

MIT 18.06 线性代数

第十七讲：正交矩阵和施密特正交化

整理自 MIT 18.06 课程及相关公开笔记

引言

本讲是关于“正交”理论的最后一讲，重点在于将普通向量组转换为性质更优的向量组。首先引入**标准正交向量组**与**标准正交矩阵**（正交矩阵）的概念；随后介绍将这些优良性质应用于投影计算与最小二乘求解时的巨大简化；最后详细讲解**施密特正交化**（Gram-Schmidt）方法，以及由此导出的 **QR 分解**。

1 标准正交向量与标准正交矩阵

1.1 标准正交向量组

一组向量 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 若满足以下条件，则称为**标准正交**（Orthonormal）：

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

即：这些向量两两互相垂直（正交，Orthogonal），且每个向量的长度均为 1（归一化，Normalized）。标准正交向量一定是线性无关的。

1.2 标准正交矩阵

将标准正交向量组作为列向量构成的矩阵记为 Q 。根据内积的定义，可以得到其最核心的性质：

$$Q^T Q = I$$

此处 Q 不一定是方阵，当 Q 为方阵时，称其为**正交矩阵**（Orthogonal Matrix）。此时由于 $Q^T Q = I$ ，可以直接推导出其逆矩阵的简洁形式：

$$Q^T = Q^{-1}$$

示例：

- **置换矩阵**：对单位阵进行行交换得到的矩阵，是典型的正交矩阵，例如：

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad Q Q^T = I$$

- **旋转矩阵**：平面上的二维旋转变换矩阵：

$$Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

- **阿达玛 (Hadamard) 矩阵**：由 +1 和 -1 构成的正交矩阵：

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad Q = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

- **长方形标准正交矩阵**：即使不是方阵，列向量也可满足标准正交条件。例如：

$$Q = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}$$

$$\text{该矩阵可通过扩展为正交矩阵 } Q = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}。$$

2 标准正交矩阵在投影与求解中的优势

2.1 投影矩阵的简化

根据上一讲关于投影矩阵的公式 $P = A(A^T A)^{-1} A^T$ ，若使用标准正交矩阵 Q 替代矩阵 A ，由于 $Q^T Q = I$ ，投影矩阵可化为：

$$P = Q(Q^T Q)^{-1} Q^T = Q Q^T$$

如果矩阵 Q 为方阵，则其列向量张成整个空间，投影矩阵进一步简化为单位阵：

$$P = Q Q^T = Q Q^{-1} = I$$

这意味着对落在整个空间内的任意向量进行投影，结果就是其自身。

2.2 最小二乘正规方程的简化

求解无解方程组 $A\mathbf{x} = \mathbf{b}$ 的最优解时，需要解正规方程 $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ 。当采用标准正交矩阵 Q 时， $Q^T Q = I$ ，正规方程立即简化为：

$$\hat{\mathbf{x}} = Q^T \mathbf{b}$$

这极大地降低了计算复杂度。此时，最优解 $\hat{\mathbf{x}}$ 的每一个分量 $\hat{x}_i$ ，恰好等于 $\mathbf{b}$ 在标准正交基向量 $\mathbf{q}_i$ 上的投影系数，即：

$$\hat{x}_i = \mathbf{q}_i^T \mathbf{b}$$

3 施密特正交化

从任意一组线性无关的向量（例如 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ ）出发，可以通过**施密特正交化**方法构造出张成同一空间的一组标准正交基 $(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ 。

3.1 正交化过程推导

1. **第一步（设定第一个方向）**：取第一个向量 $\mathbf{a}$ 的方向作为基准，得到未归一化的正交向量 $\mathbf{A}$ ：

$$\mathbf{A} = \mathbf{a}$$

对应的单位向量为 $\mathbf{q}_1 = \mathbf{A}/\|\mathbf{A}\|$ 。

2. **第二步（对第二个向量做投影修正）**：为了使 $\mathbf{b}$ 垂直于 $\mathbf{A}$ ，只需将 $\mathbf{b}$ 在 $\mathbf{A}$ 方向的投影分量减去即可。该投影即为上一讲中提到的误差向量 $\mathbf{e}$ ：

$$\mathbf{B} = \mathbf{b} - \text{proj}_{\mathbf{A}}(\mathbf{b}) = \mathbf{b} - \frac{\mathbf{A}^T \mathbf{b}}{\mathbf{A}^T \mathbf{A}} \mathbf{A}$$

容易验证： $\mathbf{A}^T \mathbf{B} = \mathbf{A}^T \mathbf{b} - \frac{\mathbf{A}^T \mathbf{b}}{\mathbf{A}^T \mathbf{A}} (\mathbf{A}^T \mathbf{A}) = 0$ ，故 $\mathbf{A}$ 与 $\mathbf{B}$ 正交。其对应单位向量为 $\mathbf{q}_2 = \mathbf{B}/\|\mathbf{B}\|$ 。

3.2 推广到三维

扩展到三维向量组 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 时，只需在已经得到两个正交向量 $\mathbf{A}$ 与 $\mathbf{B}$ 的基础上，将 $\mathbf{c}$ 同时减去它在 $\mathbf{A}$ 和 $\mathbf{B}$ 两个方向上的投影：

$$\mathbf{C} = \mathbf{c} - \frac{\mathbf{A}^T \mathbf{c}}{\mathbf{A}^T \mathbf{A}} \mathbf{A} - \frac{\mathbf{B}^T \mathbf{c}}{\mathbf{B}^T \mathbf{B}} \mathbf{B}$$

最终得到一组正交向量 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ ，分别单位化即可得到标准正交基 $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ 。

3.3 施密特正交化实例

给定二维向量 $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 与 $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ ，求其空间的一组标准正交基。

$$\mathbf{A} = \mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{A}^T \mathbf{b}}{\mathbf{A}^T \mathbf{A}} \mathbf{A} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

验证： $\mathbf{A}^T \mathbf{B} = 1 \times 0 + 1 \times (-1) + 1 \times 1 = 0$ ，二者正交。进行单位化得到标准正交矩阵：

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2] = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}$$

可以验证，此矩阵 Q 的列空间与原始矩阵 $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$ 的列空间是完全相同的。

4 QR 分解

将 Gram-Schmidt 正交化的全过程用矩阵乘法表示，可以得到类似的分解形式 $A = QR$ ：

- A 为原始列向量组成的矩阵。
- Q 为 Gram-Schmidt 过程最终得到的标准正交矩阵。
- R 是一个上三角矩阵。

由于 $Q^T Q = I$ ，在等式 $A = QR$ 两边同时左乘 Q^T ，可以得到 R 的计算公式：

$$R = Q^T A$$

R 的上三角性质源于 Gram-Schmidt 过程的迭代本质（后生成的基向量与先前的原始列向量正交，导致矩阵左下角元素为 0）。

设 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ 是 $\mathbb{R}^n$ 的一组基，Gram-Schmidt 正交化的计算过程分为两步，第一步正交化，得到一组正交基：

$$\begin{aligned}\tilde{\mathbf{q}}_1 &= \mathbf{a}_1, \\ \tilde{\mathbf{q}}_2 &= \mathbf{a}_2 - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_2}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1, \\ \tilde{\mathbf{q}}_3 &= \mathbf{a}_3 - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_3}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1 - \frac{\tilde{\mathbf{q}}_2^T \mathbf{a}_3}{\tilde{\mathbf{q}}_2^T \tilde{\mathbf{q}}_2} \tilde{\mathbf{q}}_2, \\ &\vdots \\ \tilde{\mathbf{q}}_n &= \mathbf{a}_n - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_n}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1 - \frac{\tilde{\mathbf{q}}_2^T \mathbf{a}_n}{\tilde{\mathbf{q}}_2^T \tilde{\mathbf{q}}_2} \tilde{\mathbf{q}}_2 - \dots - \frac{\tilde{\mathbf{q}}_{n-1}^T \mathbf{a}_n}{\tilde{\mathbf{q}}_{n-1}^T \tilde{\mathbf{q}}_{n-1}} \tilde{\mathbf{q}}_{n-1}.\end{aligned}$$

第二步再单位化每个向量，得到标准正交基： $\mathbf{q}_i = \frac{\tilde{\mathbf{q}}_i}{\|\tilde{\mathbf{q}}_i\|}, i = 1, 2, \dots, n$.

设 $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$, $\tilde{Q} = [\tilde{\mathbf{q}}_1 \ \tilde{\mathbf{q}}_2 \ \dots \ \tilde{\mathbf{q}}_n]$, $Q = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_n]$ ，第一步正交化，能得到 $\mathbf{a}_i (i = 1, 2, \dots, n)$ 被 $\tilde{\mathbf{q}}_1, \tilde{\mathbf{q}}_2, \dots, \tilde{\mathbf{q}}_i$ 线性表示的表示法。这可用矩阵乘法表示：

$$A = \tilde{Q} \tilde{R}, \quad \tilde{R} = \begin{bmatrix} 1 & \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_2}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} & \dots & \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_n}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \\ & 1 & \ddots & \vdots \\ & & \ddots & \frac{\tilde{\mathbf{q}}_{n-1}^T \mathbf{a}_n}{\tilde{\mathbf{q}}_{n-1}^T \tilde{\mathbf{q}}_{n-1}} \\ & & & 1 \end{bmatrix}.$$

第二步单位化可以写成 $\tilde{Q} = Q \text{diag}(\|\tilde{\mathbf{q}}_i\|)$ ，因此

$$A = Q \text{diag}(\|\tilde{\mathbf{q}}_i\|) \tilde{R} = QR,$$

其中 $R = \text{diag}(\|\tilde{\mathbf{q}}_i\|) \tilde{R}$ 是对角元素都是正数的上三角矩阵， Q 是正交矩阵。注意， $\tilde{Q}$ 一般不是正交矩阵。

如果以 A 的列向量作为平行六面体的三条同一顶点出发的边，那么可以看到，第一步正交化的过程 $A \mapsto AR_1 \mapsto AR_1R_2 \mapsto AR_1R_2R_3 = \tilde{Q}$ ，就是将这个平行六面体化成长方体的过程。参见图 3.2.2. ☺

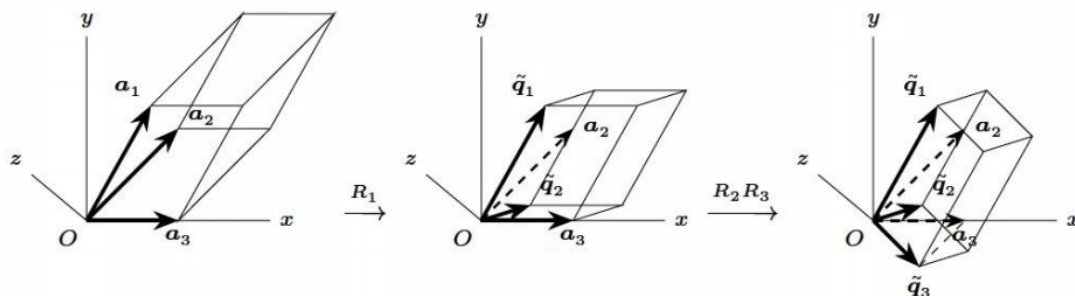


图 3.2.2 向量的正交化

4.1 具体实例

考察一个具体的 3×3 矩阵：

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3]$$

运用 Gram-Schmidt 正交化可以算出：

$$\begin{array}{lll} \mathbf{a}_1 = (1, 0, 0)^T, & \tilde{\mathbf{q}}_1 = (1, 0, 0)^T, & \|\tilde{\mathbf{q}}_1\| = 1 \\ \mathbf{a}_2 = (2, 0, 3)^T, & \tilde{\mathbf{q}}_2 = (0, 0, 3)^T, & \|\tilde{\mathbf{q}}_2\| = 3 \\ \mathbf{a}_3 = (4, 5, 6)^T, & \tilde{\mathbf{q}}_3 = (0, 5, 0)^T, & \|\tilde{\mathbf{q}}_3\| = 5 \end{array}$$

对 $\tilde{Q}$ 中的列进行单位化，得到标准正交矩阵 Q ：

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \mathbf{q}_3]$$

由 $R = Q^T A$ 可直接求出上三角矩阵：

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

从 $A = QR$ 中可以看出，矩阵 R 的每一列恰好是 A 的原始列向量在标准正交基上的线性组合系数：

- $\mathbf{a}_1 = 1\mathbf{q}_1$
- $\mathbf{a}_2 = 2\mathbf{q}_1 + 3\mathbf{q}_2$
- $\mathbf{a}_3 = 4\mathbf{q}_1 + 6\mathbf{q}_2 + 5\mathbf{q}_3$

4.2 QR 分解的代码实现

基于前文的理论推导，本节展示两种主流科学计算环境中 QR 分解的标准实现、不同输出参数的物理含义，利用列置换可以增强数值稳定性或优化稀疏矩阵计算。

MATLAB 中的 qr

MATLAB 的 `qr` 函数针对不同需求提供了丰富的调用形式，其常见的调用语法为：

```
R = qr(A) % 返回上三角因子 R
[Q, R] = qr(A) % 标准分解：A = Q*R, Q 为正交矩阵，R 为上三角
[Q, R, P] = qr(A) % 列置换分解：A*P = Q*R
[___] = qr(A, "econ") % 经济型分解（节约内存）
[Q, R, P] = qr(A, outputForm) % 控制 P 以矩阵或向量形式返回
```

示例：

```
>> A = magic(5); % 幻方矩阵
>> [Q,R] = qr(A)
Q = 5×5
```

```
-0.5234    0.5058    0.6735   -0.1215   -0.0441
-0.7081   -0.6966   -0.0177    0.0815   -0.0800
-0.1231    0.1367   -0.3558   -0.6307   -0.6646
-0.3079    0.1911   -0.4122   -0.4247    0.7200
-0.3387    0.4514   -0.4996    0.6328   -0.1774
```

```
R = 5×5
```

```
-32.4808  -26.6311  -21.3973  -23.7063  -25.8615
         0   19.8943   12.3234    1.9439    4.0856
         0         0  -24.3985  -11.6316   -3.7415
         0         0         0  -20.0982   -9.9739
         0         0         0         0  -16.0005
```

```
>> norm(A-Q*R) % 在计算机精度范围内验证 A = Q*R
ans = 9.5562e-15
```

Python (SciPy) 中的 `scipy.linalg.qr`

SciPy 的 `scipy.linalg.qr` 函数提供了与 MATLAB 几乎完全对应的功能，其核心调用形式为：

```
Q, R = qr(a, mode='full') # 标准分解：A = QR, Q 为列标准正交矩阵
Q, R = qr(a, mode='economic') # 经济型分解
R = qr(a, mode='r') # 仅返回上三角因子 R
Q, R, P = qr(a, pivoting=True) # 带列置换的分解，返回置换向量 P
```

示例:

```
>>> import numpy as np
>>> from scipy.linalg import qr
>>> rng = np.random.default_rng()
>>> a = rng.standard_normal((9, 6)) # 生成 9x6 的随机矩阵

# 1. 标准分解与验证
>>> q, r = qr(a)
>>> np.allclose(a, q @ r)
True
>>> q.shape, r.shape
((9, 9), (9, 6))

# 2. 经济型分解与仅返回 R
>>> q2, r2 = qr(a, mode='economic')
>>> q2.shape, r2.shape
((9, 6), (6, 6))

>>> r2 = qr(a, mode='r')
>>> np.allclose(r, r2)
True

# 3. 带列置换的 QR 分解
>>> q3, r3, p = qr(a, pivoting=True)
>>> np.allclose(a[:, p], q3 @ r3) # 验证  $A[:, P] = Q * R$ 
True

>>> P = np.eye(p.size)[p] # 根据置换向量 p 构建置换矩阵 P
>>> np.allclose(a, q3 @ r3 @ P.T) # 验证  $A = (Q * R) * P^T$ 
True
```

总结

- 标准正交向量组和标准正交矩阵 Q 拥有性质 $Q^T Q = I$, 当 Q 为方阵时, 还有 $Q^T = Q^{-1}$ 。
- 使用标准正交矩阵可以极大地简化投影矩阵计算公式, 求解正规方程 $\hat{\mathbf{x}} = Q^T \mathbf{b}$ 变得极其简便。
- 施密特正交化 (Gram-Schmidt) 提供了一个系统化的方法, 可以将任意线性无关向量组转化为张成同一空间的标准正交基。
- Gram-Schmidt 过程的矩阵化表示即为 QR 分解, 这也是一种非常重要的矩阵分解方法, 在数值计算中往往比普通的消元法更稳定。