

# 15

## Singular Value Decomposition

# 奇异值分解

应该是最重要的矩阵分解，没有之一



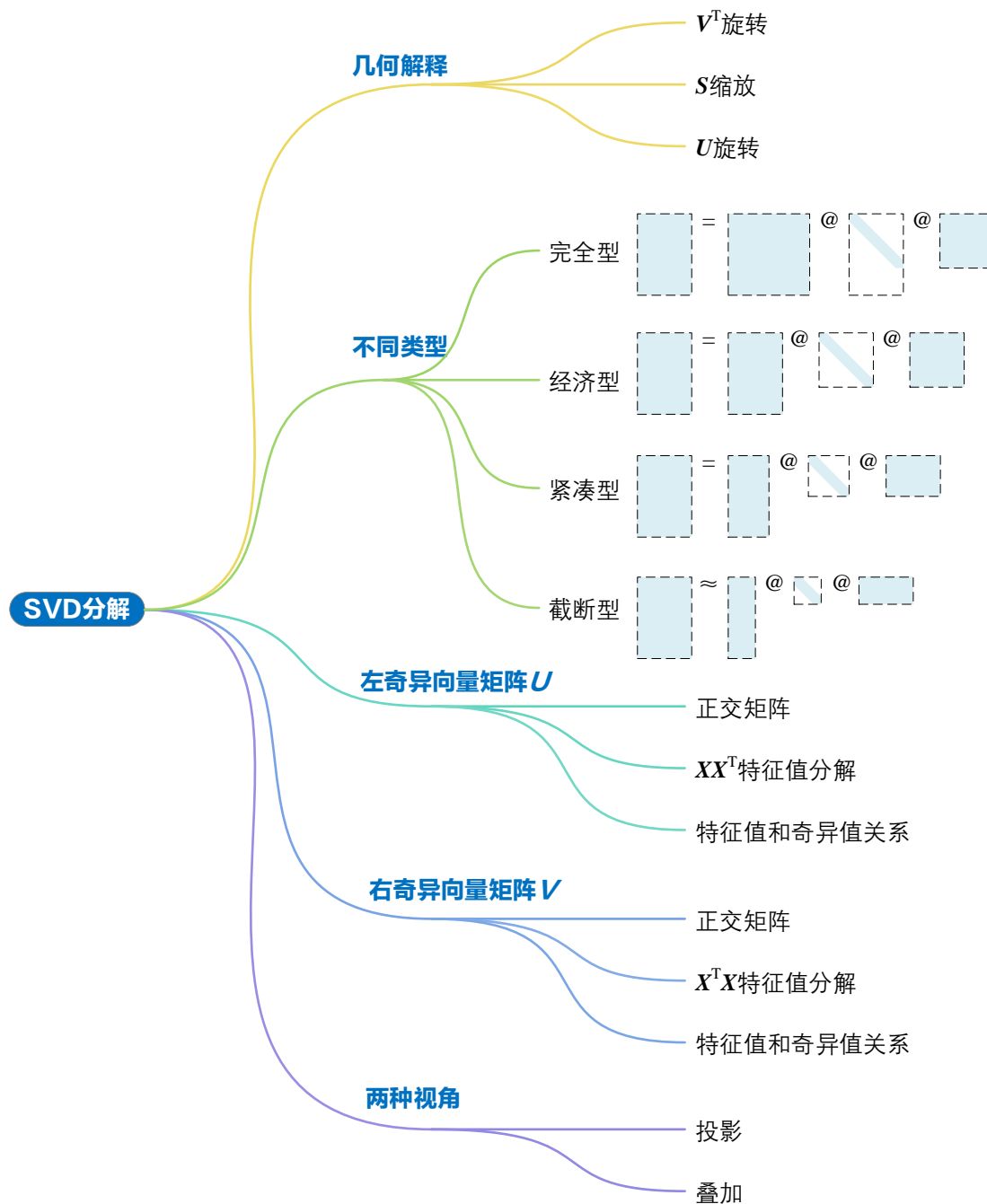
就我而言，我一无所知，但满眼的繁星让我入梦。

*For my part I know nothing with any certainty, but the sight of the stars makes me dream.*

—— 文森特·梵高 (Vincent van Gogh) | 荷兰后印象派画家 | 1853 ~ 1890



- ◀ `matplotlib.pyplot.quiver()` 绘制箭头图
- ◀ `numpy.linspace()` 在指定的间隔内, 返回固定步长的数据
- ◀ `numpy.linalg.svd()` 进行 SVD 分解
- ◀ `numpy.diag()` 以一维数组的形式返回方阵的对角线元素, 或将一维数组转换成对角阵



## 15.1 几何视角：旋转 → 缩放 → 旋转

本书第 11 章简要介绍过**奇异值分解** (Singular Value Decomposition, SVD)，这个宇宙中最重要的矩阵分解。本节将从几何视角解剖奇异值分解。

对数据矩阵  $\mathbf{X}_{n \times D}$  奇异值分解得到：

$$\mathbf{X}_{n \times D} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (1)$$

其中， $\mathbf{S}$  为对角阵，其主对角线元素  $s_j (j = 1, 2, \dots, D)$  为**奇异值** (singular value)。

**⚠ 注意**，SVD 分解得到的奇异值非负，即  $s_j \geq 0$ 。此外注意，(1) 中矩阵  $\mathbf{V}$  的转置运算。

$\mathbf{U}$  的列向量称作**左奇异向量** (left singular vector)。

$\mathbf{V}$  的列向量称作**右奇异向量** (right singular vector)。

在完全型 SVD 分解中， $\mathbf{U}$  和  $\mathbf{V}$  为正交矩阵，即  $\mathbf{U}$  和自己转置  $\mathbf{U}^T$  的乘积为单位矩阵； $\mathbf{V}$  和自己转置  $\mathbf{V}^T$  的乘积也是单位矩阵。SVD 分解有四种主要形式，完全型是其中一种。

从向量空间角度来看， $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$  为  $\mathbb{R}^n$  的规范正交基， $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$  为  $\mathbb{R}^D$  的规范正交基。

根据这三个矩阵的形态，我们知道，从几何视角来看，正交矩阵  $\mathbf{U}$  和  $\mathbf{V}$  矩阵作用是旋转，而对角矩阵  $\mathbf{S}$  的作用是缩放。

大家可能会问这和特征值分解对应的“旋转 → 缩放 → 旋转”有何不同？

特征值分解中，三步几何变换是旋转 ( $\mathbf{V}^{-1}$ ) → 缩放 ( $\mathbf{A}$ ) → 旋转 ( $\mathbf{V}$ )。

奇异值分解中，三步几何变换是旋转 ( $\mathbf{V}^T$ ) → 缩放 ( $\mathbf{S}$ ) → 旋转 ( $\mathbf{U}$ )。一个明显的区别是， $\mathbf{V}^T$  的旋转发生在  $\mathbb{R}^D$  空间， $\mathbf{U}$  的旋转则发生在  $\mathbb{R}^n$  空间。

### 几何视角

为了方便解释，我们用  $2 \times 2$  矩阵  $\mathbf{A}$  做例子。

利用矩阵  $\mathbf{A}$  完成  $\mathbf{z} \rightarrow \mathbf{x}$  线性映射，即  $\mathbf{x} = \mathbf{A}\mathbf{z}$ 。利用 SVD 分解，将  $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$  代入映射运算得到：

$$\mathbf{A}\mathbf{z} = \mathbf{U} \underset{\text{Rotate Scale Rotate}}{\mathbf{S}} \mathbf{V}^T \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (2)$$

图 1 所示为几何角度解释奇异值分解， $\mathbf{A}$  乘  $\mathbf{x}$ ，相当于先用  $\mathbf{V}^T$  旋转，再用  $\mathbf{S}$  缩放，最后用  $\mathbf{U}$  旋转。

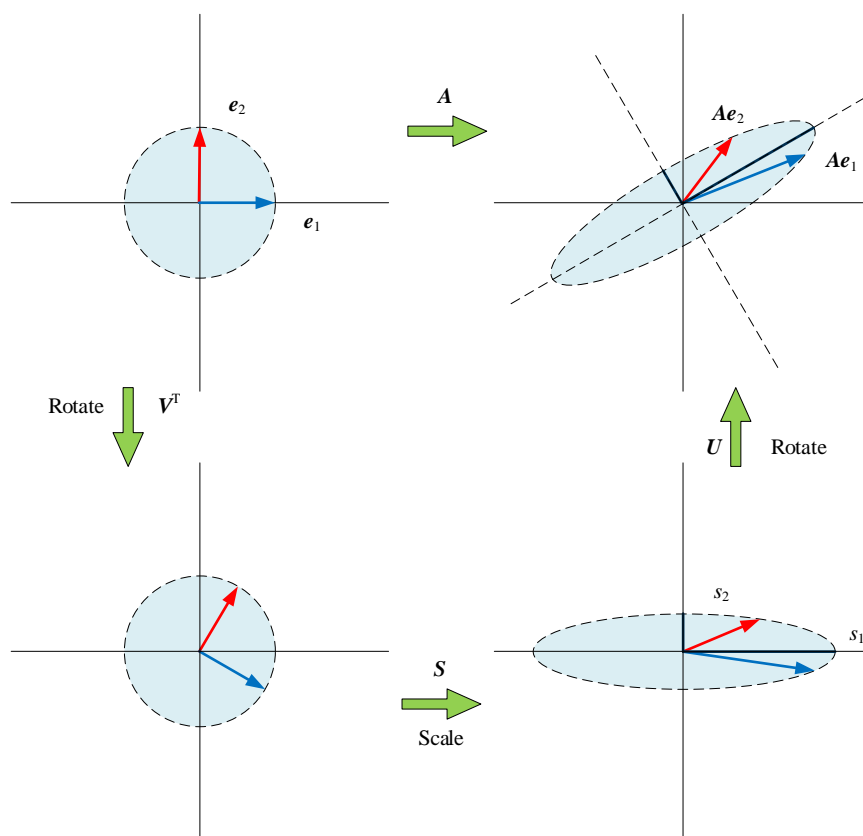


图 1. 几何角度解释奇异值分解

### 举个实例

下面用具体实例解释图 1。给定如下  $2 \times 2$  矩阵  $A$ ：

$$A = \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \quad (3)$$

对矩阵  $A$  进行 SVD 分解：

$$A = USV^T = \underbrace{\begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}}_S \underbrace{\begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}}_{V^T} \quad (4)$$

即，

$$U = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}, \quad S = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad V = \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \quad (5)$$

▲ 注意，如果特征值分解和奇异值分解的对象都是可对角矩阵，两个分解得到的结果等价。但是，奇异值分解的强大之处在于，任何实数矩阵都可以奇异值分解。

给定  $e_1$  和  $e_2$  两个单位向量：

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (6)$$

$\mathbf{e}_1$  和  $\mathbf{e}_2$  经过  $A$  转换分别得到：

$$\begin{aligned} A\mathbf{e}_1 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\ A\mathbf{e}_2 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix} \end{aligned} \quad (7)$$

图 2 所示为转换前后的结果对比。请大家注意转换前后向量的方向和长度(模)的变化。

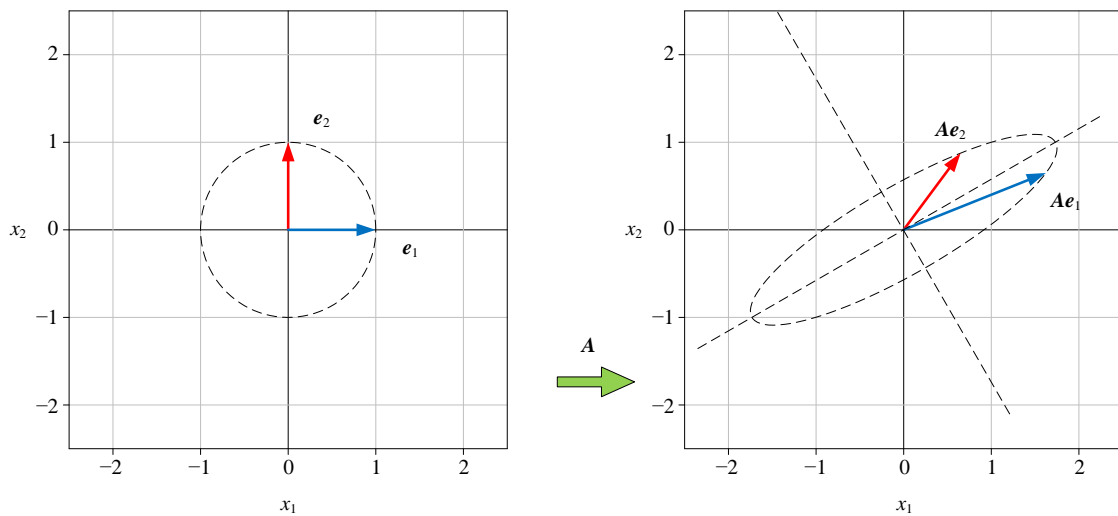


图 2.  $\mathbf{e}_1$  和  $\mathbf{e}_2$  经过  $A$  线性转换

## 分步几何变换

(7) 等价于“旋转 ( $V^T$ )  $\rightarrow$  缩放 ( $S$ )  $\rightarrow$  旋转 ( $U$ )”，具体如图 3 所示。

$\mathbf{e}_1$  和  $\mathbf{e}_2$  两个向量先通过  $V^T$  进行旋转，得到：

$$\begin{aligned} V^T \mathbf{e}_1 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} \\ V^T \mathbf{e}_2 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} \end{aligned} \quad (8)$$

在 (8) 基础上，再用对角矩阵  $S$  进行缩放，得到：

$$\begin{aligned} SV^T e_1 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} \\ SV^T e_2 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.433 \end{bmatrix} \end{aligned} \quad (9)$$

在之前“旋转”和“缩放”两步基础上，最后再利用  $U$  进行旋转，得到：

$$\begin{aligned} USV^T e_1 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\ USV^T e_2 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1 \\ 0.433 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix} \end{aligned} \quad (10)$$

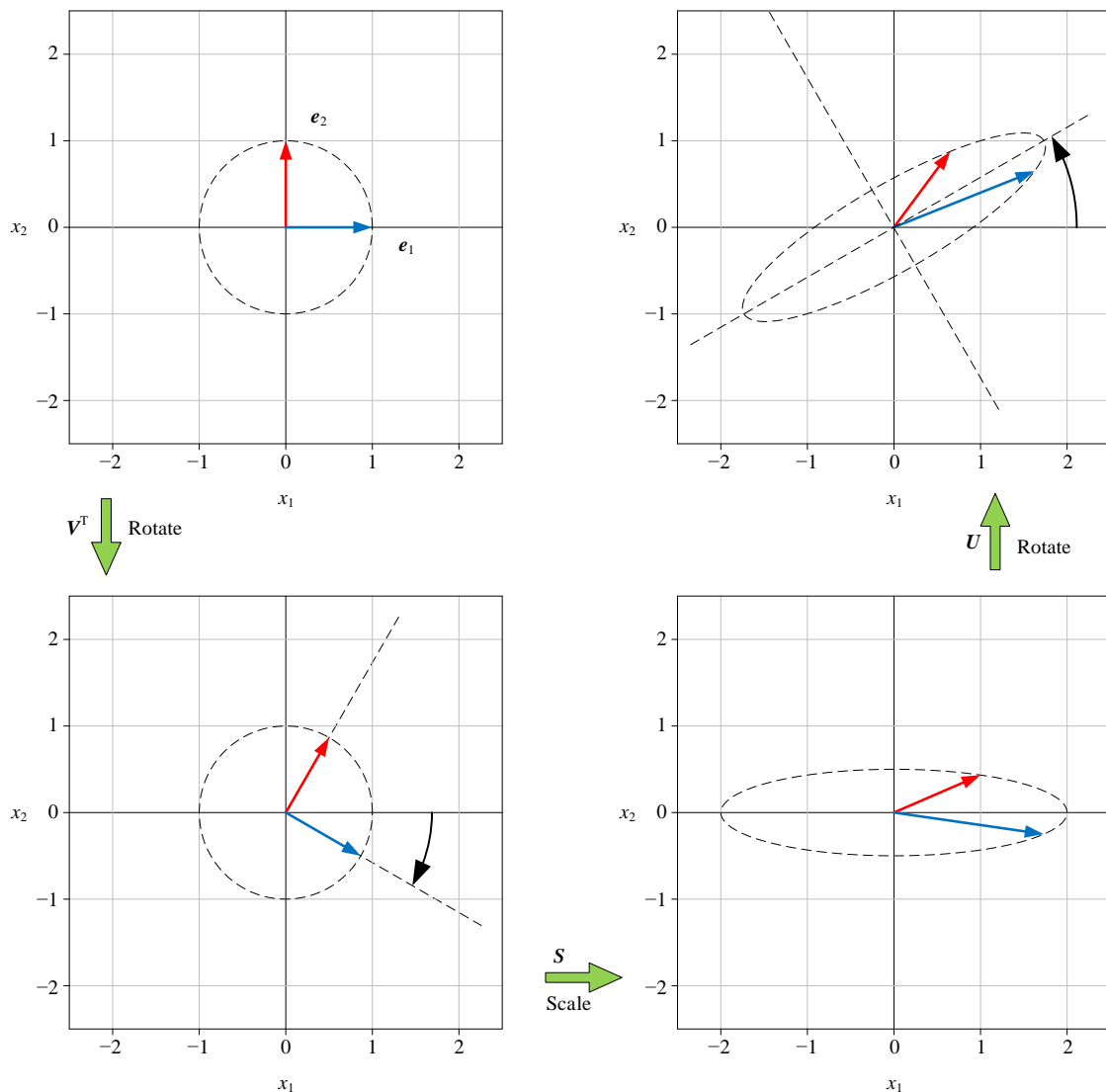


图 3.  $e_1$  和  $e_2$  分别经过  $V^T$ 、 $S$  和  $U$  转换



Bk4\_Ch15\_01.py 绘制图 3 所有子图。

## 15.2 不同类型 SVD 分解

SVD 分解分为**完全型** (full)、**经济型** (economy-size, thin)、**紧凑型** (compact) 和**截断型** (truncated) 四大类。

本节将简要介绍完全型和经济型两种奇异值分解之间的关系。下一章将深入讲解这四种 SVD 分解。

### 完全型

图 4 所示为完全型 SVD 分解热图，其中左奇异值矩阵  $U$  为方阵，形状为  $n \times n$ 。 $S$  的形状和  $X$  相同，为  $n \times D$ 。 $S$  的主对角线元素  $s_i$  为奇异值，具体形式为：

$$S = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \quad (11)$$

约定俗成，这  $D$  个奇异值的关系为  $s_1 \geq s_2 \geq \dots \geq s_D$ 。

如图 4 所示， $S$  可以分块为上下两个子块——对角方阵、全 0 矩阵  $O$ 。

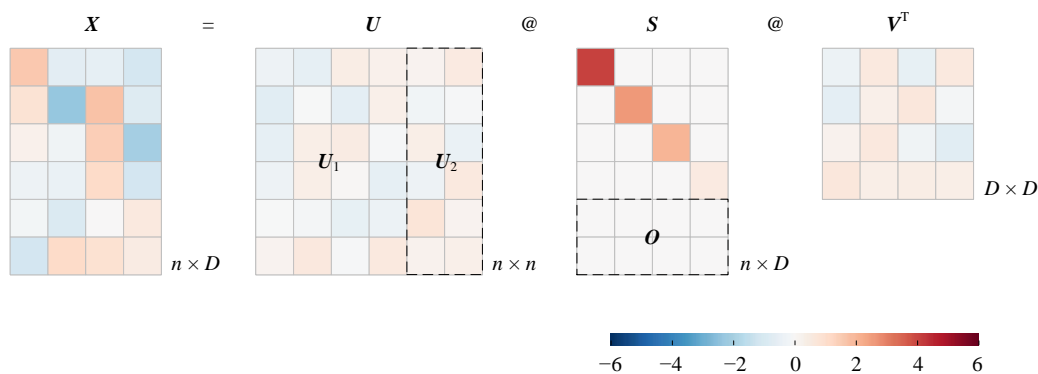


图 4. 矩阵  $X$  的完全型 SVD 分解

注意，一般情况，数据矩阵为“细高”长方形，偶尔大家也会见到“宽矮”长方形的数据矩阵。  
(1) 中  $\mathbf{X}$  为细高长方形，对  $\mathbf{X}$  转置便得到宽矮长方形矩阵  $\mathbf{X}^T$ 。如图 5 所示，相应的， $\mathbf{X}^T$  的 SVD 分解为：

$$\mathbf{X}^T = \mathbf{V} \mathbf{S}^T \mathbf{U}^T \quad (12)$$

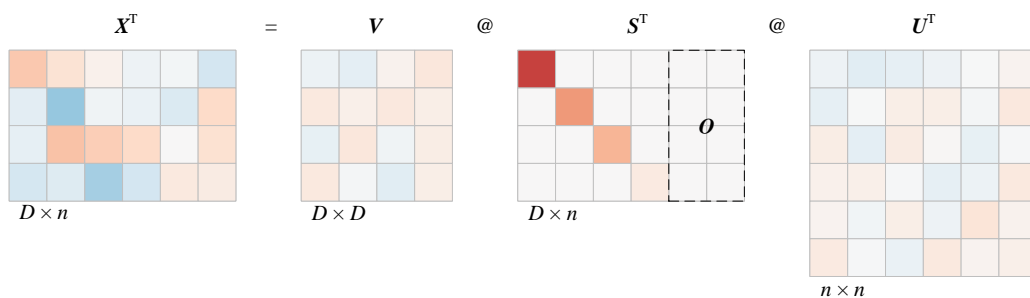


图 5. 矩阵  $\mathbf{X}^T$  的完全型 SVD 分解

## 经济型

图 6 所示为经济型 SVD 分解结果热图。可以发现，左奇异值矩阵  $\mathbf{U}$  形状和  $\mathbf{X}$  相同，均为  $n \times D$ 。而  $\mathbf{S}$  为方阵，形状为  $D \times D$ 。从图 4 到图 6，利用的是分块矩阵乘法，这个话题留到下一章讨论。

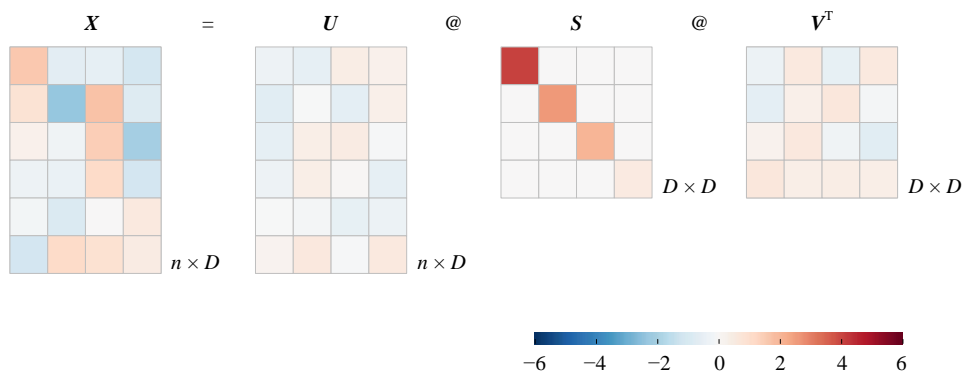


图 6. 经济型 SVD 分解

在经济型 SVD 分解中， $\mathbf{S}$  为对角方阵：

$$\mathbf{S} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \quad (13)$$

当  $\mathbf{S}$  为对角方阵时，(12) 可以写成：



$$\mathbf{X}^T = \mathbf{V} \mathbf{S} \mathbf{U}^T \quad (14)$$



Bk4\_Ch15\_02.py 中 Bk4\_Ch15\_02\_A 部分绘制图 4 和图 6。

## 15.3 左奇异向量矩阵 $\mathbf{U}$

$\mathbf{U}$  的列向量称作**左奇异向量** (left singular vector),  $\mathbf{U}$  和自己转置  $\mathbf{U}^T$  的乘积为单位矩阵:

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (15)$$

如图 7 所示, 对于完全型 SVD 分解,  $\mathbf{U}$  为方阵。

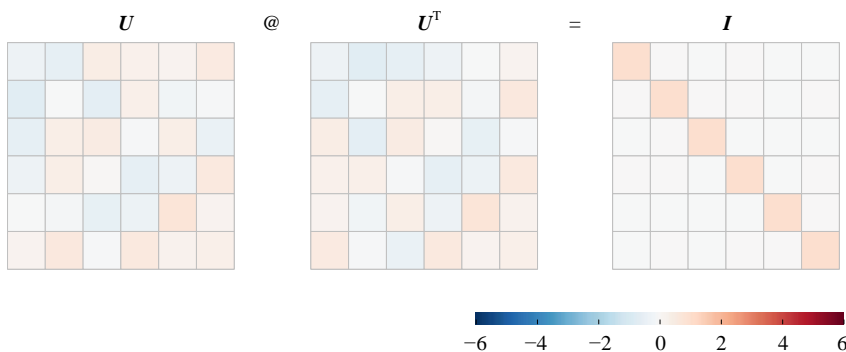


图 7.  $\mathbf{U}$  和自己转置  $\mathbf{U}^T$  的乘积为单位矩阵

### 特征值分解

本书前文提到过两次, 细高的长方形矩阵  $\mathbf{X}$  不能进行特征值分解。但是, 它的格拉姆矩阵  $\mathbf{X}^T \mathbf{X}$  和  $\mathbf{X} \mathbf{X}^T$  都是对称矩阵, 可以进行特征值分解。下面, 我们先分析  $\mathbf{X} \mathbf{X}^T$ 。

图 8 所示为  $\mathbf{X}$  和自己转置  $\mathbf{X}^T$  相乘得到第一个格拉姆矩阵  $\mathbf{X} \mathbf{X}^T$  的热图,  $\mathbf{X} \mathbf{X}^T$  为  $n \times n$  方阵。

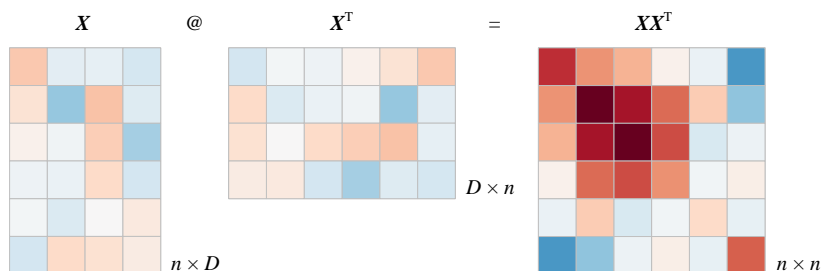
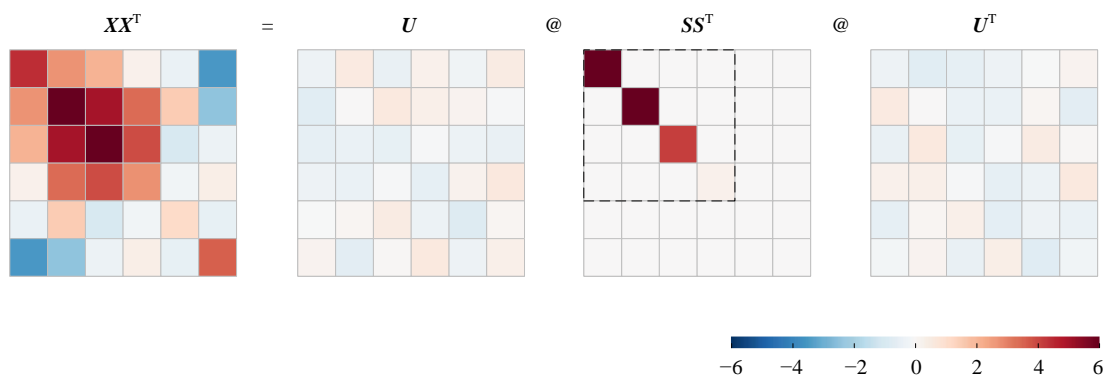


图 8.  $X$  和自己转置  $X^T$  的乘积热图

对方阵  $XX^T$  进行特征值分解，可以发现  $U$  的列向量是特征向量，而  $SS^T$  是  $XX^T$  的特征值矩阵：

$$\begin{aligned} XX^T &= (USV^T)(USV^T)^T \\ &= US(V^TV)S^TU^T \\ &= USS^TU^T \end{aligned} \quad (16)$$

图 9 所示为  $X^TX$  特征值分解热图。

图 9. 对  $X^TX$  特征值分解

$SS^T$  主对角线为特征值，对  $SS^T$  展开得到：

$$SS^T = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2 \\ & & & & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \\ & & & & \ddots \\ & & & & & \lambda_n \end{bmatrix} \quad (17)$$

观察上式，发现当  $j = 1 \sim D$  时，特征值  $\lambda_j$  和奇异值  $s_j$  存在如下关系：

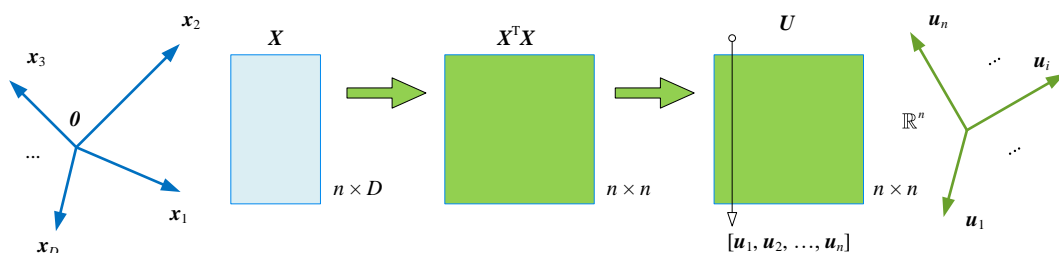
$$\lambda_j = s_j^2 \quad (18)$$

特别地，在经济型 SVD 分解中， $S$  为对角方阵， $XX^T$  特征值分解可以写成：

$$XX^T = US^2U^T \quad (19)$$

## 向量空间

如图 10 所示， $XX^T$  进行特征值分解得到正交矩阵  $U = [u_1, u_2, \dots, u_n]$  是个规范正交基，张起的空间为  $\mathbb{R}^n$ 。

图 10. 对 Gram 矩阵  $XX^T$  特征值分解得到规范正交基  $U$ 

### 类比 QR 分解

数据矩阵  $X$  进行 QR 分解得到：

$$X = QR \quad (20)$$

对于完全型 QR 分解， $Q$  为正交矩阵，也是一个规范正交基  $[q_1, q_2, \dots, q_D]$ 。

对  $X$  进行完全型 SVD 分解，把结果写成：

$$X = U(SV^T) \quad (21)$$

对比 (20) 和 (21)， $Q$  和  $U$  都是正交矩阵，形状虽然相同，但是两者显然是不同的规范正交基。

对于 QR 分解， $x_1$  和  $q_1$  平行。打个比方， $x_1$  像是一个锚，确定了  $[q_1, q_2, \dots, q_D]$  的空间位置。

而 SVD 分解则引入了一个优化视角——逐个最大化奇异值。



本书第 18 章将深入介绍这个优化视角。

对比 (20) 和 (21)， $R$  则对应  $SV^T$ 。特别地， $SV^T$  结果正交，即  $SV^T(SV^T)^T = SV^T VS = S^2$ 。



Bk4\_Ch15\_02.py 中 Bk4\_Ch15\_02\_B 部分绘制图 7。请读者自行编写代码绘制图 8 和图 9。

## 15.4 右奇异向量矩阵 $V$

$V$  的列向量称作**右奇异向量** (right singular vector)， $V$  和其转置  $V^T$  的乘积也是单位矩阵：

$$V^T V = I \quad (22)$$

图 11 所示为上式运算对应热图。

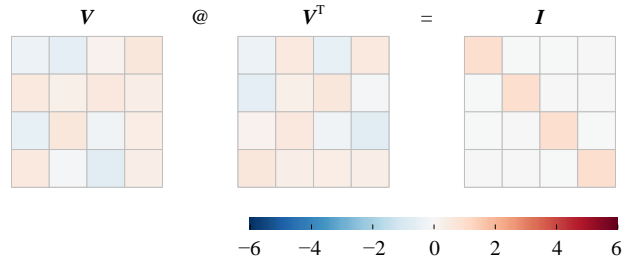


图 11.  $V$  和其转置  $V^T$  的乘积也是单位矩阵

## 特征值分解

图 12 所示为转置  $X^T$  和  $X$  相乘得到第二个格拉姆矩阵  $X^T X$  的热图， $X^T X$  为  $D \times D$  方阵。

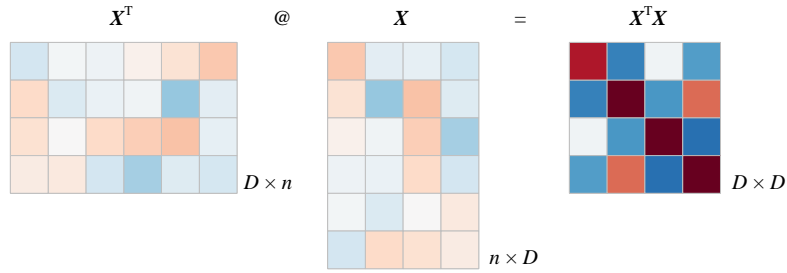


图 12. 转置  $X^T$  和  $X$  乘积热图

对  $X^T X$  特征值分解得到：

$$\begin{aligned} X^T X &= (USV^T)^T (USV^T) \\ &= VS^T (U^T U) S V^T \\ &= VS^T S V^T \end{aligned} \quad (23)$$

$V$  是  $X^T X$  的特征向量矩阵， $S^T S$  为特征值矩阵。图 13 所示为对  $X^T X$  进行特征值分解热图。

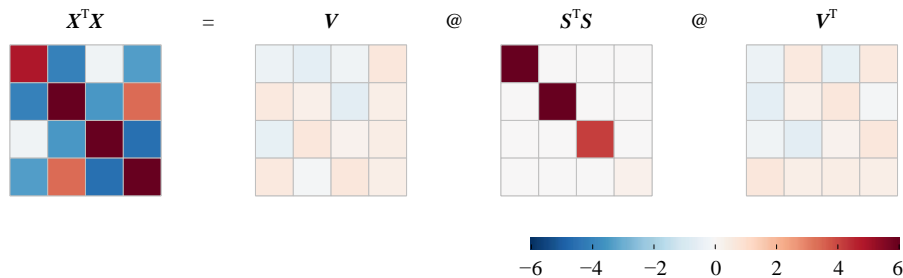


图 13. 对  $X^T X$  进行特征值分解

特别地，在经济型 SVD 分解中， $S$  为对角方阵；因此  $X^T X$  特征值分解可以写成：

$$X^T X = VS^2V^T \quad (24)$$

前文提过，经济型 SVD 分解中， $S$  为对角方阵， $S$  主对角元素  $s_i$  为奇异值：

$$S = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \quad (25)$$

对  $X^T X$  进行特征值分解， $S^2$  为特征值矩阵，奇异值和特征值也存在如下平方关系：

$$S^2 = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} \quad (26)$$

(25) 和 (26) 奇异值和特征值之间的关系如图 14 所示。



本书第 24 章将总结分解对象不同时，奇异值和特征值之间的联系和差异。

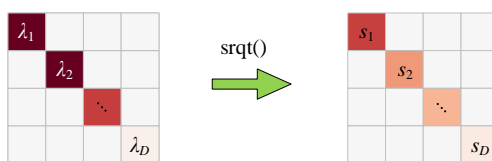


图 14. 奇异值和特征值之间关系

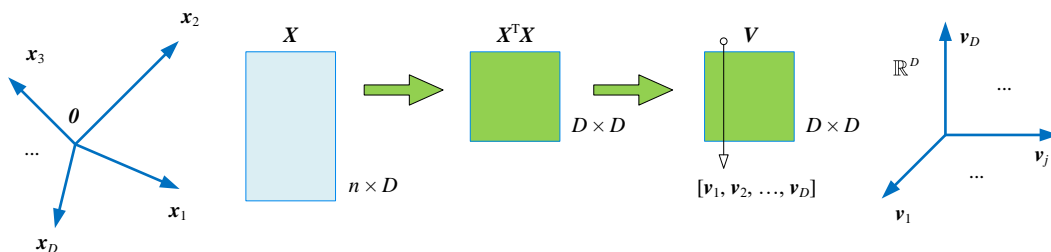
## 向量空间

如图 10 所示， $X^T X$  进行特征值分解得到正交矩阵  $V = [v_1, v_2, \dots, v_D]$ ，它也是个规范正交基，张起的空间为  $\mathbb{R}^D$ 。

奇异值分解强大之处在于，它可以分解各种形状实数矩阵，并且一次性获得  $U = [u_1, u_2, \dots, u_n]$  和  $V = [v_1, v_2, \dots, v_D]$  两个规范正交基。



大家可能好奇， $U$  和  $V$  这两个规范正交基有怎么样的联系、区别、应用？这个问题留给本书最后三章回答。

图 15. 对 Gram 矩阵  $X^T X$  特征值分解得到规范正交基  $V$ 

Bk4\_Ch15\_02.py 中 Bk4\_Ch15\_02\_C 部分绘制图 11。请读者自行编写代码绘制图 12 和图 13。

## 15.5 两个视角：投影和数据叠加

本节用两个数据视角观察 SVD 分解。这两个视角对应两种不同的矩阵乘法展开方式。

### 投影

对于经济型 SVD 分解，将 (1) 等式左右两侧右乘  $V$ ，可以得到：

$$X_{n \times D} V = US \quad (27)$$

将  $V$  和  $U$  本身分别写成左右排列的列向量：

$$X_{n \times D} \begin{bmatrix} v_1 & v_2 & \cdots & v_D \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_D \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \quad (28)$$

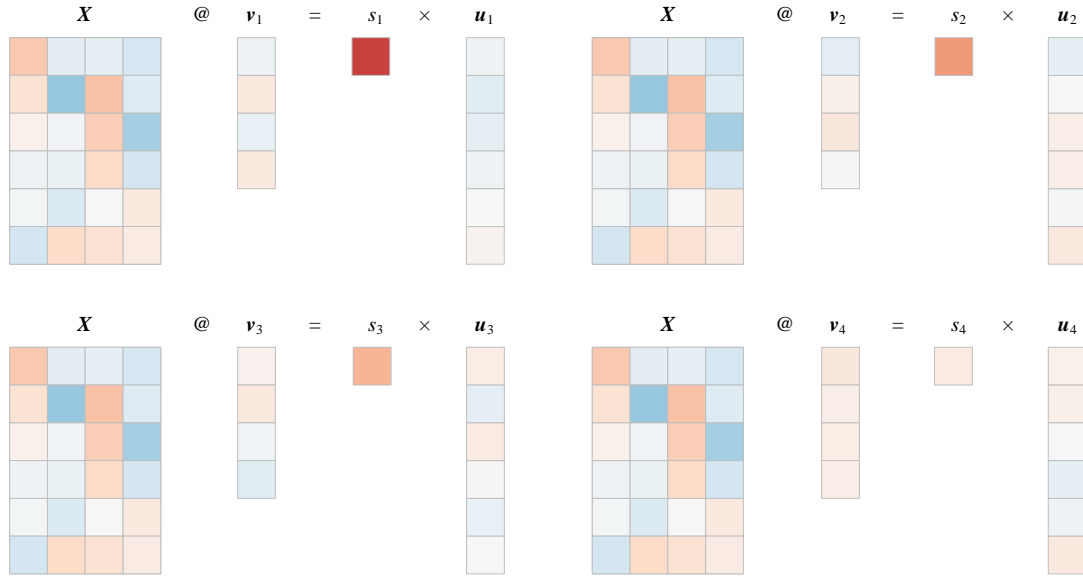
(28) 进一步展开得到：

$$\begin{bmatrix} Xv_1 & Xv_2 & \cdots & Xv_D \end{bmatrix} = \begin{bmatrix} s_1 u_1 & s_2 u_2 & \cdots & s_D u_D \end{bmatrix} \quad (29)$$

因此，

$$Xv_j = s_j u_j \quad (30)$$

上式可以理解为  $X$  向  $v_j$  投影，结果为  $s_j u_j$ 。对应运算热图如图 16 所示。

图 16.  $X$  向  $v_j$  映射结果为  $s_j u_j$ 

(30) 左右都是向量，等式两侧分别求模，即  $L^2$  范数，得到：

$$\|Xv_j\| = \|s_j u_j\| = s_j \quad (31)$$

也就是说  $Xv_j$  的模为对应奇异值  $s_j$ 。由于奇异值  $s_1$  到  $s_4$  从大到小排列，也就是说  $Xv_1$  的模最大。这个角度对于理解[主成分分析](#) (principal component analysis, PCA) 极为重要。

## 叠加

第二种展开方式如下：

$$\begin{aligned} X_{m \times D} &= \begin{bmatrix} u_1 & u_2 & \cdots & u_D \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_D^T \end{bmatrix} \\ &= \begin{bmatrix} s_1 u_1 & s_2 u_2 & \cdots & s_D u_D \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_D^T \end{bmatrix} = s_1 u_1 v_1^T + s_2 u_2 v_2^T + \cdots + s_D u_D v_D^T \end{aligned} \quad (32)$$

举个例子，对于  $D=4$  时：

$$X = s_1 u_1 v_1^T + s_2 u_2 v_2^T + s_3 u_3 v_3^T + s_4 u_4 v_4^T \quad (33)$$

(33) 中奇异值  $s_1$  到  $s_4$  从大到小排列，即  $s_1 \geq s_2 \geq s_3 \geq s_4$ 。

**▲** 注意， $s_j u_j v_j^T$  的秩为 1。

如图 17 所示，可以发现对应 (33) 等式右侧从左到右的四项相当于逐步还原  $X$ 。特别地，请大家注意图 17 左侧四副热图由上到下颜色逐渐变浅。下一章会深入介绍通过叠加还原原始数据矩阵。

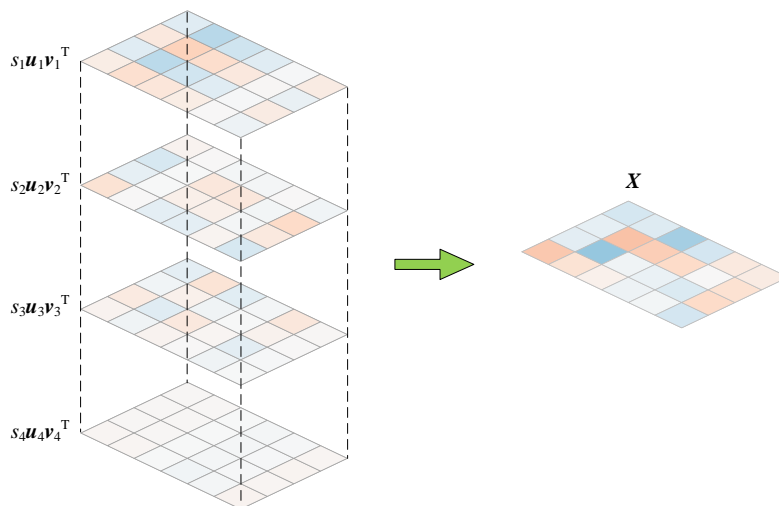


图 17. 四幅热图叠加还原原始图像

## 张量积

再进一步，利用 (30) 给出的关系，我们将 (33) 写成张量积之和的形式：

$$\begin{aligned}
 X &= s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2^T + s_3 \mathbf{u}_3 \mathbf{v}_3^T + s_4 \mathbf{u}_4 \mathbf{v}_4^T \\
 &= X \mathbf{v}_1 \mathbf{v}_1^T + X \mathbf{v}_2 \mathbf{v}_2^T + X \mathbf{v}_3 \mathbf{v}_3^T + X \mathbf{v}_4 \mathbf{v}_4^T \\
 &= X (\mathbf{v}_1 \mathbf{v}_1^T + \mathbf{v}_2 \mathbf{v}_2^T + \mathbf{v}_3 \mathbf{v}_3^T + \mathbf{v}_4 \mathbf{v}_4^T) \\
 &= X (\mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{v}_2 \otimes \mathbf{v}_2 + \mathbf{v}_3 \otimes \mathbf{v}_3 + \mathbf{v}_4 \otimes \mathbf{v}_4)
 \end{aligned} \tag{34}$$

这就是本书第 10 章讲解的“二次投影”再“层层叠加”。

➡ 能完成类似 (34) 投影的规范正交基有无数组，为什么  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$  脱颖而出？ $\mathbf{V}$  的特殊性体现在哪？回答这个问题需要优化方面的知识，这是本书第 18 章要探讨的话题。



Bk4\_Ch15\_02.py 中 Bk4\_Ch15\_02\_D 部分绘制本节图像。



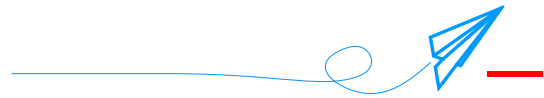


图 18 四副子图总结本章主要内容。请大家特别注意，奇异值分解对应“旋转 → 缩放 → 旋转”，不同于特征值分解的“旋转 → 缩放 → 旋转”。

任何实数矩阵都可以进行奇异值分解，但是只有可对角矩阵才能进行特征值分解。此外，奇异值分解得到的两个正交矩阵  $U$  和  $V$  一般形状不同。

格拉姆矩阵是奇异值分解和特征值分解的桥梁，这一点本书后面还要提到。请大家注意，特征值和奇异值之间的关系。

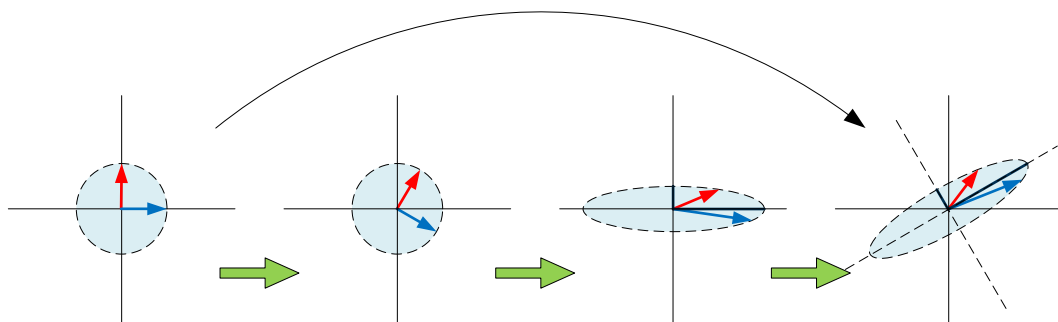


图 18. 总结本章重要内容的四副图