15 Singular Value Decomposition 奇异值分解

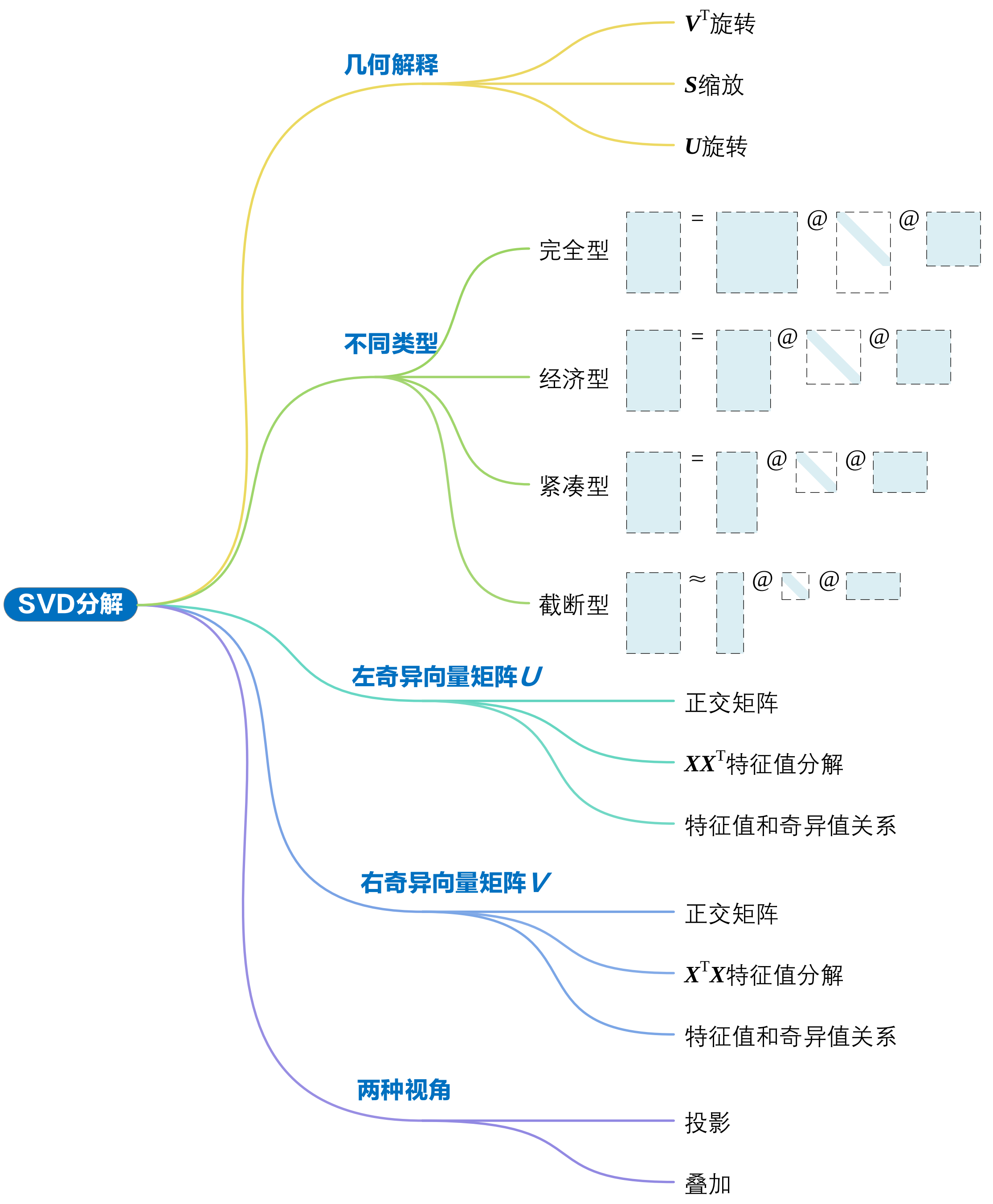
最重要的矩阵分解，没有之一

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


SVD分解
几何解释
V^T旋转
S缩放
U旋转
不同类型
完全型
经济型
紧凑型
截断型
=
≈
@
左奇异向量矩阵U
正交矩阵
XX^T特征值分解
特征值和奇异值关系
右奇异向量矩阵V
正交矩阵
X^TX特征值分解
特征值和奇异值关系
两种视角
投影
叠加

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) —— 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 $\boldsymbol{X}_{n\times D}$ 奇异值分解得到：

$$\boldsymbol{X}_{n\times D} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{1}$$

其中，$\boldsymbol{S}$ 为对角阵，其主对角线元素 s_j ($j = 1, 2, \ldots, D$) 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 $\boldsymbol{V}$ 的转置运算。

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

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

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

从向量空间角度来看，$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 为 $\mathbb{R}^n$ 的规范正交基，$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 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 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 矩阵作用是旋转，而对角矩阵 $\boldsymbol{S}$ 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 ($\boldsymbol{V}^{-1}$) → 缩放 ($\boldsymbol{\Lambda}$) →旋转 ($\boldsymbol{V}$)。

奇异值分解中，三步几何变换是旋转 ($\boldsymbol{V}^{\mathrm{T}}$) → 缩放 ($\boldsymbol{S}$) →旋转 ($\boldsymbol{U}$)。一个明显的区别是，$\boldsymbol{V}^{\mathrm{T}}$ 的旋转发生在 $\mathbb{R}^D$ 空间，$\boldsymbol{U}$ 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 × 2 矩阵 $\boldsymbol{A}$ 做例子。

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

$$\boldsymbol{A}\boldsymbol{z} = \underset{\text{Rotate}}{\boldsymbol{U}}\ \underset{\text{Scale}}{\boldsymbol{S}}\ \underset{\text{Rotate}}{\boldsymbol{V}^{\mathrm{T}}}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2}$$

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

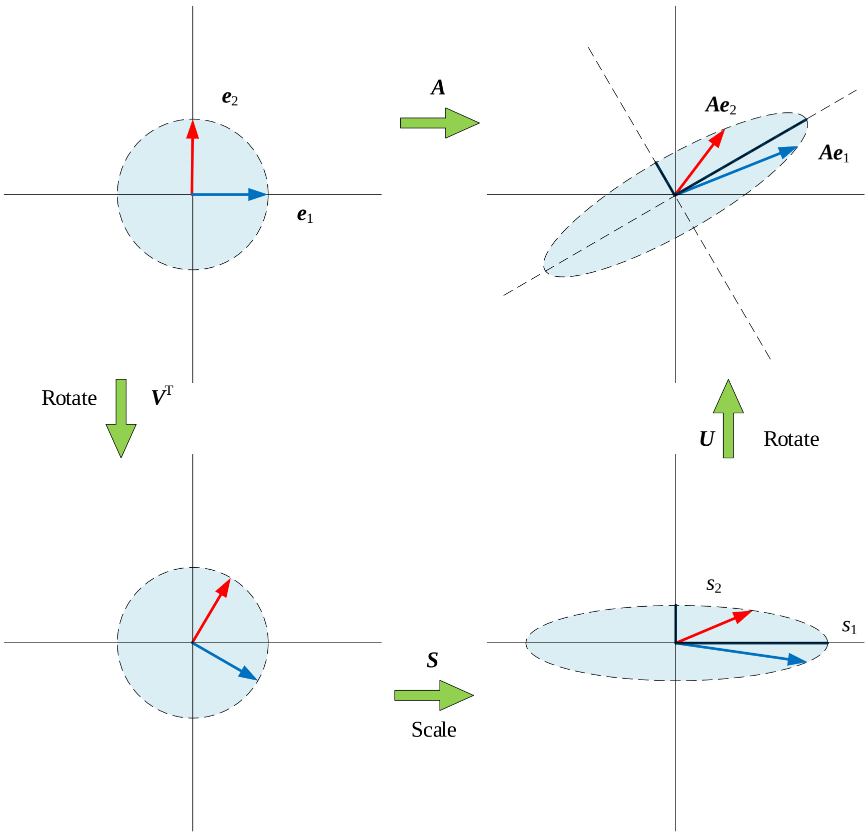


图 1. 几何角度解释奇异值分解

举个实例

下面用具体实例解释图 1。给定如下 2 × 2 矩阵 A:

$$A = \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \tag{3}$$

对矩阵 A 进行 SVD 分解:

$$A = USV^{\mathrm{T}} = \underbrace{\begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}}_{V^{\mathrm{T}}} \tag{4}$$

即,

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

⚠ 注意，如果特征值分解和奇异值分解的对象都是可对角化矩阵，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:

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{6}$$

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

$$\begin{aligned} \boldsymbol{A}\boldsymbol{e}_1 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\ \boldsymbol{A}\boldsymbol{e}_2 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix} \end{aligned} \tag{7}$$

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。

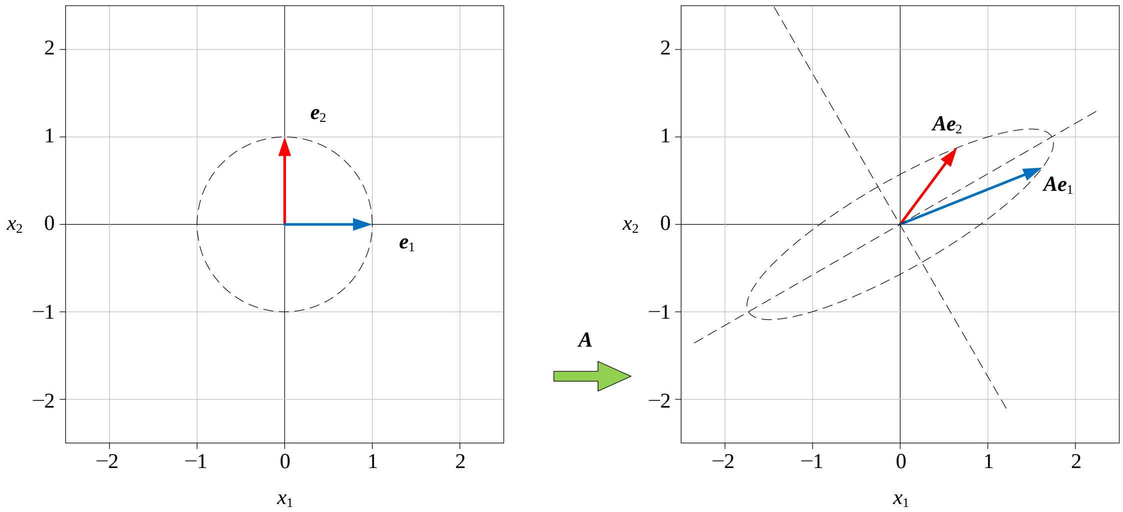


图 2. $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 经过 $\boldsymbol{A}$ 线性转换

分步几何变换

(7) 等价于“旋转 ($\boldsymbol{V}^{\mathrm{T}}$) → 缩放 ($\boldsymbol{S}$) → 旋转 ($\boldsymbol{U}$)”，具体如图 3 所示。

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

$$\begin{aligned} \boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} \\ \boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} \end{aligned} \tag{8}$$

在 (8) 基础上，再用对角矩阵 $\boldsymbol{S}$ 进行缩放，得到：

$$\begin{aligned} \boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} \\ \boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.433 \end{bmatrix} \end{aligned} \tag{9}$$

在之前“旋转 ($\boldsymbol{V}^{\mathrm{T}}$)”和“缩放 ($\boldsymbol{S}$)”两步基础上，最后再利用 $\boldsymbol{U}$ 进行旋转，得到：

$$\begin{aligned} \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}\begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\ \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}\begin{bmatrix} 1 \\ 0.433 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix} \end{aligned} \tag{10}$$

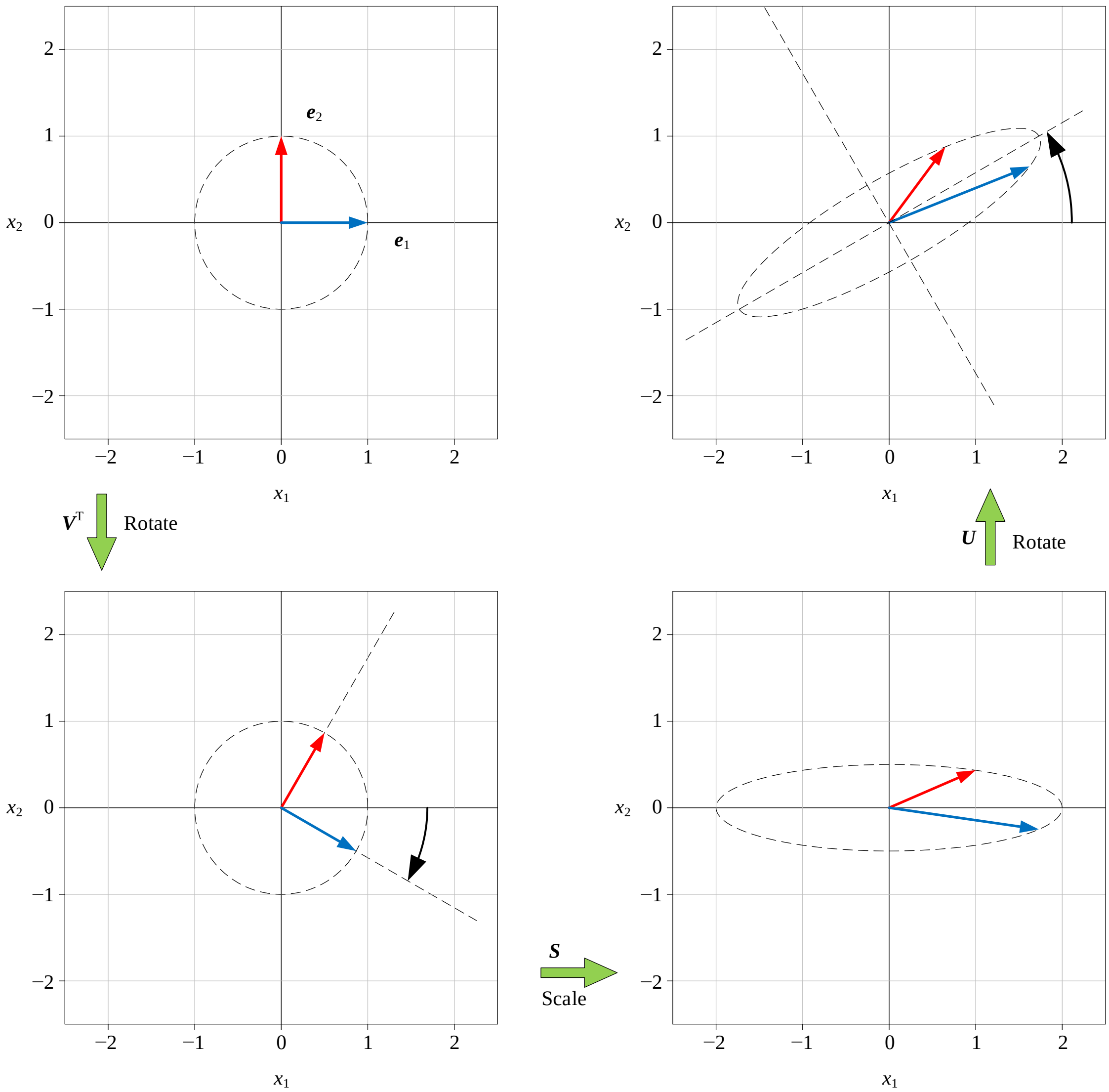


图 3. $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 分别经过 $\boldsymbol{V}^{\mathrm{T}}$、$\boldsymbol{S}$ 和 $\boldsymbol{U}$ 转换

```
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 $\boldsymbol{U}$ 为方阵，形状为 $n \times n$。$\boldsymbol{S}$ 的形状和 $\boldsymbol{X}$ 相同，为 $n \times D$。$\boldsymbol{S}$ 的主对角线元素 s_j 为奇异值，具体形式为：

$$\boldsymbol{S} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \\ & & & \\ & & & \end{bmatrix} \tag{11}$$

约定俗成，这 D 个奇异值的大小关系为 $s_1 \geqslant s_2 \geqslant \dots \geqslant s_D$。

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

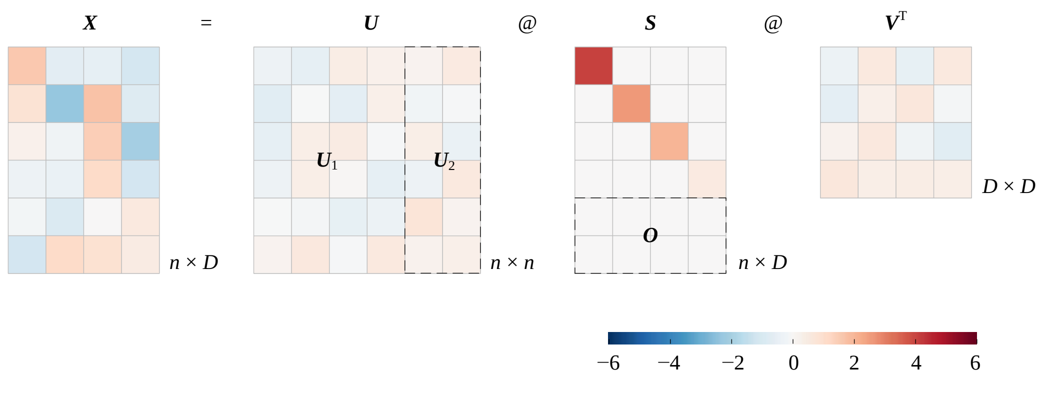


图 4. 矩阵 $\boldsymbol{X}$ 的完全型 SVD 分解

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 $\boldsymbol{X}$ 为细高长方形，对 $\boldsymbol{X}$ 转置便得到宽矮长方形矩阵 $\boldsymbol{X}^{\mathrm{T}}$。如图 5 所示，相应的，$\boldsymbol{X}^{\mathrm{T}}$ 的 SVD 分解为：

$$\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \tag{12}$$

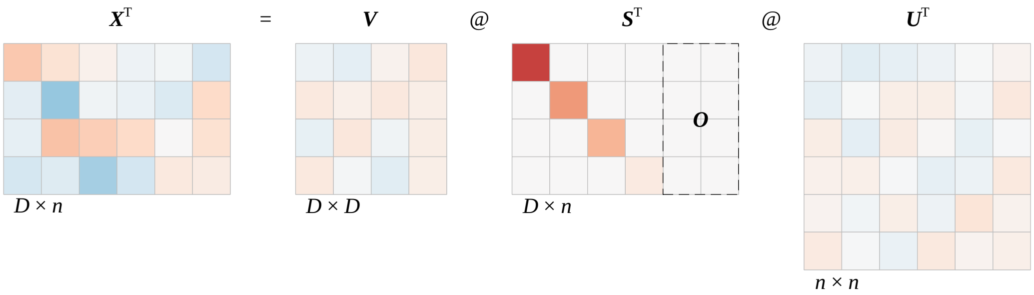


图 5. 矩阵 $\boldsymbol{X}^{\mathrm{T}}$ 的完全型 SVD 分解

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 $\boldsymbol{U}$ 形状和 $\boldsymbol{X}$ 相同，均为为 $n \times D$。而 $\boldsymbol{S}$ 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。

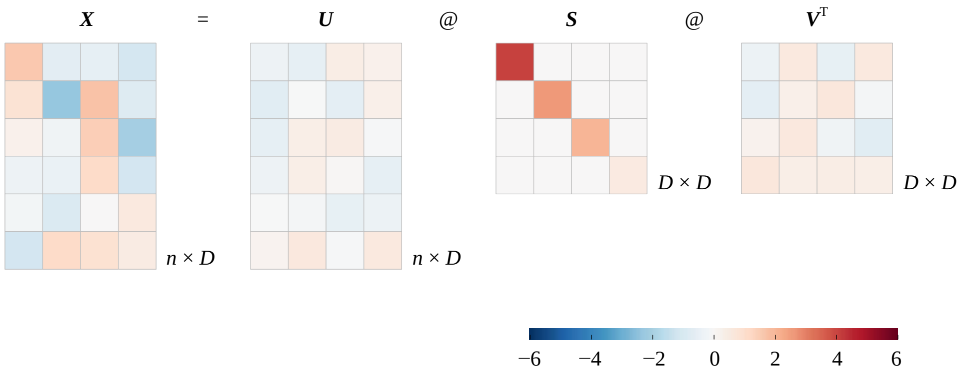


图 6. 经济型 SVD 分解

在经济型 SVD 分解中，$\boldsymbol{S}$ 为对角方阵：

$$\boldsymbol{S} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \tag{13}$$

当 $\boldsymbol{S}$ 为对角方阵时，(12) 可以写成：

$$X^{\mathrm{T}} = VSU^{\mathrm{T}} \tag{14}$$

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 U

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

$$U^{\mathrm{T}}U = I \tag{15}$$

如图 7 所示，对于完全型 SVD 分解，U 为方阵。

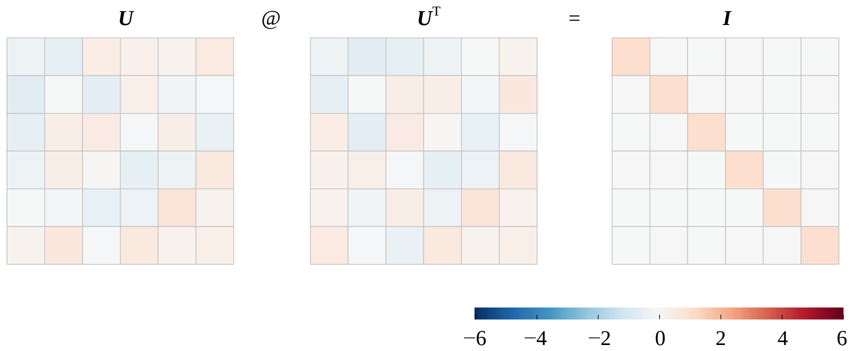


图 7. U 和自己转置 U^{T} 的乘积为单位矩阵

特征值分解

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

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

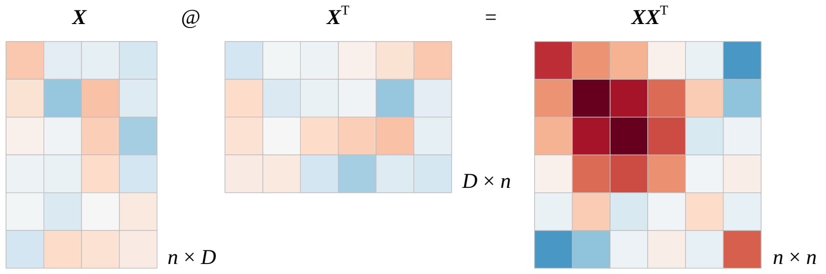


图 8. $\boldsymbol{X}$ 和自己转置 $\boldsymbol{X}^{\mathrm{T}}$ 的乘积热图

对方阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 进行特征值分解，可以发现 $\boldsymbol{U}$ 的列向量是特征向量，而 $\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 是 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的特征值矩阵：

$$
\begin{aligned}
\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} &= \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)\left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}} \\
&= \boldsymbol{U}\boldsymbol{S}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right)\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\
&= \boldsymbol{U}\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}
\end{aligned}
\tag{16}
$$

图 9 所示为 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解热图。

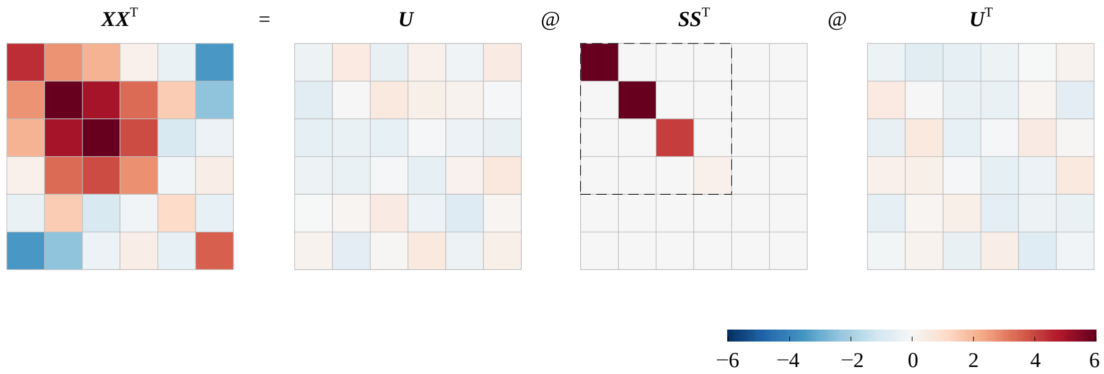


图 9. 对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解

$\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 主对角线为特征值，对 $\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 展开得到：

$$
\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \\ & & & \end{bmatrix} \begin{bmatrix} s_1 & & & & \\ & s_2 & & & \\ & & \ddots & & \\ & & & s_D & \end{bmatrix} = \begin{bmatrix} s_1^2 & & & & \\ & s_2^2 & & & \\ & & \ddots & & \\ & & & s_D^2 & \\ & & & & 0 \\ & & & & & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \lambda_D & & \\ & & & & \ddots & \\ & & & & & \lambda_n \end{bmatrix}
\tag{17}
$$

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

$$
\lambda_j = s_j^2
\tag{18}
$$

剩余的特征值均为 0。

向量空间

如图 10 所示，$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 进行特征值分解得到正交矩阵 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 是个规范正交基，张起的空间为 $\mathbb{R}^n$。

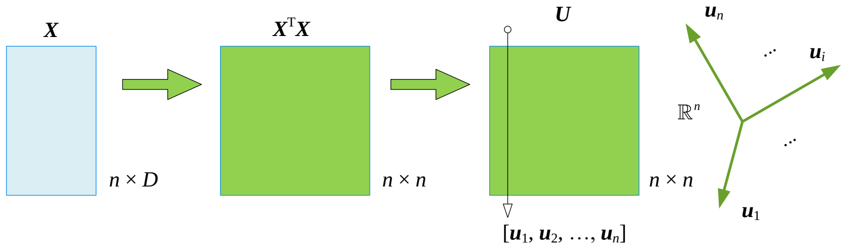


图 10. 对 Gram 矩阵 XX^T 特征值分解得到规范正交基 U

类比 QR 分解

数据矩阵 X 进行 QR 分解得到：

$$X = QR \tag{19}$$

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

对 X 进行完全型 SVD 分解，把结果写成：

$$X = U\left(SV^T\right) \tag{20}$$

对比 (19) 和 (20)，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 $[q_1, q_2, \ldots, q_D]$ 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 (19) 和 (20)，R 则对应 SV^T。特别地，SV^T 结果正交，即 $SV^T(SV^T)^T = SV^TVS^T = SS^T$。

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：

$$V^TV = I \tag{21}$$

图 11 所示为上式运算对应热图。

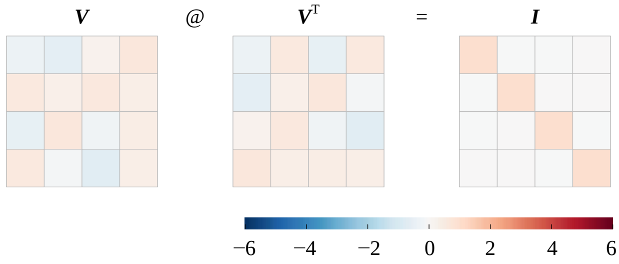


图 11. $\boldsymbol{V}$ 和其转置 $\boldsymbol{V}^{\mathrm{T}}$ 的乘积也是单位矩阵

特征值分解

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

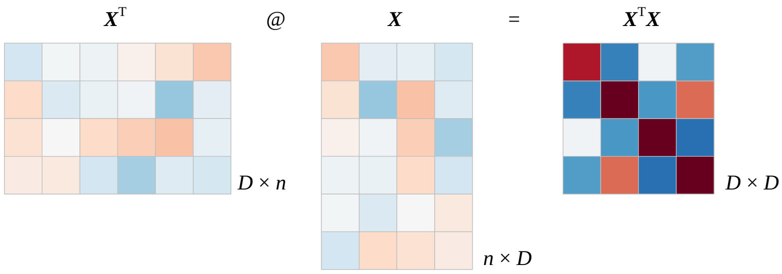


图 12. 转置 $\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{X}$ 乘积热图

对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解得到：

$$
\begin{aligned}
\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} &= \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right) \\
&= \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \\
&= \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}
\end{aligned}
\tag{22}
$$

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

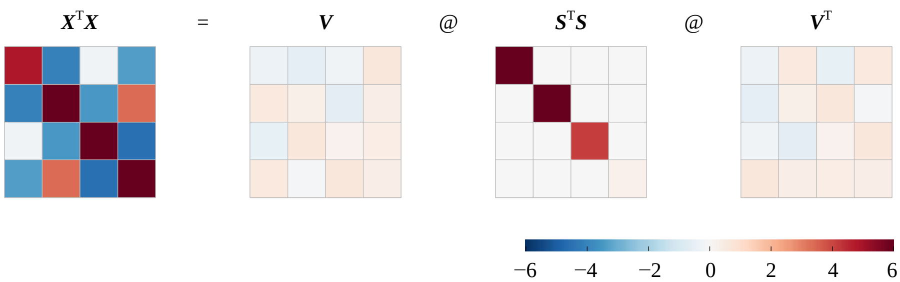


图 13. 对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 进行特征值分解

如图 14 所示，对 X^TX 进行特征值分解，S^TS 为特征值矩阵，奇异值和特征值也存在如下平方关系：

$$S^{\mathrm{T}}S=\begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2 \end{bmatrix}=\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} \tag{23}$$

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。

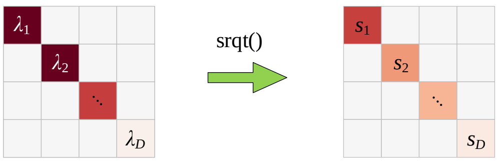


图 14. 奇异值和特征值之间关系

向量空间

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

奇异值分解不但可以分解各种形状实数矩阵，并且一次性获得 $U = [u_1, u_2, \ldots, u_n]$ 和 $V = [v_1, v_2, \ldots, v_D]$ 两个规范正交基。

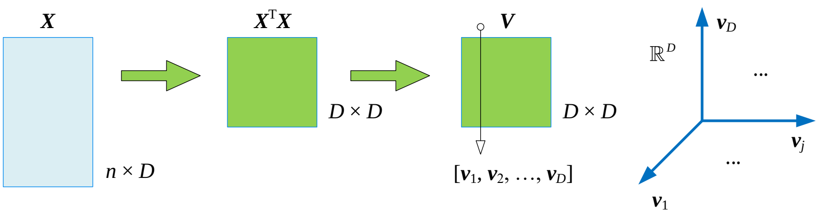


图 15. 对 Gram 矩阵 X^TX 特征值分解得到规范正交基 V

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 $\boldsymbol{V}$，可以得到：

$$\boldsymbol{X}_{n\times D}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} \tag{24}$$

将 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 本身分别写成左右排列的列向量：

$$\boldsymbol{X}_{n\times D}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D\end{bmatrix} \tag{25}$$

(25) 进一步展开得到：

$$\begin{bmatrix}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{X}\boldsymbol{v}_D\end{bmatrix} = \begin{bmatrix}s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D\end{bmatrix} \tag{26}$$

因此，

$$\boldsymbol{X}\boldsymbol{v}_j = s_j\boldsymbol{u}_j \tag{27}$$

上式可以理解为 $\boldsymbol{X}$ 向 $\boldsymbol{v}_j$ 投影，结果为 $s_j\boldsymbol{u}_j$。对应运算热图如图 16 所示。注意，$\boldsymbol{v}_j$ 和 $\boldsymbol{u}_j$ 都是单位向量，即两者的模都是 1。

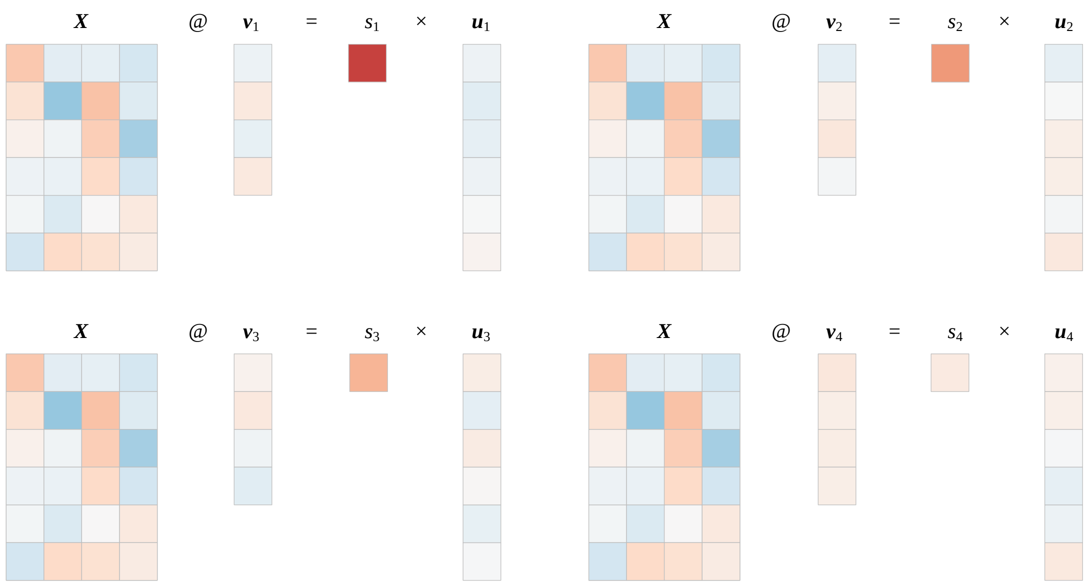


图 16. $\boldsymbol{X}$ 向 $\boldsymbol{v}_j$ 映射结果为 $s_j\boldsymbol{u}_j$

(27) 左右都是向量，等式两侧分别求模，即 L^2 范数，得到：

$$\left\|\boldsymbol{X}\boldsymbol{v}_j\right\| = \left\|s_j\boldsymbol{u}_j\right\| = s_j\underbrace{\left\|\boldsymbol{u}_j\right\|}_{1} = s_j \tag{28}$$

也就是说 $\boldsymbol{X}\boldsymbol{v}_j$ 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 $\boldsymbol{X}\boldsymbol{v}_1$ 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：

$$\begin{aligned}\boldsymbol{X}_{n\times D} &= \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} \\ &= \begin{bmatrix}s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} = s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + s_D\boldsymbol{u}_D\boldsymbol{v}_D^{\mathrm{T}}\end{aligned} \tag{29}$$

举个例子，对于 $D = 4$ 时：

$$\boldsymbol{X} = s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + s_3\boldsymbol{u}_3\boldsymbol{v}_3^{\mathrm{T}} + s_4\boldsymbol{u}_4\boldsymbol{v}_4^{\mathrm{T}} \tag{30}$$

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

⚠ 注意，$s_j\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}$ 的秩为 1。

如图 17 所示，可以发现对应 (30) 等式右侧从左到右的四项相当于逐步还原 $\boldsymbol{X}$。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换换原始数据矩阵。

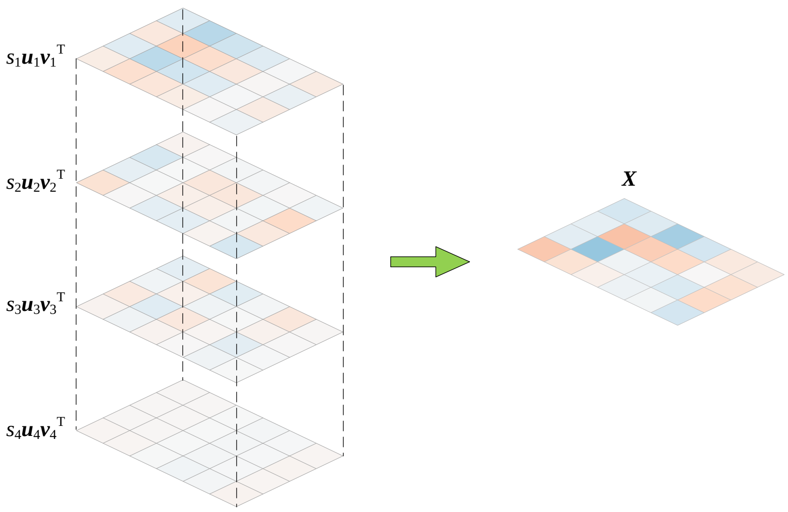


图 17. 四幅热图叠加还原原始图像

张量积

再进一步，利用 (27) 给出的关系，我们将 (30) 写成张量积之和的形式：

$$\begin{aligned}\boldsymbol{X} &= s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + s_3\boldsymbol{u}_3\boldsymbol{v}_3^{\mathrm{T}} + s_4\boldsymbol{u}_4\boldsymbol{v}_4^{\mathrm{T}} \\ &= \boldsymbol{X}\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_4\boldsymbol{v}_4^{\mathrm{T}} \\ &= \boldsymbol{X}\left(\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{T}} + \boldsymbol{v}_4\boldsymbol{v}_4^{\mathrm{T}}\right) \\ &= \boldsymbol{X}\left(\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_3 + \boldsymbol{v}_4 \otimes \boldsymbol{v}_4\right)\end{aligned} \tag{31}$$

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 (31) 投影的规范正交基有无数组，为什么 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 脱颖而出？$\boldsymbol{V}$ 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 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 一般形状不同。

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

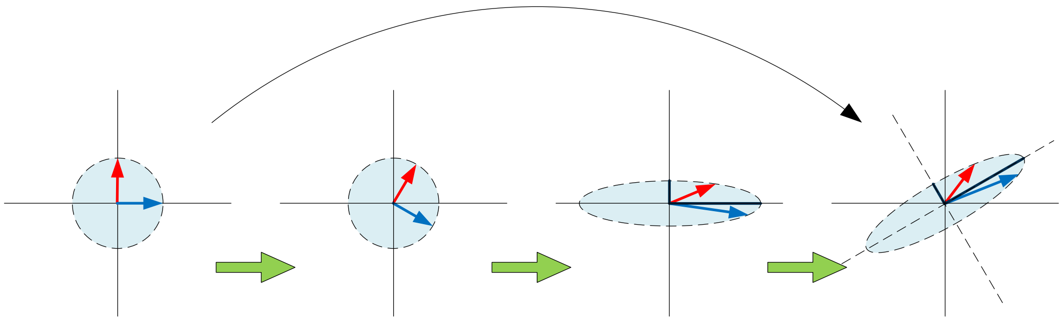


图 18. 总结本章重要内容的四副图

数值线性代数是本书完全没有涉及的板块。

本书有关矩阵分解这个版块介绍了 LU 分解、Cholesky 分解、QR 分解、特征值分解、奇异值分解等等原理和应用，也介绍如何利用 Python 函数完成矩阵分解。但是本书没有提到计算机如何完成这些矩阵分解，也就是 Python 库中这些函数的底层算法实现，这就是数值线性代数研究的问题。

大家如果对这个话题感兴趣的话，可以参考 Holger Wendland 的 *Numerical Linear Algebra: An Introduction*。