

雷达信号处理中的矩阵方法

目 录

3			
4	第 1 章 绪论		1
5	1.1 雷达的基本概念		1
6	1.2 雷达工作原理		2
7	1.2.1 基本组成		2
8	1.2.2 工作频率		2
9	1.2.3 目标探测		4
10	1.3 电磁波传播特性		4
11	1.3.1 电磁波的反射		5
12	1.3.2 电磁波的折射		6
13	1.3.3 自由空间路径损耗与大气衰减		8
14	1.3.4 色散现象与多径效应		9
15	1.4 雷达方程		10
16	1.5 小结		13
17	第 2 章 矩阵方法基础		15
18	2.1 矩阵分解		15
19	2.1.1 特征分解		15
20	2.1.2 奇异值分解		19
21	2.1.3 稀疏分解		22
22	2.2 矩阵微分		24
23	2.2.1 实矩阵微分		24
24	2.2.2 复矩阵微分		28
25	2.3 张量及相关运算		29
26	2.3.1 张量的定义		29
27	2.3.2 张量的运算		31
28	2.3.3 张量分解		33
29	2.4 常见统计量的矩阵表示		35
30	第 3 章 雷达发射与接收信号模型		39

31	3.1 发射信号模型	39
32	3.2 阵列接收信号模型	42
33	3.2.1 一维均匀线性阵列	42
34	3.2.2 二维均匀线性阵列	45
35	3.3 多普勒接收信号模型	47
36	3.4 空时联合接收信号模型	50
37	3.5 噪声、杂波和多径模型	51
38	第 4 章 雷达信号检测	55
39	4.1 脉冲压缩	55
40	4.2 脉冲积累	60
41	4.3 目标检测	63
42	4.3.1 门限选择与检测性能	63
43	4.3.2 恒虚警率检测	67
44	第 5 章 目标参数估计	71
45	5.1 波束形成方法	71
46	5.1.1 经典波束形成	72
47	5.1.2 和差比幅测角	76
48	5.1.3 最小方差无失真响应	79
49	5.2 子空间方法	81
50	5.2.1 MUSIC 方法	82
51	5.2.2 ESPRIT 方法	87
52	5.3 稀疏表示方法	91
53	5.4 多域联合估计	94
54	5.4.1 空时自适应处理	94
55	5.4.2 张量分解方法	99
56	第 6 章 目标跟踪	103
57	6.1 状态估计	103
58	6.1.1 最小均方滤波	103
59	6.1.2 卡尔曼滤波	108
60	6.2 数据关联	112
61	6.2.1 联合概率数据关联	112

62	6.2.2 图匹配方法	114
63	附录 A 基础知识	121
64	A.1 符号规定	121
65	A.2 矩阵函数	121
66	A.3 卷积与相关的矩阵表示	121
67	A.4 拉格朗日乘数法	124
68	A.5 分块矩阵求逆公式	125
69	附录 B 向量范数与矩阵范数	129
70	B.1 向量范数	129
71	B.2 矩阵范数	132
72	附录 C 张量乘法性质	135

第 2 章 矩阵方法基础

2.1 矩阵分解

矩阵分解是线性代数中的一个重要概念，它将一个矩阵分解为多个简单的矩阵的乘积。常见的矩阵分解方法包括特征分解、奇异值分解、稀疏分解等。这些分解方法在数值计算、信号处理和机器学习等领域有着广泛的应用。

2.1.1 特征分解

特征分解 (Eigenvalue Decomposition) 是线性代数中的一个重要概念，对于一个方阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ ，其特征值和特征向量是满足以下方程的标量和向量：

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \quad (2.1)$$

其中， λ 是特征值，非零向量 $\mathbf{v}$ 是对应的特征向量。

如果矩阵 $\mathbf{A}$ 一共有 n 个线性无关的特征向量，构成如下的特征向量矩阵

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix},$$

那么有

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}, \quad (2.2)$$

其中 $\mathbf{\Lambda}$ 是一个对角矩阵，其对角线上的元素为对应的特征值，即

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

对 (2.2) 两边同时左乘 $\mathbf{V}^{-1}$ ，可以得到

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}. \quad (2.3)$$

式 (2.3) 就是矩阵的特征分解形式，而能够被分解的矩阵被称为可对角化矩阵。进一步地，对于对称矩阵，则有定理 2.1 所述的特征分解形式。

定理 2.1 (对称矩阵的特征分解) 如果矩阵 $\mathbf{A}$ 是一个对称矩阵，则其特征向量构成的

矩阵为正交矩阵, 即满足

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}.$$

此时, 矩阵有如下特征分解

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T.$$

证明 设 $\mathbf{A}$ 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 对应的特征向量为 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 。由于 $\mathbf{A}$ 是对称矩阵, 所以

$$\mathbf{A} \mathbf{v}_i = \mathbf{A}^T \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

显然, 对于任意的 $i \neq j$, 都有如下两个等式成立:

$$\mathbf{v}_i^T \mathbf{A} \mathbf{v}_j = \mathbf{v}_i^T (\mathbf{A} \mathbf{v}_j) = \lambda_j \mathbf{v}_i^T \mathbf{v}_j,$$

$$\mathbf{v}_i^T \mathbf{A} \mathbf{v}_j = (\mathbf{A}^T \mathbf{v}_i)^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j.$$

因此有

$$\lambda_j \mathbf{v}_i^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j.$$

当 $\lambda_i \neq \lambda_j$ 时, $\mathbf{v}_i^T \mathbf{v}_j = 0$, 即特征向量正交。当 $\lambda_i = \lambda_j$ 时, 可以验证对于任意的 $\alpha, \beta \in \mathbb{C}$, 以两者为系数线性组合得到的向量

$$\mathbf{v} = \alpha \mathbf{v}_i + \beta \mathbf{v}_j,$$

也是特征向量, 对应的特征值仍然是 $\lambda_i = \lambda_j$ 。也就是说, 由 $\mathbf{v}_i$ 和 $\mathbf{v}_j$ 张成的平面上的任意向量都是特征向量。从这个平面上可以选出两个正交的特征向量 $\mathbf{v}'_i$ 和 $\mathbf{v}'_j$, 作为新的特征向量。这样, 就可以确保所有的特征向量互相之间是正交的。如对所有特征向量都进行归一化, 令其模长为一, 则有特征向量构成的矩阵满足

$$\mathbf{V} \mathbf{V}^T = \mathbf{I}.$$

■

然而并不是所有的矩阵都可以进行特征分解, 对于部分矩阵, 其特征向量可能线性相关。此时, 矩阵 $\mathbf{V}$ 不可逆, 因此无法得到特征分解。比如, 下方的 2×2 矩阵:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

设其特征值为 λ , 则有

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \left(\begin{bmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{bmatrix} \right) = (1 - \lambda)^2 = 0.$$

因此，该矩阵的两个特征值都是 1，对应的特征向量相同，都为

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

对于这样不可对角化的矩阵，一方面我们可以进行广义特征分解，将矩阵分解为若尔当标准型 (Jordan form)：

定理 2.2 (广义特征分解) 对于任意的 $n \times n$ 矩阵 $\mathbf{A}$ ，存在一个可逆矩阵 $\mathbf{P}$ 和一个分块对角矩阵 $\mathbf{J}$ ，使得

$$\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}.$$

分块对角矩阵 $\mathbf{J}$ 有如下形式

$$\mathbf{J} = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{bmatrix},$$

其中每个 J_i 是一个若尔当块 (Jordan block)，其形式为

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}.$$

考虑到本书中基本不涉及广义特征分解，这里不再给出定理 2.2 证明。除了广义特征分解外，我们还可以使用其他的矩阵分解方法来处理不可对角化的矩阵，比如下一小节给出的奇异值分解。

例 2.1 设有三座城市，且这三座城市之间人口流动的情况可以用一个矩阵来表示：

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.4 \\ 0.2 & 0.3 & 0.5 \end{bmatrix},$$

其中， P_{ij} 表示单位时间内，从城市 j 流入到城市 i 的人口比例。比如 $P_{11} = 0.5$ ，表

明经过单位时间, 有 50% 的人口仍然留在城市 1, $P_{12} = 0.2$ 表示第 2 个城市 20% 的人口会流动到城市 1。请求解以下问题:

1. 假设初始时刻, 三个城市的人口都为 1000 人, 请计算经过一个单位时间后, 各个城市的人口分布情况。
2. 请计算经过足够长的时间后, 各个城市的人口分布情况。

解

1. 记初始时刻三个城市的人口数量为如下向量

$$\mathbf{x}_0 = \begin{bmatrix} 1000 \\ 1000 \\ 1000 \end{bmatrix}.$$

则经过一个单位时间后, 各个城市的人口数量为

$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0 = \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.4 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} 1000 \\ 1000 \\ 1000 \end{bmatrix} = \begin{bmatrix} 800 \\ 1200 \\ 1000 \end{bmatrix}.$$

2. 设 $\mathbf{x}_n$ 为经过 n 个单位时间后, 各个城市的人口分布情况。则有

$$\mathbf{x}_n = \mathbf{P}\mathbf{x}_{n-1} = \mathbf{P}^n\mathbf{x}_0.$$

又因为 $\mathbf{P}$ 有如下特征分解

$$\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

其中

$$\mathbf{V} \approx \begin{bmatrix} 0.3995 & -0.8090 & 0.3090 \\ 0.7068 & 0.3090 & -0.8090 \\ 0.5839 & 0.5000 & 0.5000 \end{bmatrix}, \quad \mathbf{\Lambda} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.3618 & 0 \\ 0 & 0 & 0.1382 \end{bmatrix}.$$

因此, 当 n 趋向于无穷大时, $\mathbf{P}^n$ 的极限为

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \lim_{n \rightarrow \infty} \mathbf{V}\mathbf{\Lambda}^n\mathbf{V}^{-1} = \mathbf{V} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}^{-1} \approx \begin{bmatrix} 0.2364 & 0.2364 & 0.2364 \\ 0.4182 & 0.4182 & 0.4182 \\ 0.3455 & 0.3455 & 0.3455 \end{bmatrix}.$$

于是, 经过足够长的时间后, 各个城市的人口分布情况为

$$\mathbf{x}_\infty = \lim_{n \rightarrow \infty} \mathbf{x}_n = \lim_{n \rightarrow \infty} \mathbf{P}^n \mathbf{x}_0 \approx \begin{bmatrix} 709.1 \\ 1254.5 \\ 1036.4 \end{bmatrix}.$$

2.1.2 奇异值分解

奇异值分解 (Singular Value Decomposition, SVD) 是线性代数中一个非常重要的矩阵分解工具, 被广泛应用于信号处理、统计学、机器学习、图像压缩等领域。相比于特征分解, 奇异值分解不要求矩阵是可逆的, 可以应用于任意形状的矩阵, 具体见下方的定理。

定理 2.3 对于任意的矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 都存在一个奇异值分解:

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T,$$

其中, $\mathbf{U} \in \mathbb{R}^{m \times m}$ 是一个正交矩阵, $\mathbf{V} \in \mathbb{R}^{n \times n}$ 是一个正交矩阵, $\mathbf{S} \in \mathbb{R}^{m \times n}$ 是一个对角矩阵, 对角线上的元素为 $\mathbf{A}$ 的奇异值 (Singular Values)。

证明 不妨假设 $m \geq n$, 则 $\mathbf{A}^T \mathbf{A}$ 为一个 $n \times n$ 的对称矩阵。根据特征分解定理, 存在一个正交矩阵 $\mathbf{V}$ 和一个对角矩阵 $\mathbf{\Lambda}$, 使得

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T.$$

可以验证, $\mathbf{\Lambda}$ 的对角元素必然是非负的, 因此不妨将矩阵 $\mathbf{A}^T \mathbf{A}$ 的特征值记作 λ_i^2 。令 $\mathbf{v}_i$ 为 $\mathbf{V}$ 的第 i 列向量, 并令

$$\mathbf{u}_i = \frac{\mathbf{A} \mathbf{v}_i}{\|\mathbf{A} \mathbf{v}_i\|} = \frac{\mathbf{A} \mathbf{v}_i}{\sqrt{\mathbf{v}_i^T \mathbf{A}^T \mathbf{A} \mathbf{v}_i}} = \frac{1}{\lambda_i} \mathbf{A} \mathbf{v}_i.$$

则有

$$\mathbf{u}_i^T \mathbf{A} \mathbf{v}_i = \frac{1}{\lambda_i} \mathbf{v}_i^T \mathbf{A}^T \mathbf{A} \mathbf{v}_i = \frac{1}{\lambda_i} \lambda_i^2 = \lambda_i. \quad (2.4)$$

并且, 可以验证对于任意的 $i \neq j$ 且 $i, j \leq n$, 都有

$$\mathbf{u}_i^T \mathbf{A} \mathbf{v}_j = \frac{1}{\lambda_i} \mathbf{v}_i^T \mathbf{A}^T \mathbf{A} \mathbf{v}_j = 0. \quad (2.5)$$

注意到, $\mathbf{u}_j = \frac{1}{\lambda_j} \mathbf{A} \mathbf{v}_j$, 因此还可以得出 $\mathbf{u}_i$ 和 $\mathbf{u}_j$ 是正交的。

不妨构造如下的矩阵

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n & \mathbf{u}_{n+1} & \cdots & \mathbf{u}_m \end{bmatrix} \in \mathbb{R}^{m \times m},$$

对于其中的列向量, 当 $1 \leq i \leq n$ 时, 有 $\mathbf{u}_i = \frac{\mathbf{A} \mathbf{v}_i}{\|\mathbf{A} \mathbf{v}_i\|}$; 而当 $n+1 \leq i \leq m$ 时, 可以令 $\mathbf{u}_i$ 为 $\mathbf{A}$ 的零空间中的任意向量, 只需满足构成的矩阵 $\mathbf{U}$ 是一个正交矩阵即可。此

时, 可以验证, 对于任意的 $i \neq j$ 且 $i \leq n, j > n$, 都有

$$\mathbf{u}_i^T \mathbf{A} \mathbf{v}_j = 0. \quad (2.6)$$

综上所述, 结合 (2.4) 到 (2.6), 可以得到

$$\mathbf{u}_i^T \mathbf{A} \mathbf{v}_j = \begin{cases} \lambda_i, & i = j, \\ 0, & i \neq j. \end{cases}$$

因此, 对于构造的矩阵 $\mathbf{U}$, 可以验证其满足

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \begin{bmatrix} \sqrt{\Lambda} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{m \times n},$$

其中 $\sqrt{\Lambda}$ 是 Λ 的开平方, 也是一个对角矩阵, 其对角线上的元素为 λ_i , 而 $\mathbf{0}$ 是一个 $(m - n) \times n$ 的全零矩阵。记该分块矩阵为 $\mathbf{S}$, 并对上式两边同时左乘 $\mathbf{U}$, 右乘 $\mathbf{V}^T$, 即可得到

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T.$$

■

有意思的是, 对于对称矩阵, 其奇异值分解和特征分解是等价的。不妨设有对称矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 其 SVD 分解为

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T.$$

又因为 $\mathbf{A}^T = \mathbf{A}$, 因此有

$$\mathbf{A} = \mathbf{A}^T = (\mathbf{U} \mathbf{S} \mathbf{V}^T)^T = \mathbf{V} \mathbf{S}^T \mathbf{U}^T.$$

注意到 $\mathbf{S}$ 是一个 $n \times n$ 的对角矩阵, 所以 $\mathbf{S}^T = \mathbf{S}$, 至此可以得到如下等式

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{V} \mathbf{S} \mathbf{U}^T.$$

观察上式, 不难发现 $\mathbf{U} = \mathbf{V}$ 。此时, 对称矩阵 $\mathbf{A}$ 的特征分解可以写为 $\mathbf{A} = \mathbf{V} \mathbf{S} \mathbf{V}^T$, 因此对称矩阵的奇异值分解和特征分解是等价的。

除此以外, SVD 分解实际上提供了一种对矩阵进行低秩近似的方法。考虑如下优化问题:

$$\begin{cases} \min_{\mathbf{B}} & \|\mathbf{A} - \mathbf{B}\|_F^2 \\ \text{s.t} & \text{rank}(\mathbf{B}) = k \end{cases} \quad (2.7)$$

其中 $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数。通过计算 $\mathbf{A}$ 的 SVD 分解, 并保留最大的 k 个奇异值和对应的奇异向量, 即截断 SVD (Truncated SVD) 分解, 可以得到该优化问

题的最优解：

$$\mathbf{B} = \sum_{i=1}^k \lambda_i \mathbf{u}_i \mathbf{v}_i^T, \quad (2.8)$$

其中 λ_i 是矩阵 $\mathbf{A}$ 第 i 大的奇异值。该具体证明不再展开，请读者自行推导。

例 2.2 对于如图 2.2 所示的灰度图像，可以将其看作是一个 200×300 的矩阵，请利用奇异值分解对其进行压缩处理。



图 2.1 原始灰度图像

解 直接对图像矩阵进行奇异值分解，得到

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T.$$

其中 $\mathbf{S}$ 的对角线元素从大到小排列绘制的曲线如图 2.2 所示。从图中可以看到，奇异值数值大小从大到小迅速下降，这表明大部分信息都集中在较大的奇异值对应的特征向量上。

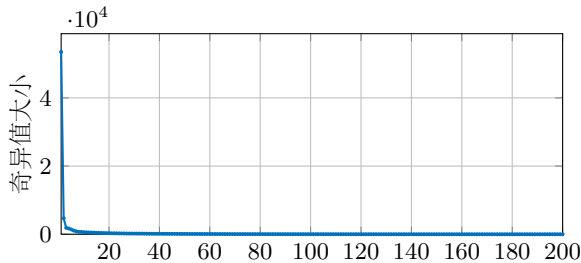


图 2.2 奇异值大小分布

因此，我们可以通过保留前 k 个奇异值来实现图像的压缩。具体而言，设 k 为保留的奇异值个数，则压缩后的图像矩阵为原始矩阵的低秩近似：

$$\mathbf{A}_{\text{compressed}} = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T,$$

其中 $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ 是 $\mathbf{U}$ 的前 k 列， $\mathbf{S}_k \in \mathbb{R}^{k \times k}$ 是 $\mathbf{S}$ 的前 k 行和前 k 列， $\mathbf{V}_k \in \mathbb{R}^{n \times k}$ 是 $\mathbf{V}$ 的前 k 列。

通过调整 k 的值，可以控制压缩后的图像质量和文件大小，如图 2.3 所示。可以看到，随着 k 的增大，图像质量逐渐提高，当 $k = 50$ 时，图像质量已经非常接近原始

图像，此时，对应的三个矩阵的元素个数大约只有原始矩阵元素个数的 41%。

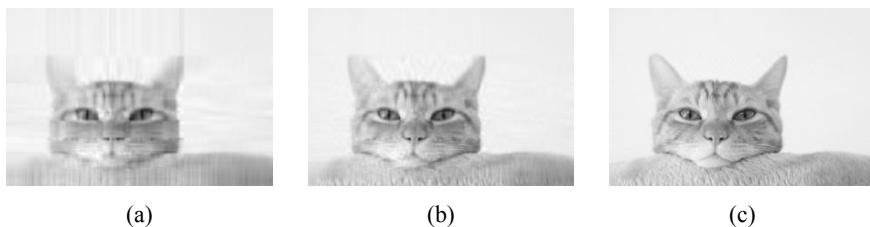


图 2.3 奇异值分解压缩后的图像 (a) $k = 10$ (b) $k = 20$ (c) $k = 50$

2.1.3 稀疏分解

稀疏分解 (Sparse Decomposition) 是指将一个观测向量表示为一个字典矩阵与一个稀疏系数向量的乘积。该方法在信号处理、图像处理和机器学习等领域具有广泛应用，尤其在特征提取、压缩编码和降维等任务中表现出强大的建模能力。

稀疏分解对应的优化问题可以表示为

$$\begin{cases} \min_{\mathbf{x}} & \|\mathbf{x}\|_0 \\ \text{s.t} & \mathbf{D}\mathbf{x} = \mathbf{y} \end{cases}, \quad (2.9)$$

其中, $\mathbf{y} \in \mathbb{R}^m$ 是观测向量, $\mathbf{D} \in \mathbb{R}^{m \times n}$ 是字典矩阵。一般来说, $n \gg m$, 这意味着方程 $\mathbf{D}\mathbf{x} = \mathbf{y}$ 可能有多个甚至无穷多个解。而稀疏分解则是希望找到其中最稀疏的那个解, 即使得系数向量 $\mathbf{x}$ 的非零元素个数最少。当存在噪声时, 等式 $\mathbf{D}\mathbf{x} = \mathbf{y}$ 可能无法完全满足, 此时可以将约束条件改为

$$\begin{cases} \min_{\mathbf{x}} & \|\mathbf{x}\|_0 \\ \text{s.t} & \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon^2 \end{cases}. \quad (2.10)$$

然而, 优化问题 (2.9) 中的 ℓ_0 -范数是非凸且不连续的, 使得该问题属于 NP-hard 问题, 难以直接求解。因此, 通常采用以下近似方法进行求解:

1. **基追踪 (Basis Pursuit, BP)**: 将 ℓ_0 -范数替换为其凸包 ℓ_1 -范数, 转化为如下凸优化问题:

$$\begin{cases} \min_{\mathbf{x}} & \|\mathbf{x}\|_1 \\ \text{s.t.} & \mathbf{D}\mathbf{x} = \mathbf{y} \text{ 或 } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon. \end{cases} \quad (2.11)$$

在满足一定条件, 如限制性等距条件 (Restricted Isometry Property) 时, 该方法能够准确恢复原始的稀疏解。并且, 可以通过成熟的线性规划 (Linear Programming) 或二阶锥规划 (Second-Order Cone Programming) 等算法高效求解。

2. **LASSO (Least Absolute Shrinkage and Selection Operator)**: 通过引入 ℓ_1 -范数正则项, 构造如下优化模型:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (2.12)$$

其中, $\lambda > 0$ 是正则化参数, 用于权衡数据拟合精度与解的稀疏性。LASSO 更适用于存在观测噪声的实际应用场景, 可以通过坐标下降法 (Coordinate Descent)、交替方向乘子法 (Alternating Direction Method of Multipliers, ADMM) 等算法高效求解。

除了上述方法, 稀疏分解还可以采用匹配追踪 (Matching Pursuit)、正交匹配追踪 (Orthogonal Matching Pursuit, OMP)、迭代阈值法 (Iterative Thresholding) 等方法进行求解, 这些方法在计算效率和精度之间提供了不同的折中选择。

将稀疏分解拓展到矩阵的情形, 可以得到 k -SVD (k -Sparse Singular Value Decomposition) 模型。该模型假设矩阵的奇异值分解可以表示为一个稀疏系数向量与字典矩阵的乘积。具体而言, 给定一个矩阵 $\mathbf{Y} \in \mathbb{R}^{m \times n}$, 对应的 k -SVD 优化问题可以表示为

$$\begin{cases} \min_{\mathbf{D}, \mathbf{X}} & \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \\ \text{s.t.} & \|\mathbf{x}_i\|_0 = 1, \quad \forall i = 1, \dots, n \end{cases}, \quad (2.13)$$

其中, $\mathbf{D} \in \mathbb{R}^{m \times k}$ 是字典矩阵, $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n] \in \mathbb{R}^{k \times n}$ 是稀疏系数矩阵。 k -SVD 是对聚类算法 k -均值 (k -Means) 的一种推广, 其中字典矩阵的列向量对应聚类中心, 而稀疏系数矩阵 $\mathbf{X}$ 则包含了聚类结果。

除此以外, 稀疏分解的矩阵推广还有另一种形式, 即鲁棒主成分分析 (Robust Principal Component Analysis, RPCA)。RPCA 假设观测矩阵可以分解为一个低秩矩阵和一个稀疏矩阵的和。具体而言, 给定一个矩阵 $\mathbf{Y} \in \mathbb{R}^{m \times n}$, RPCA 的优化问题可以表示为

$$\begin{cases} \min_{\mathbf{L}, \mathbf{S}} & \|\mathbf{L}\|_* + \lambda \|\text{vec}(\mathbf{S})\|_1 \\ \text{s.t.} & \mathbf{Y} = \mathbf{L} + \mathbf{S} \end{cases}, \quad (2.14)$$

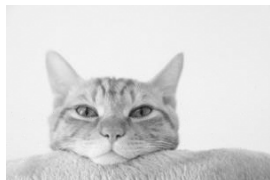
其中, $\mathbf{L} \in \mathbb{R}^{m \times n}$ 是低秩矩阵, $\|\cdot\|_*$ 表示核范数, $\mathbf{S} \in \mathbb{R}^{m \times n}$ 是稀疏矩阵, $\|\text{vec}(\cdot)\|_1$ 表示将矩阵展开为向量后的 ℓ_1 -范数, 也即矩阵所有元素的绝对值之和。由于稀疏项的存在, RPCA 能够有效提取数据中的异常, 同时保留主要的低秩结构。

例 2.3 请对如图 2.4 所示的包含椒盐噪声的图像矩阵进行去噪处理。

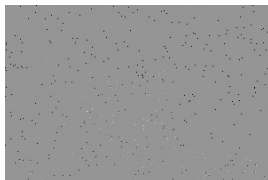
解 不难发现, 椒盐噪声对应的矩阵是一个稀疏矩阵, 而原始图像矩阵则是一个低秩矩阵。因此, 我们可以使用鲁棒主成分分析 (RPCA) 来对其进行去噪处理。算法结果如图 2.5 所示, 可以看到, 所分离出的低秩矩阵基本对应无噪声的图像, 而稀疏矩阵则对应椒盐噪声部分。



图 2.4 包含椒盐噪声的图像



(a)



(b)

图 2.5 鲁棒主成分分析结果 (a) 低秩矩阵 (b) 稀疏矩阵

2.2 矩阵微分

2.2.1 实矩阵微分

普通的求导想必大家都很熟悉，比如对于标量函数 $f(x)$ ，其关于 x 的导数可以表示为 $\frac{\partial f(x)}{\partial x}$ 。而函数不仅仅可以是关于标量的函数，也可以是关于向量甚至矩阵的函数。比如，一个二元函数 $f(x_1, x_2)$ 则可以看作是关于向量 $\mathbf{x} = [x_1 \ x_2]^T$ 的一个函数 $f(\mathbf{x})$ 。自然地，我们也可以求解其关于向量 $\mathbf{x}$ 的导数。具体而言，我们有如下定义。

定义 2.1 (标量关于向量求导) 对于向量 $\mathbf{x} \in \mathbb{R}^{n \times 1}$ ，有映射 $f: \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}$ ，则定义其关于向量 $\mathbf{x}$ 的导数为

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

换句话说， $f(\mathbf{x})$ 关于 $\mathbf{x}$ 的导数可以通过计算 $f(\mathbf{x})$ 关于 $\mathbf{x}$ 中每一个元素的偏导数，并将结果排列成和 $\mathbf{x}$ 同样维度的向量。类似地，我们也可以定义标量对矩阵的求导，具体定义如下。

定义 2.2 (标量关于矩阵求导) 对于矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$ ，有映射 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ ，则定

义其关于矩阵 $\mathbf{A}$ 导数为

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f(\mathbf{A})}{\partial a_{11}} & \frac{\partial f(\mathbf{A})}{\partial a_{12}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{1n}} \\ \frac{\partial f(\mathbf{A})}{\partial a_{21}} & \frac{\partial f(\mathbf{A})}{\partial a_{22}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{A})}{\partial a_{m1}} & \frac{\partial f(\mathbf{A})}{\partial a_{m2}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{mn}} \end{bmatrix}.$$

此外，函数本身也有可能不是一个标量，而是一个向量，因此我们也需要定义向量对向量的求导。具体定义如下：

定义 2.3（向量关于向量求导） 对于向量 $\mathbf{x} \in \mathbb{R}^{m \times 1}$ ，有映射 $\mathbf{f}: \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$ ，则定义其关于向量 $\mathbf{x}$ 导数为

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}} \\ \frac{\partial f_2(\mathbf{x})}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial \mathbf{x}} \end{bmatrix}.$$

可以看到，对应的求导规则就是将函数的每一个元素分别对向量 $\mathbf{x}$ 求导，并将结果按照 $\mathbf{f}$ 的形式排列。比如， $m \times 1$ 列向量关于 $n \times 1$ 列向量的导数为一个 $mn \times 1$ 的列向量。如果是 $1 \times m$ 的行向量关于 $n \times 1$ 列向量的导数，那么根据规则，结果则是一个 $n \times m$ 的矩阵。

接下来，本节给出一些常见的矩阵微分公式，这些公式在本书的后续章节中会经常用到。

例 2.4 设有向量 $\mathbf{x} \in \mathbb{R}^{n \times 1}$ 、 $\mathbf{y} \in \mathbb{R}^{n \times 1}$ 和矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ ，计算如下导数

1. $f = \mathbf{x}^T \mathbf{y}$ 关于 $\mathbf{x}$ 的导数
2. $f = \mathbf{x}^T \mathbf{y}$ 关于 $\mathbf{y}$ 的导数
3. $f = \mathbf{x}^T \mathbf{A} \mathbf{y}$ 关于 $\mathbf{x}$ 的导数
4. $f = \mathbf{x}^T \mathbf{A} \mathbf{y}$ 关于 $\mathbf{A}$ 的导数
5. $\mathbf{f} = \mathbf{x}^T \mathbf{A}$ 关于 $\mathbf{x}$ 的导数
6. $\mathbf{f} = (\mathbf{x} - \mathbf{y})^T$ 关于 $\mathbf{x}$ 的导数

解

234 1. 注意到 $f = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$, 因此有

$$235 \quad \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{y}.$$

236 2. 同理, 对于 $f = \mathbf{x}^T \mathbf{y}$ 有

$$237 \quad \frac{\partial f}{\partial \mathbf{y}} = \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial y_2} \\ \vdots \\ \frac{\partial f}{\partial y_n} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}.$$

238 3. 同样地, 将函数写成求和的形式 $f = \mathbf{x}^T \mathbf{A} \mathbf{y} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i y_j$, 因此有

$$239 \quad \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j} y_j \\ \sum_{j=1}^n a_{2j} y_j \\ \vdots \\ \sum_{j=1}^n a_{nj} y_j \end{bmatrix} = \mathbf{A} \mathbf{y}.$$

240 4. 同理, 对于 $f = \mathbf{x}^T \mathbf{A} \mathbf{y}$ 有

$$241 \quad \frac{\partial f}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial a_{n1}} & \frac{\partial f}{\partial a_{n2}} & \cdots & \frac{\partial f}{\partial a_{nn}} \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{bmatrix} = \mathbf{x} \mathbf{y}^T.$$

242 5. 最后, 对于 $f = \mathbf{x}^T \mathbf{A}$, 其第 j 个元素为 $f_j = \sum_{i=1}^n a_{ij} x_i$, 因此有

$$243 \quad \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}} & \frac{\partial f_2}{\partial \mathbf{x}} & \cdots & \frac{\partial f_n}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \mathbf{A}.$$

6. 对于 $\mathbf{f} = \mathbf{x} - \mathbf{y}$, 有

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial(x_1 - y_1)}{\partial \mathbf{x}} & \frac{\partial(x_2 - y_2)}{\partial \mathbf{x}} & \dots & \frac{\partial(x_n - y_n)}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}.$$

从上面的例子中, 我们可以得到一些经验结论:

1. 如果向量位于函数表达式的左侧且有转置, 那么导数则直接是右侧变量, 比如 $\frac{\partial \mathbf{x}^T \mathbf{y}}{\partial \mathbf{x}} = \mathbf{y}$ 和 $\frac{\partial \mathbf{x}^T \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$ 。
2. 如果向量位于函数表达式的右侧, 那么导数则是左侧变量加转置, 比如 $\frac{\partial \mathbf{x}^T \mathbf{y}}{\partial \mathbf{y}} = \mathbf{x}$ 和 $\frac{\partial \mathbf{A} \mathbf{y}}{\partial \mathbf{y}} = \mathbf{A}^T$ 。

矩阵微分有四个常用的性质: 线性、乘积、商和链式法则, 具体见性质 2.1。

性质 2.1 (矩阵微分的四个性质) 以下性质中的前三个与标量函数求导的性质类似, 只有最后一个链式法则略有不同。

1. 线性

$$\frac{\partial(a f(\mathbf{x}) + b g(\mathbf{x}))}{\partial \mathbf{x}} = a \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + b \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}. \quad (2.15)$$

2. 乘积

$$\frac{\partial f(\mathbf{x}) g(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} g(\mathbf{x}) + f(\mathbf{x}) \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}. \quad (2.16)$$

3. 商

$$\frac{\partial \frac{f(\mathbf{x})}{g(\mathbf{x})}}{\partial \mathbf{x}} = \frac{f'(\mathbf{x}) g(\mathbf{x}) - f(\mathbf{x}) g'(\mathbf{x})}{g^2(\mathbf{x})}. \quad (2.17)$$

4. 链式法则

$$\frac{\partial f(\mathbf{g}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^T(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial f(\mathbf{g})}{\partial \mathbf{g}}. \quad (2.18)$$

例 2.5 计算如下函数关于向量 $\mathbf{x}$ 的导数

$$f(\mathbf{x}) = \|\mathbf{A} \mathbf{x} - \mathbf{y}\|^2.$$

解 方法 1: 将函数展开, 有

$$f(\mathbf{x}) = \|\mathbf{A} \mathbf{x} - \mathbf{y}\|^2 = (\mathbf{A} \mathbf{x} - \mathbf{y})^T (\mathbf{A} \mathbf{x} - \mathbf{y}) = \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} - 2 \mathbf{y}^T \mathbf{A} \mathbf{x} + \mathbf{y}^T \mathbf{y}.$$

因此有

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2 \mathbf{A}^T \mathbf{A} \mathbf{x} - 2 \mathbf{A}^T \mathbf{y}.$$

方法 2: 记 $\mathbf{g}(\mathbf{x}) = \mathbf{Ax} - \mathbf{y}$, 则 $f(\mathbf{x})$ 可以写为如下的复合函数形式

$$f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \mathbf{g}(\mathbf{x}).$$

根据链式法则, 有

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^T(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial f(\mathbf{g})}{\partial \mathbf{g}} = \frac{(\mathbf{Ax} - \mathbf{y})^T}{\partial \mathbf{x}} \frac{\partial \mathbf{g}^T \mathbf{g}}{\partial \mathbf{g}} = 2\mathbf{I}\mathbf{g} = 2\mathbf{A}^T(\mathbf{Ax} - \mathbf{y}).$$

2.2.2 复矩阵微分

在本课程中, 还有可能涉及到复数矩阵的微分。由于在实际应用中, 大部分函数都是关于复向量或者复矩阵的实值函数, 因此我们只针对这种函数给出其导数的定义, 具体如下。

定义 2.4 对于复数 $z = x + jy$, 有映射 $g: \mathbb{C} \rightarrow \mathbb{R}$, 则定义其关于复数 z 的导数为

$$\frac{\partial g(z)}{\partial z} = \frac{\partial g(z)}{\partial x} + j \frac{\partial g(z)}{\partial y}.$$

例 2.6 设有复数 $z = x + jy$, 计算 $g(z) = z\bar{z}$ 关于 z 的导数。

解 将目标函数展开, 有

$$g(z) = z\bar{z} = (x + jy)(x - jy) = x^2 + y^2.$$

因此根据定义有

$$\frac{\partial g(z)}{\partial z} = \frac{\partial g(z)}{\partial x} + j \frac{\partial g(z)}{\partial y} = 2x + j2y = 2z.$$

利用上述定义, 我们同样得到类似的实值函数关于复向量和复矩阵的导数。下面我们通过一些简单的例子, 给出一些常用的经验公式。

例 2.7 计算如下函数关于复向量 $\mathbf{z}$ 的导数

$$g(\mathbf{z}) = \mathbf{z}^H \mathbf{z}.$$

解 记 $\mathbf{z} = \mathbf{x} + j\mathbf{y}$, 其中 $\mathbf{x}$ 和 $\mathbf{y}$ 分别是实部和虚部向量, 那么目标函数可以展开为

$$g(\mathbf{z}) = \mathbf{z}^H \mathbf{z} = (\mathbf{x} - j\mathbf{y})(\mathbf{x} + j\mathbf{y}) = \mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}.$$

因此根据定义有

$$\frac{\partial g(\mathbf{z})}{\partial \mathbf{z}} = \frac{\partial g(\mathbf{z})}{\partial \mathbf{x}} + j \frac{\partial g(\mathbf{z})}{\partial \mathbf{y}} = 2\mathbf{x} + j2\mathbf{y} = 2\mathbf{z}.$$

例 2.8 计算如下函数关于复向量 $\mathbf{z}$ 的导数

$$g(\mathbf{z}) = \mathbf{z}^H \mathbf{R} \mathbf{z},$$

其中 $\mathbf{R}$ 是一个共轭对称矩阵。

解 记 $z = x + jy$, $R = P + jQ$, 那么目标函数可以展开为

$$\begin{aligned} g(z) &= z^H R z = (x - jy)^T (P + jQ) (x + jy) \\ &= x^T P x + y^T P y + y^T Q x - x^T Q y + j(x^T Q x - y^T P x + x^T P y + y^T Q y). \end{aligned}$$

注意到, $R = R^H$, 即

$$P + jQ = P^T - jQ^T,$$

因此, 我们有

$$\begin{cases} P &= P^T \\ Q &= -Q^T \end{cases}.$$

此外, 对于任意一个向量 x , 我们有

$$\begin{cases} x^T Q x = x^T (-Q^T) x = -x^T Q^T x \\ x^T Q x = (x^T Q x)^T = x^T Q^T x, \end{cases}$$

也就是说 $x^T Q x = -x^T Q x = 0$ 。因此, $g(z)$ 可以简化为

$$g(z) = x^T P x + y^T P y + y^T Q x - x^T Q y.$$

根据定义, 我们有

$$\begin{aligned} \frac{\partial g(z)}{\partial z} &= \frac{\partial g(z)}{\partial x} + j \frac{\partial g(z)}{\partial y} = 2P x - 2Q y + j(2P y + 2Q x) \\ &= 2(P + jQ)(x + jy) = 2Rz. \end{aligned}$$

2.3 张量及相关运算

2.3.1 张量的定义

张量 (Tensor) 是线性代数与多线性代数中的一个基本概念, 可视为标量 (0 阶张量)、向量 (1 阶张量) 和矩阵 (2 阶张量) 在更高维度上的推广, 如图 2.6 所示。形式上, 一个 n 阶张量可以表示为一个多维数组:

$$\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n},$$

其中每个维度称为一个“模” (mode), 维度的个数即为张量的“阶” (order)。

对于 n 阶张量 $\mathcal{T}$, 其中元素可以用 n 个索引来表示, 通常记为 $t_{i_1 i_2 \cdots i_n}$, 其中 i_k 表示第 k 个模的索引。比如, 图 2.7 所示的 3 阶张量 $\mathcal{T} \in \mathbb{R}^{4 \times 4 \times 4}$, 其中蓝色立方体对应的元素的索引为 $(3, 4, 2)$, 即 t_{342} 。

实际上, 张量这种形式的数据结构在实际应用中非常常见, 比如一张彩色图像就可以看作是一个 3 阶张量, 这个三阶张量的三个模分别是高度、宽度和颜色通

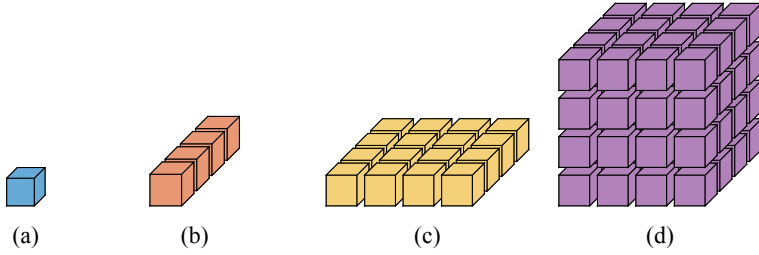


图 2.6 张量的示意图 (a) 0 阶张量 (b) 1 阶张量 (c) 2 阶张量 (d) 3 阶张量

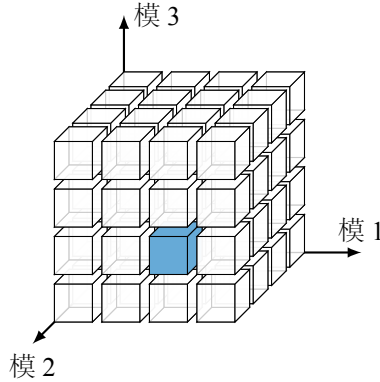


图 2.7 张量索引示意图

道。图 2.8a 展示了一张 200×300 像素大小的彩色图像，其对应的张量表示为 $\mathcal{T} \in \mathbb{R}^{200 \times 300 \times 3}$ ，其中第三个模的三个索引分别对应红色通道图 2.8b、绿色通道图 2.8c 和蓝色通道图 2.8d。



图 2.8 一张彩色图像的张量表示 (a) 原图 (b) 红色通道 (c) 绿色通道 (d) 蓝色通道

为了方便对张量进行分析，通常可以将其展开为矩阵的形式，对应的操作被称作模-k 展开 (Mode-k Flattening)。具体而言，对于一个 n 阶张量 $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ ，其模-k 展开可以表示为一个矩阵 $\mathbf{T}_k \in \mathbb{R}^{I_k \times (I_1 I_2 \cdots I_{k-1} I_{k+1} \cdots I_n)}$ ，其中矩阵的每一列对应张量 $\mathcal{T}$ 在模 k 上的一个纤维 (Fiber)，即固定其他模的索引后，模 k 上的所有元素构

成的向量：

$$\begin{bmatrix} t_{i_1 \dots i_{k-1} 1 i_{k+1} \dots i_n} & t_{i_1 \dots i_{k-1} 2 i_{k+1} \dots i_n} & \cdots & t_{i_1 \dots i_{k-1} I_k i_{k+1} \dots i_n} \end{bmatrix}^T.$$

比如，图 2.8a 所示的张量 $\mathcal{T} \in \mathbb{R}^{200 \times 300 \times 3}$ 可以通过模-1 展开得到一个矩阵 $\mathbf{T}_1 \in \mathbb{R}^{200 \times (300 \times 3)}$ ，而通过模-2 展开则得到一个矩阵 $\mathbf{T}_2 \in \mathbb{R}^{300 \times (200 \times 3)}$ 。



图 2.9 张量模-k 展开示意图 (a) 模-1 展开 (b) 模-2 展开

除了图像以外，视频也可以看作是一个 4 阶张量，对应的四个模分别为高度、宽度、颜色通道和时间。比如，一段 10 秒的视频，分辨率为 1920×1080 像素，帧率为 30 帧/秒，那么对应的张量可以表示为 $\mathcal{T} \in \mathbb{R}^{1920 \times 1080 \times 3 \times 300}$ 。此外，本书涉及到的多域雷达数据也可以看作是高阶张量，利用张量的相关知识，可以对这些数据进行有效的分析和处理。

张量的模-k 展开还可以进一步推广为多模展开 (Multi-Mode Flattening)，即同时对多个模进行展开。具体而言，对于一个 n 阶张量 $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ ，其模- $\{k_1, k_2, \dots, k_m\}$ 展开可以表示为一个矩阵 $\mathbf{T}_{k_1 k_2 \dots k_m} \in \mathbb{R}^{(I_{k_1} I_{k_2} \cdots I_{k_m}) \times (I_{l_1} I_{l_2} \cdots I_{l_{n-m}})}$ ，其中 $\{l_1, l_2, \dots, l_{n-m}\} = \{1, 2, \dots, n\} \setminus \{k_1, k_2, \dots, k_m\}$ ，即剩余的模。

2.3.2 张量的运算

本节首先介绍一种生成张量的运算，即外积 (Outer Product)，然后介绍张量与向量、矩阵之间的乘法，即 k 模积 (k -Mode Product)，以及张量乘法与克罗内克积 (Kronecker Product) 之间的关系。

内积 (Inner Product) 是线性代数中一个基本的运算，其可以将两个向量映射为一个标量。具体而言，对于两个 n 维向量 $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ ，其内积定义为

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^n a_i b_i.$$

而外积则刚好相反，同样是两个 n 维向量 $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ ，通过外积可以生成一个 $n \times n$ 的矩阵 $\mathbf{X}$ ，该矩阵可以表示为

$$\mathbf{X} = \mathbf{a} \circ \mathbf{b} = \mathbf{a} \mathbf{b}^T.$$

换句话说，矩阵 $\mathbf{X}$ 的第 i, j 个元素为 $x_{ij} = a_i b_j$ 。此外，外积还可以推广到任意个数

的向量之间, 具体定义如下。

定义 2.5 给定 n 个向量 $\mathbf{x}^{(1)} \in \mathbb{R}^{I_1}, \mathbf{x}^{(2)} \in \mathbb{R}^{I_2}, \dots, \mathbf{x}^{(n)} \in \mathbb{R}^{I_n}$, 它们的外积定义为

$$\mathcal{T} = \mathbf{x}^{(1)} \circ \mathbf{x}^{(2)} \circ \dots \circ \mathbf{x}^{(n)} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}.$$

对于张量 $\mathcal{T}$ 中的元素, 其第 $i_1, i_2, \dots, i_n$ 个元素有如下表达式

$$t_{i_1, i_2, \dots, i_n} = x_{i_1}^{(1)} x_{i_2}^{(2)} \dots x_{i_n}^{(n)},$$

其中, $x_{i_k}^{(k)}$ 表示向量 $\mathbf{x}^{(k)}$ 的第 i_k 个元素。

特别地, 如果将同一个向量 $\mathbf{x}$ 进行 n 次外积, 可以得到一个 n 阶张量, 记作

$$\mathcal{T} = \mathbf{x} \circ \mathbf{x} \circ \dots \circ \mathbf{x} = \mathbf{x}^{\circ n}.$$

并且, 该张量是一个对称张量, 即任意交换 $i_1, i_2, \dots, i_n$ 的顺序, 其对应的元素不变, 比如

$$t_{i_1, i_2, \dots, i_n} = t_{i_n, i_{n-1}, \dots, i_1} = x_{i_1} x_{i_2} \dots x_{i_n}.$$

而 k 模积同样是最基本的张量运算之一, 它可以看作是矩阵乘法的推广, 其定义如下。

定义 2.6 给定张量 $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ 与矩阵 $\mathbf{U} \in \mathbb{R}^{J \times I_k}$, 两者的 k 模积操作可以表示为

$$\mathcal{T} \times_k \mathbf{U} \in \mathbb{R}^{I_1 \times \dots \times I_{k-1} \times J \times I_{k+1} \times \dots \times I_N}, \quad (2.19)$$

k 模积结果中的元素具有如下表达式

$$(\mathcal{T} \times_k \mathbf{U})_{i_1 \dots i_{k-1} j i_{k+1} \dots i_N} = \sum_{i_k=1}^{I_k} t_{i_1 \dots i_{k-1} i_k i_{k+1} \dots i_N} u_{j i_k}, \quad (2.20)$$

其中 $t_{i_1 i_2 \dots i_N}, u_{j i_k}$ 分别为张量 $\mathcal{T}$ 和矩阵 $\mathbf{U}$ 对应位置的元素

我们知道, 矩阵乘以一个向量可以看作是用该向量中的元素对矩阵的各个列向量进行线性组合. 同样地, 张量与向量的 k 模积也可以看作作用向量中的元素对张量沿着第 k 个维度的切片进行线性组合. 例如, 对于一个三阶张量 $\mathcal{A} \in \mathbb{R}^{3 \times 3 \times 4}$, 该张量 3 模积一个 1×4 大小的向量将会得到一个 3×3 的矩阵, 具体操作如图 2.10 所示. 在这个例子中, 线性组合的对象不再是向量, 而是张量沿着第三个维度的切片, 也就是矩阵。

类似地, 张量与一个矩阵的 k 模积则可以看作是: 使用该矩阵的每一行对张量的第 k 个维度的切片进行线性组合得到一个新的切片, 然后将这些切片组合成一个新的张量. 例如, 对于一个三阶张量 $\mathcal{A} \in \mathbb{R}^{3 \times 3 \times 4}$, 其 3 模积一个 2×4 大小的矩阵会得到一个 $3 \times 3 \times 2$ 的张量, 具体操作如图 2.11 所示。

例 2.9 利用 k 模积, 将彩色图像张量 $\mathcal{T} \in \mathbb{R}^{h \times w \times 3}$, 转换为灰度图像矩阵 $\mathbf{T} \in \mathbb{R}^{h \times w}$ 。

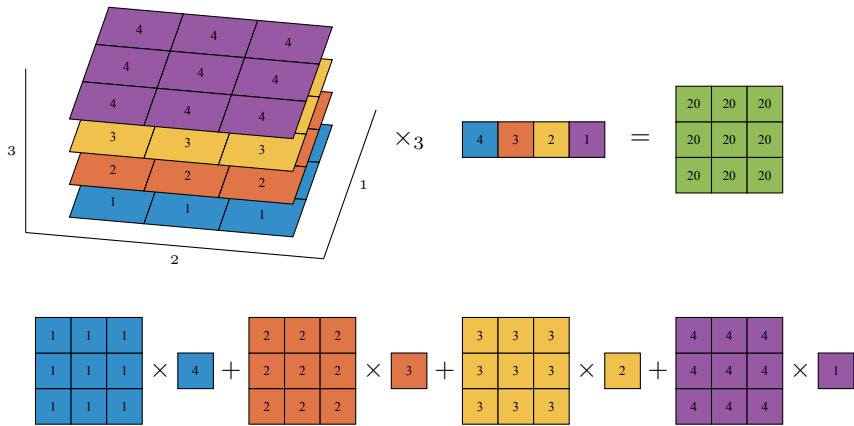


图 2.10 张量与向量 k 模积示意图

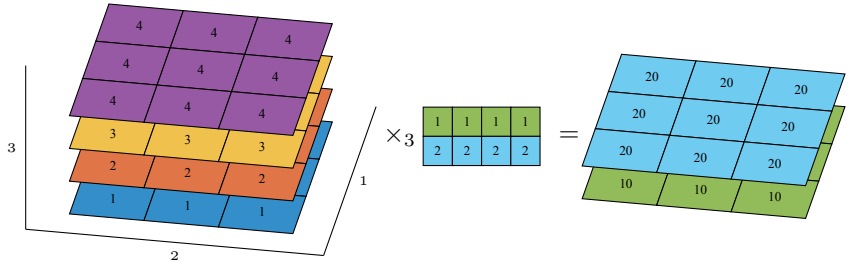


图 2.11 张量与矩阵 k 模积示意图

解 对于一个普通的彩色图像，其对应的灰度图像可以看作是将彩色图像的三个颜色通道进行加权求和得到的，即

$$Y = 0.299R + 0.587G + 0.114B,$$

其中 R, G, B 分别表示彩色图像红色、绿色和蓝色通道。因此，可以构建如下向量

$$\mathbf{w} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \end{bmatrix}.$$

并利用 k 模积将彩色图像张量转换为灰度图像矩阵，即

$$\mathbf{T} = \mathcal{T} \times_3 \mathbf{w}.$$

2.3.3 张量分解

类似于矩阵的奇异值分解 (SVD)，张量分解旨在将高维张量表示为低维张量的组合。常见的张量分解方法包括 CANDECOMP/PARAFAC (CP 分解) 和 Tucker 分解

等。

给定一个张量 $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ ，其 CP 分解可以表示为

$$\mathcal{T} \approx \sum_{r=1}^R \mathbf{x}_1^{(r)} \circ \mathbf{x}_2^{(r)} \circ \cdots \circ \mathbf{x}_N^{(r)},$$

其中， $\mathbf{x}_1^{(r)} \in \mathbb{R}^{I_1}, \mathbf{x}_2^{(r)} \in \mathbb{R}^{I_2}, \dots, \mathbf{x}_N^{(r)} \in \mathbb{R}^{I_N}$ 为基向量， R 为分解秩。简而言之，张量 CP 分解旨在将高维张量表示为多个秩一张量的和。为了避免向量大小带来的模糊问题，我们可以规定所有的基向量的模长都为 1，此时 CP 分解的可以写为

$$\mathcal{T} \approx \sum_{r=1}^R \lambda_r \mathbf{x}_1^{(r)} \circ \mathbf{x}_2^{(r)} \circ \cdots \circ \mathbf{x}_N^{(r)},$$

其中， λ_r 为标量系数，表示第 r 个秩一张量的权重。

而 Tucker 分解则将张量表示为核心张量与一组因子矩阵的乘积，其形式为

$$\mathcal{T} \approx \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \cdots \times_N \mathbf{X}_N,$$

其中， $\mathcal{G} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ 为核心张量， $\mathbf{X}^{(n)} \in \mathbb{R}^{J_n \times I_n}$ 为第 n 个维度的因子矩阵。

根据性质 C.6，可以将 CP 分解写为如下的形式

$$\mathcal{T} \approx \mathcal{D} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \cdots \times_N \mathbf{X}_N,$$

其中， $\mathcal{D} \in \mathbb{R}^{R \times R \times \cdots \times R}$ 为对角张量，对角线元素为 λ_r 。而 $\mathbf{X}_n \in \mathbb{R}^{I_n \times R}$ 则是所有的 $\mathbf{x}_n^{(r)}$ 构成的矩阵：

$$\mathbf{X}_n = \begin{bmatrix} \mathbf{x}_n^{(1)} & \mathbf{x}_n^{(2)} & \cdots & \mathbf{x}_n^{(R)} \end{bmatrix}.$$

这表明 Tucker 分解可以看作是对 CP 分解的一种推广。

例 2.10 对图 2.8a 所示的彩色图像进行 CP 分解，并重构出低秩近似图像。

解 令 $R = 200$ ，对图 2.12 进行 CP 分解，对应的各个秩一分量的权重如图 2.12 所示。可以发现，类似于 SVD 分解，权重下降速度极快，这意味着图像的主要信息集中在少数几个秩一分量中。

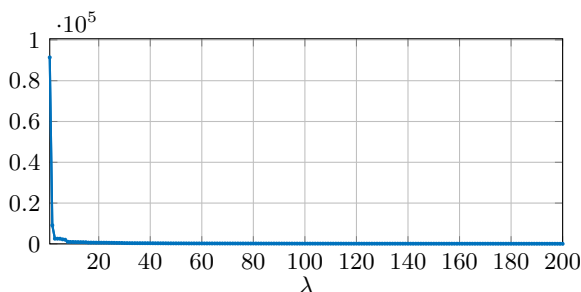


图 2.12

如图 2.13 所示，当使用的秩一分量数量接近 50 时，重构图像的质量已经非常接近原始图像。

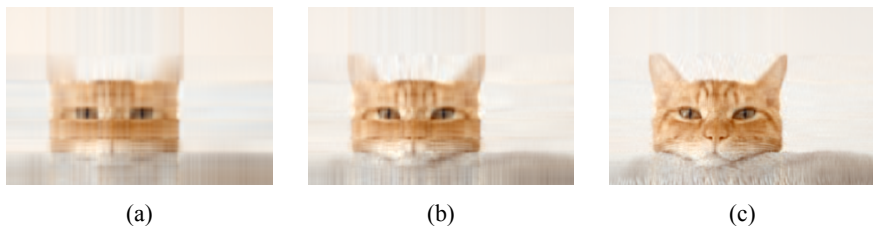


图 2.13 重构图像 (a) 10 个秩一分量 (b) 20 个秩一分量 (c) 50 个秩一分量

2.4 常见统计量的矩阵表示

在概率论中，我们经常会遇到一些常见的统计量，比如均值、方差、协方差等。这些统计量在本课程中也会经常用到。不同的是，本课程中的对象都是采集到的离散数据，对应向量或矩阵而不是随机变量。因此，我们需要提前了解对于向量和矩阵，如何计算这些统计量。

设有向量 $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^T$ ，可以看作是某个随机变量的 n 个观测。下面，我们利用求和的形式，给出该随机变量的样本均值和方差的计算公式。首先是样本均值：

$$\bar{x} = \text{Mean}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i.$$

在获得均值的基础上，我们可以进一步计算样本方差：

$$\sigma^2 = \text{Var}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

如果样本均值为零，则样本方差计算公式可以简化为

$$\sigma^2 = \text{Var}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

这两个统计量是最常见的，使用最为广泛的统计量，分别反映的数据的平均趋势（直流分量）和离散程度（功率）。

除此之外，还有其他统计量，比如偏度（三阶统计量）和峰度（四阶统计量）等，这都属于高阶统计量。其中，样本偏度的计算公式为

$$\text{Skew}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{\sigma^3}.$$

而样本峰度的计算公式为

$$\text{Kurt}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{\sigma^4}.$$

偏度通常用于衡量数据分布的对称性，而峰度则用于衡量数据分布偏离正态分布的程度。当数据的均值为 0，方差为 1 时，这两个统计量的计算公式可以简化为

$$\text{Skew}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i^3, \quad \text{Kurt}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i^4.$$

需要注意的是，上述公式都是针对样本的计算公式。当直接考虑随机变量 X 时，对应的计算公式只需将求平均换为期望即可，比如：

$$\bar{X} = \mathbb{E}[X].$$

例 2.11 给定一个正态分布的随机变量 X ，设其均值为 0，方差为 1，请计算其偏度和峰度。

解 对于均值为 0，方差为 1 的正态分布，其概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

由于正态分布的对称性，其偏度为

$$\text{Skew}(X) = \mathbb{E}[X^3] = \int_{-\infty}^{+\infty} x^3 f(x) dx = 0.$$

此外，根据 Wick's 定理¹，我们可以得到正态分布的峰度为

$$\text{Kurt}(X) = \mathbb{E}[X^4] = 3\mathbb{E}^2[X^2] = 3.$$

因此，对于标准正态分布，其偏度和峰度分别为

$$\text{Skew}(X) = 0, \quad \text{Kurt}(X) = 3.$$

接下来，我们尝试使用矩阵来表示这些统计量。均值对应的矩阵表示较为简单，即

$$\bar{x} = \frac{1}{n} \mathbf{x}^T \mathbf{1}, \quad (2.21)$$

其中 $\mathbf{1}$ 为一个全一的向量，其大小与 $\mathbf{x}$ 相同。当均值为 0 时，方差的矩阵表示如下：

$$\sigma^2 = \frac{1}{n} \mathbf{x}^T \mathbf{x}. \quad (2.22)$$

然而，三阶和四阶样本统计量显然没有办法通过简单的矩阵运算来表示，此时就需要引入张量。注意到，(2.21) 和 (2.22) 可以表示为如下的张量形式

$$\bar{x} = \frac{1}{n} \mathbf{1} \times_1 \mathbf{x}^T, \quad \sigma^2 = \frac{1}{n} \mathbf{I} \times_1 \mathbf{x}^T \times_2 \mathbf{x}^T, \quad (2.23)$$

¹需要注意的是 $\mathbb{E}[X^4] = 3\mathbb{E}^2[X^2]$ 仅在正态分布下成立。

其中 $\mathbf{I}$ 为单位矩阵。根据 (2.23)，可以自然地推导出三阶和四阶样本统计量的张量表示形式（均值为 0，方差为 1），如下

$$\begin{aligned}\text{Skew}(\mathbf{x}) &= \frac{1}{n} \mathcal{I}_3 \times_1 \mathbf{x}^T \times_2 \mathbf{x}^T \times_3 \mathbf{x}^T, \\ \text{Kurt}(\mathbf{x}) &= \frac{1}{n} \mathcal{I}_4 \times_1 \mathbf{x}^T \times_2 \mathbf{x}^T \times_3 \mathbf{x}^T \times_4 \mathbf{x}^T,\end{aligned}\quad (2.24)$$

其中 $\mathcal{I}_k \in \mathbb{R}^{n \times n \times \cdots \times n}$ 为 k 阶对角张量，其对角元素都为 1，即

$$\mathcal{I}_k(i, i, \dots, i) = 1, \quad i = 1, 2, \dots, n.$$

至此，我们可以给出任意一个均值为 0，方差为 1 的向量的任意阶样本统计量的张量表示形式：

$$S_k(\mathbf{x}) = \frac{1}{n} \mathcal{I}_k \times_1 \mathbf{x}^T \times_2 \mathbf{x}^T \cdots \times_k \mathbf{x}^T, \quad k \geq 3. \quad (2.25)$$

在实际应用中，往往需要处理多维数据，此时，我们更关心数据在某个投影下的统计信息。设有数据 $\mathbf{X} \in \mathbb{R}^{n \times p}$ ，其中 n 为样本数， p 为特征数。令投影向量为 $\mathbf{v} \in \mathbb{R}^{p \times 1}$ ，则投影后的为 $\mathbf{X}\mathbf{v} \in \mathbb{R}^{n \times 1}$ 。对于该向量，我们同样可以利用样本统计量的计算公式来计算相应的统计量。但注意到，对于均值，我们有

$$\text{Mean}(\mathbf{X}\mathbf{v}) = \frac{1}{n} (\mathbf{X}\mathbf{v})^T \mathbf{1} = \frac{1}{n} \mathbf{v}^T \mathbf{X}^T \mathbf{1} = \mathbf{v}^T \left(\frac{1}{n} \mathbf{X}^T \mathbf{1} \right).$$

这意味着，我们可以提前计算 $\mathbf{u} = \frac{1}{n} \mathbf{X}^T \mathbf{1}$ 并将其存储，而对于任意投影方向的均值计算，我们只需要计算 $\mathbf{v}^T \mathbf{u}$ 即可。当数据 $\mathbf{X}$ 的样本数远大于特征数 p 时，这种方法可以大大提高计算效率。

类似地，对于方差（设均值为 0），我们有

$$\text{Var}(\mathbf{X}\mathbf{v}) = \frac{1}{n} (\mathbf{X}\mathbf{v})^T (\mathbf{X}\mathbf{v}) = \frac{1}{n} \mathbf{v}^T \mathbf{X}^T \mathbf{X} \mathbf{v} = \mathbf{v}^T \boldsymbol{\Sigma} \mathbf{v},$$

其中 $\boldsymbol{\Sigma} = \frac{1}{n} \mathbf{X}^T \mathbf{X}$ 被称作协方差矩阵。可以看到，在得到协方差矩阵的基础上，任意方向上的方差都可以通过 $\mathbf{v}^T \boldsymbol{\Sigma} \mathbf{v}$ 来快速得到。换句话说，协方差矩阵包含了数据的所有二阶统计量信息。

对于高阶统计量，是否也存在某个矩阵或者张量包含了数据所有的高阶统计量信息呢？答案是肯定的，这里以三阶统计量偏度为例，注意到

$$\begin{aligned}\text{Skew}(\mathbf{X}\mathbf{v}) &= \frac{1}{n} \mathcal{I}_3 \times_1 (\mathbf{X}\mathbf{v})^T \times_2 (\mathbf{X}\mathbf{v})^T \times_3 (\mathbf{X}\mathbf{v})^T \\ &= \left(\frac{1}{n} \mathcal{I}_3 \times_1 \mathbf{X}^T \times_2 \mathbf{X}^T \times_3 \mathbf{X}^T \right) \times_1 \mathbf{v}^T \times_2 \mathbf{v}^T \times_3 \mathbf{v}^T \\ &= \mathcal{S} \times_1 \mathbf{v}^T \times_2 \mathbf{v}^T \times_3 \mathbf{v}^T,\end{aligned}$$

其中 $\mathcal{S} = \frac{1}{n} \mathcal{I}_3 \times_1 \mathbf{X}^T \times_2 \mathbf{X}^T \times_3 \mathbf{X}^T$ 是一个三阶张量，其大小为 $p \times p \times p$ 。

根据张量 k 模积的定义, 我们可以得到 $S(i, j, k)$ 的表达式如下

$$S(i, j, k) = \frac{1}{n} \sum_{a,b,c=1}^n \mathcal{I}_3(a, b, c) x_{ai} x_{bj} x_{ck} = \frac{1}{n} \sum_{l=1}^n x_{li} x_{lj} x_{lk},$$

其中 x_{ij} 表示矩阵 $\mathbf{X}$ 的第 i, j 个元素, 并且张量 $\mathcal{S}$ 是一个对称张量。进一步地, 令 S_l 为一个 $p \times p \times p$ 的张量, 其第 i, j, k 个元素为

$$S_l(i, j, k) = x_{li} x_{lj} x_{lk}.$$

根据外积的定义, 不难发现 S_l 是由向量 $\mathbf{x}_l = [x_{l1} \ x_{l2} \ \cdots \ x_{lp}]^T$ 的外积构成的, 即

$$S_l = \mathbf{x}_l \circ \mathbf{x}_l \circ \mathbf{x}_l.$$

并且向量 $\mathbf{x}_l$ 中的元素刚好对应矩阵 $\mathbf{X}$ 的第 l 行中的元素。此时, 张量 $\mathcal{S}$ 可以表示为

$$\mathcal{S} = \sum_{l=1}^n S_l = \frac{1}{n} \sum_{l=1}^n \mathbf{x}_l \circ \mathbf{x}_l \circ \mathbf{x}_l.$$

因此, 任意投影方向上的偏度可以表示为

$$\text{Skew}(\mathbf{X}\mathbf{v}) = \mathcal{S} \times_1 \mathbf{v}^T \times_2 \mathbf{v}^T \times_3 \mathbf{v}^T.$$

方便起见, 可以将上面的式子简化为

$$\text{Skew}(\mathbf{X}\mathbf{v}) = \mathcal{S}\mathbf{v}^3,$$

其中, $\mathcal{S}$ 被称作**协偏度张量**, 其包含了数据所有的三阶统计量信息。

有意思的是, 对于均值和方差, 同样有类似的表达式:

$$\mathbf{u} = \frac{1}{n} \sum_{l=1}^n \mathbf{x}_l, \quad \Sigma = \frac{1}{n} \sum_{l=1}^n \mathbf{x}_l \circ \mathbf{x}_l.$$

任意方向上的均值和方差可以表示为

$$\text{Mean}(\mathbf{X}\mathbf{v}) = \mathbf{u} \times_1 \mathbf{v}^T, \quad \text{Var}(\mathbf{X}\mathbf{v}) = \Sigma \times_1 \mathbf{v}^T \times_2 \mathbf{v}^T.$$

因此, 对于数据 $\mathbf{X}$, 我们可以类似地计算其任意的高阶协方差张量 (均值为零, 方差为一):

$$\mathcal{S}_k = \frac{1}{n} \sum_{l=1}^n \mathbf{x}_l^{\circ k},$$

其中 $\mathcal{S}_k$ 为数据的 k 阶统计张量。基于 $\mathcal{S}_k$, 数据在任意方向上的投影的统计量可以表示为

$$S_k(\mathbf{X}\mathbf{v}) = \mathcal{S}_k \mathbf{v}^k.$$