

当$A^T A$可逆时，唯一解为：

$$\boldsymbol{x}^* = (A^T A)^{-1} A^T \boldsymbol{b}$$

2.4 解的存在性和唯一性

定理 2.2 (最小二乘解的存在性和唯一性). (i) 最小二乘问题总是有解。

(ii) 当A的列线性无关（即$\text{rank}(A) = n$）时，解唯一。

(iii) 当A的列线性相关时，解不唯一，但最小范数解唯一。

3 矩阵分解理论

矩阵分解是解决最小二乘问题的核心工具。不同的分解方法揭示了矩阵的不同结构特征。

3.1 奇异值分解(SVD)

奇异值分解

对于任意矩阵$A \in \mathbb{R}^{m\times n}$，存在正交矩阵$U \in \mathbb{R}^{m\times m}$和$V \in \mathbb{R}^{n\times n}$，使得

$$A = U\Sigma V^T$$

其中$\Sigma \in \mathbb{R}^{m\times n}$是对角矩阵，对角元素$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$（$r = \text{rank}(A)$）称为奇异值。

SVD的几何意义深刻：它将任意线性变换分解为三个基本变换的复合：旋转、缩放、再旋转。

SVD的几何意义

SVD分解$A = U\Sigma V^T$可以理解为：

- V^T：在输入空间进行旋转，将输入向量对齐到主要方向
- Σ：沿主要方向进行缩放，缩放因子为奇异值
- U：在输出空间进行旋转，将结果对齐到输出空间的主要方向

3.2 Moore-Penrose伪逆

当A^TA不可逆时，我们需要更一般的工具来求解最小二乘问题。

定义 3.1 (Moore-Penrose伪逆)**.** 矩阵$A \in \mathbb{R}^{m\times n}$的Moore-Penrose伪逆$A^+ \in \mathbb{R}^{n\times m}$定义为满足以下四个Penrose条件的唯一矩阵：

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $(AA^+)^T = AA^+$
4. $(A^+A)^T = A^+A$

使用SVD可以显式构造伪逆：

定理 3.1 (伪逆的SVD表示)**.** 设$A = U\Sigma V^T$是A的SVD分解，则

$$A^+ = V\Sigma^+U^T$$

其中$\Sigma^+ \in \mathbb{R}^{n\times m}$定义为：当$\Sigma_{ii} \neq 0$时，$\Sigma^+_{ii} = 1/\Sigma_{ii}$；否则$\Sigma^+_{ii} = 0$。

3.3 QR分解

QR分解在数值计算中具有更好的稳定性。

定义 3.2 (QR分解)**.** 对于列满秩矩阵$A \in \mathbb{R}^{m\times n}$（$m \geq n$），存在正交矩阵$Q \in \mathbb{R}^{m\times m}$和上三角矩阵$R \in \mathbb{R}^{n\times n}$，使得

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

通常写作$A = Q_1R$，其中$Q_1 \in \mathbb{R}^{m\times n}$是$Q$的前$n$列。

4 投影理论

投影理论是理解最小二乘问题几何结构的关键。

4.1 正交投影

正交投影算子

设$S \subseteq \mathbb{R}^m$是子空间，正交投影算子$P_S : \mathbb{R}^m \to S$定义为：对任意$\boldsymbol{v} \in \mathbb{R}^m$，$P_S\boldsymbol{v}$是$S$中最接近$\boldsymbol{v}$的向量。

定理 4.1 (投影算子的性质)**.** 正交投影算子P_S具有以下性质：

(i) $P_S^2 = P_S$（幂等性）

(ii) $P_S^T = P_S$（自伴随性）

(iii) $\text{Range}(P_S) = S$

(iv) $\text{Null}(P_S) = S^{\perp}$

(v) 对任意$\boldsymbol{v} \in \mathbb{R}^m$，有$\boldsymbol{v} - P_S\boldsymbol{v} \perp S$

4.2 投影算子的矩阵表示

定理 4.2 (列空间投影算子)**.** 设$A \in \mathbb{R}^{m\times n}$，则$\boldsymbol{b}$在$\text{Col}(A)$上的投影算子为：

$$P_{\text{Col}(A)} = A(A^TA)^{-1}A^T \quad \text{（当}A\text{列满秩时）}$$

一般情况下：

$$P_{\text{Col}(A)} = AA^+$$

定理 4.3 (零空间投影算子)**.** 零空间$\text{Null}(A)$的投影算子为：

$$P_{\text{Null}(A)} = I - A^+A$$

5 广义最小二乘问题

5.1 约束最小二乘

考虑带等式约束的最小二乘问题：

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 \quad \text{subject to} \quad C\boldsymbol{x} = \boldsymbol{d}$$

定理 5.1 (约束最小二乘解)**.** 约束最小二乘问题的解为：

$$\boldsymbol{x}^* = A^+\boldsymbol{b} + (I - A^+A)C^+(\boldsymbol{d} - CA^+\boldsymbol{b})$$

其中第一项是无约束最小二乘解，第二项是约束修正项。

5.2 加权最小二乘

当观测具有不同精度时，使用加权最小二乘：

$$\min_{\boldsymbol{x}} \|W^{1/2}(A\boldsymbol{x} - \boldsymbol{b})\|_2^2 = \min_{\boldsymbol{x}} (A\boldsymbol{x} - \boldsymbol{b})^T W (A\boldsymbol{x} - \boldsymbol{b})$$

其中W是正定权重矩阵。

定理 5.2 (加权最小二乘解)**.** 加权最小二乘问题的解为：

$$\boldsymbol{x}^* = (A^TWA)^{-1}A^TW\boldsymbol{b}$$

6 矩阵最小二乘问题

现在我们将理论扩展到矩阵情况，这在现代应用中越来越重要。

6.1 Frobenius范数最小化

矩阵最小二乘问题

给定矩阵$A \in \mathbb{R}^{m\times p}$，$B \in \mathbb{R}^{q\times n}$，$C \in \mathbb{R}^{m\times n}$，矩阵最小二乘问题是寻找矩阵$X \in \mathbb{R}^{p\times q}$使得

$$\min_X \|AXB - C\|_F^2$$

其中$\|\cdot\|_F$是Frobenius范数。

定理 6.1 (矩阵最小二乘解). 矩阵最小二乘问题$\min_X \|AXB - C\|_F^2$的解为：

$$X^* = A^+ CB^+$$

当A和B分别列满秩和行满秩时，解为：

$$X^* = (A^T A)^{-1} A^T CB^T (BB^T)^{-1}$$

证明. 使用向量化技巧。注意到$\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$，其中$\otimes$是Kronecker积。

因此原问题等价于：

$$\min_{\text{vec}(X)} \|(B^T \otimes A)\text{vec}(X) - \text{vec}(C)\|_2^2$$

这是标准的向量最小二乘问题，解为：

$$\text{vec}(X^*) = (B^T \otimes A)^+ \text{vec}(C)$$

利用Kronecker积伪逆的性质$(B^T \otimes A)^+ = (B^T)^+ \otimes A^+ = B \otimes A^+$（当$A$和$B^T$列满秩时），得到

$$\text{vec}(X^*) = (B \otimes A^+)\text{vec}(C) = \text{vec}(A^+ CB^+)$$

因此$X^* = A^+ CB^+$。 □

6.2 多重约束的矩阵问题

考虑同时满足左乘和右乘约束的问题：

$$\min_X \|X\|_F^2 \quad \text{subject to} \quad AX = C, \quad XB = D$$

这种问题在统计学中的条件分布分析中经常出现。

定理 6.2 (双约束矩阵最小二乘). 问题$\min_X \|X\|_F^2$在约束$AX = C$和$XB = D$下的解（当约束相容时）为：

$$X^* = A^+ C + DB^+ - A^+ ADB^+$$

其中约束相容条件为$A^+ CB = D$。

7 矩阵分解在最小二乘中的核心作用

7.1 分解的统一观点

所有主要的矩阵分解都可以理解为对特定几何结构的揭示：

矩阵分解的本质作用

- **SVD**：揭示矩阵的完整几何结构，包括四个基本子空间的正交基
- **QR分解**：将列空间的基正交化，保持数值稳定性
- **特征值分解**：对称矩阵的主轴方向分析
- **Cholesky分解**：正定矩阵的平方根分解

在最小二乘问题中，这些分解的作用是：

1. **降维**：将高维问题分解为低维子问题
2. **去耦**：将耦合的变量分离为独立的部分
3. **投影**：构造到各个子空间的投影算子
4. **稳定**：提供数值稳定的计算方法

7.2 四个基本子空间

对于矩阵$A \in \mathbb{R}^{m\times n}$，SVD分解揭示了四个基本子空间：

定理 7.1 (四个基本子空间). 设$A = U\Sigma V^T$是SVD分解，$r = \mathrm{rank}(A)$，则：

$$\mathrm{Col}(A) = \{u_1, u_2, \dots, u_r\} \tag{4}$$
$$\mathrm{Null}(A^T) = \{u_{r+1}, u_{r+2}, \dots, u_m\} \tag{5}$$
$$\mathrm{Row}(A) = \{v_1, v_2, \dots, v_r\} \tag{6}$$
$$\mathrm{Null}(A) = \{v_{r+1}, v_{r+2}, \dots, v_n\} \tag{7}$$

其中u_i和v_i分别是U和V的列向量。

这四个子空间满足正交关系：$\mathrm{Col}(A) \perp \mathrm{Null}(A^T)$，$\mathrm{Row}(A) \perp \mathrm{Null}(A)$。

7.3 投影的矩阵表示

定理 7.2 (SVD投影表示). 使用SVD分解，各子空间的投影算子为：

$$P_{\mathrm{Col}(A)} = \sum_{i=1}^{r} u_i u_i^T = U_r U_r^T \tag{8}$$
$$P_{\mathrm{Null}(A^T)} = \sum_{i=r+1}^{m} u_i u_i^T = U_{m-r} U_{m-r}^T \tag{9}$$
$$P_{\mathrm{Row}(A)} = \sum_{i=1}^{r} v_i v_i^T = V_r V_r^T \tag{10}$$
$$P_{\mathrm{Null}(A)} = \sum_{i=r+1}^{n} v_i v_i^T = V_{n-r} V_{n-r}^T \tag{11}$$

其中U_r, V_r分别是U, V的前r列，U_{m-r}, V_{n-r}是剩余列。

8 在概率论中的应用：条件分布构造

8.1 高斯分布的线性约束

考虑高斯随机向量$\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$在线性约束$D\boldsymbol{z} = \boldsymbol{b}$下的条件分布。

这个问题的解决思路体现了最小二乘理论的完美应用：

条件分布的构造原理

线性约束下高斯分布的条件分布可以分解为：

$$\boldsymbol{z}|D\boldsymbol{z} = \boldsymbol{b} = \text{（确定性最小范数解）} + \text{（投影到零空间的噪声）}$$

具体地：

$$\boldsymbol{z}|D\boldsymbol{z} = \boldsymbol{b} \stackrel{d}{=} D^{+}\boldsymbol{b} + P_{\text{Null}(D)}\tilde{\boldsymbol{z}}$$

其中$\tilde{\boldsymbol{z}}$是$\boldsymbol{z}$的独立拷贝。

8.2 构造的几何直观

这个构造的几何意义是：

1. $D^{+}\boldsymbol{b}$是满足约束$D\boldsymbol{z} = \boldsymbol{b}$的最小范数解
2. $P_{\text{Null}(D)}\tilde{\boldsymbol{z}}$是投影到$D$零空间的独立高斯噪声
3. 两部分正交，因此可以独立处理

8.3 矩阵情况的推广

对于矩阵约束问题：给定$A \sim \mathcal{N}(0, I)$（元素独立），约束$AQ = Y$和$M^T A = X^T$，条件分布为：

$$A|\text{约束} \stackrel{d}{=} E + \mathcal{P}(\tilde{A})$$

其中：

- E是满足约束的最小Frobenius范数解
- $\mathcal{P}$是投影到约束零空间的投影算子
- $\tilde{A}$是A的独立拷贝

最小范数解的构造使用了我们前面发展的理论：

$$E = \text{argmin}\{\|A\|_F^2 : AQ = Y, A^T M = X\}$$

投影算子的构造为：

$$\mathcal{P}(A) = P_{M^\perp} A P_{Q^\perp}$$

其中$P_{M^\perp} = I - M(M^T M)^{-1} M^T$，$P_{Q^\perp} = I - Q(Q^T Q)^{-1} Q^T$。

9 数值方法与计算考虑

9.1 条件数与数值稳定性

定义 9.1 (条件数)**.** 矩阵A的条件数定义为：

$$\kappa(A) = \|A\|\|A^+\| = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

其中$\sigma_{\max}, \sigma_{\min}$分别是最大和最小非零奇异值。

当$\kappa(A)$很大时，最小二乘问题是病态的，数值解可能不可靠。

9.2 正则化方法

当问题病态时，常用正则化方法：

定理 9.1 (Ridge回归)**.** Ridge回归问题

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2$$

的解为：

$$\boldsymbol{x}_\lambda = (A^TA + \lambda I)^{-1}A^T\boldsymbol{b}$$

正则化的几何意义是在原解附近寻找范数较小的解，这在统计学中对应于引入先验信息。

10 总结与展望

10.1 理论体系的统一性

本讲义展示了最小二乘理论的统一性：

- **几何统一**：所有问题都可以理解为投影问题
- **代数统一**：矩阵分解提供了统一的解析工具
- **数值统一**：相同的数值方法适用于不同的问题变种

10.2 矩阵分解的核心地位

矩阵分解在最小二乘理论中的核心作用体现在：

1. **结构揭示**：分解揭示了问题的内在几何结构
2. **解的构造**：分解直接给出了解的显式表达式
3. **性质分析**：分解使我们能够分析解的存在性、唯一性等性质
4. **数值计算**：分解提供了数值稳定的计算方法

10.3 现代应用

这些理论在现代统计学、机器学习、信号处理中有广泛应用：

- **高维统计**：稀疏回归、变量选择
- **机器学习**：主成分分析、矩阵补全
- **信号处理**：压缩感知、图像重构
- **概率论**：条件分布、贝叶斯推断

理解这些基础理论对于掌握现代数据科学方法至关重要。

参考文献

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 1996.

[2] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. SIAM, 1997.

[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.

[4] A. Björck, *Numerical Methods for Least Squares Problems*. SIAM, 1996.

[5] G. W. Stewart, *Matrix Algorithms Volume I: Basic Decompositions*. SIAM, 1998.

[6] D. S. Watkins, *Fundamentals of Matrix Computations*, 3rd ed. Wiley, 2010.