16 Dive into Singular Value Decomposition
深入奇异值分解

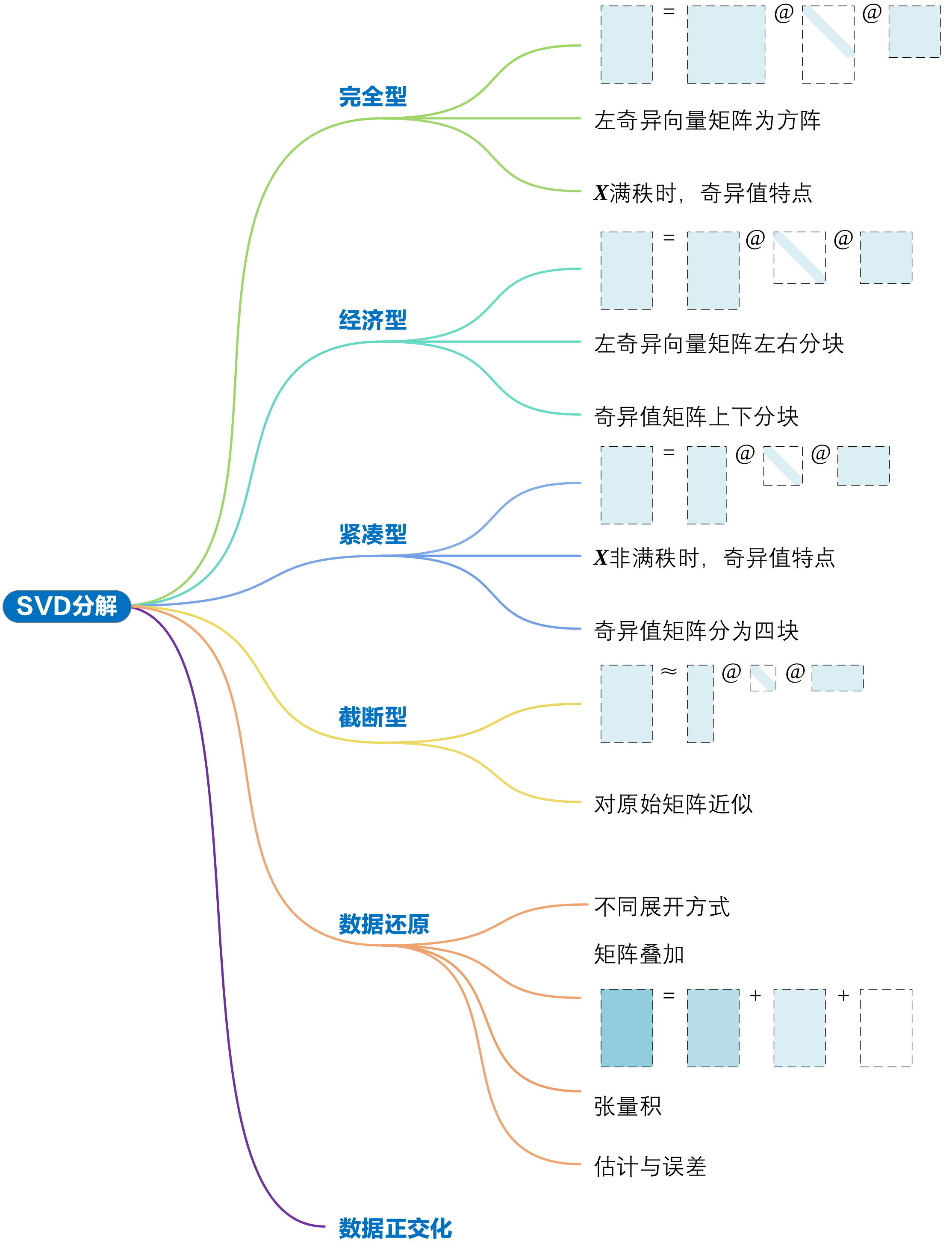
四种类型、数据还原、正交化

人不过是一根芦苇，世界最脆弱的生灵；但是，人是会思考的芦苇。

Man is but a reed, the most feeble thing in nature, but he is a thinking reed.

—— 布莱兹·帕斯卡 (Blaise Pascal) | 法国哲学家、科学家 | 1623 ~ 1662

- `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
完全型
=
@
@
左奇异向量矩阵为方阵
X满秩时，奇异值特点
经济型
=
@
@
左奇异向量矩阵左右分块
奇异值矩阵上下分块
紧凑型
=
@
@
X非满秩时，奇异值特点
奇异值矩阵分为四块
截断型
≈
@
@
对原始矩阵近似
数据还原
不同展开方式
矩阵叠加
=
+
+
张量积
估计与误差
数据正交化

16.1 完全型：*U* 为方阵

上一章介绍过奇异值分解有四种类型：

- 完全型 (full)；
- 经济型 (economy-size, thin)；
- 紧凑型 (compact)；
- 截断型 (truncated)。

本章将深入介绍这四种奇异值分解。

首先回顾完全型 SVD 分解。图 1 所示为矩阵 $\boldsymbol{X}_{6\times4}$ 进行完全 SVD 分解的结果热图。一般情况，丛书常见的数据矩阵 $\boldsymbol{X}$ 形状 $n > D$，即细高型。

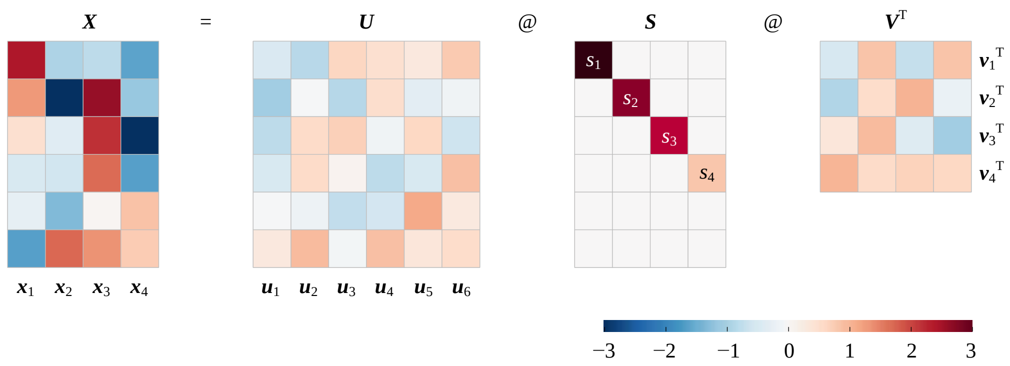


图 1. 数据 $\boldsymbol{X}$ 完全型 SVD 分解矩阵热图

完全型 SVD 分解中，左奇异向量矩阵 $\boldsymbol{U}$ 为方阵，形状为 $n \times n$。$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n]$ 是张成 $\mathbb{R}^n$ 空间的规范正交基。

$\boldsymbol{S}_{n\times D}$ 的形状和 $\boldsymbol{X}$ 相同，为 $n \times D$。虽然 $\boldsymbol{S}_{n\times D}$ 也是对角阵，但是它不是方阵。

如果 $\boldsymbol{X}$ 满秩，$\operatorname{rank}(\boldsymbol{X}) = D$，$\boldsymbol{S}$ 的主对角线元素 (奇异值 s_i) 一般排列顺序为：

$$s_1 \ge s_2 \ge \cdots s_D > 0 \tag{1}$$

右奇异向量矩阵 $\boldsymbol{V}$ 形状为 $D \times D$。$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n]$ 是张成 $\mathbb{R}^D$ 空间的规范正交基。

本章大量使用分块矩阵乘法法则，大家如果感到吃力，请回顾本书第 6 章。

Bk4_Ch16_01.py 中 Bk4_Ch16_01_A 部分绘制图 1。

16.2 经济型：S 去掉零矩阵，变方阵

在完全型 SVD 分解基础上，长方对角阵 $\boldsymbol{S}_{n\times D}$ 上下分块为一个对角方阵和一个零矩阵 $\boldsymbol{O}$:

$$\boldsymbol{S}_{n\times D}=\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}=\begin{bmatrix} \boldsymbol{S}_{D\times D} \\ \boldsymbol{O}_{(n-D)\times D} \end{bmatrix} \tag{2}$$

将 $\boldsymbol{U}_{n\times n}$ 写成左右分块矩阵 $[\boldsymbol{U}_{n\times D}, \boldsymbol{U}_{n\times(n-D)}]$，其中 $\boldsymbol{U}_{n\times D}$ 和 $\boldsymbol{X}$ 形状相同。

利用分块矩阵乘法，完全型 SVD 分解可以简化成经济型 SVD 分解:

$$\begin{aligned}\boldsymbol{X}_{n\times D}&=\begin{bmatrix}\boldsymbol{U}_{n\times D} & \boldsymbol{U}_{n\times(n-D)}\end{bmatrix}\begin{bmatrix} \boldsymbol{S}_{D\times D} \\ \boldsymbol{O}_{(n-D)\times D} \end{bmatrix}\boldsymbol{V}^{\mathrm{T}} \\ &=\left(\boldsymbol{U}_{n\times D}\boldsymbol{S}_{D\times D}+\boldsymbol{U}_{n\times(n-D)}\boldsymbol{O}_{(n-D)\times D}\right)\boldsymbol{V}^{\mathrm{T}} \\ &=\boldsymbol{U}_{n\times D}\boldsymbol{S}_{D\times D}\boldsymbol{V}^{\mathrm{T}}\end{aligned} \tag{3}$$

图 2 和图 3 比较完全型和经济型 SVD 分解结果热图。图 2 中阴影部分为消去的矩阵子块。比较完全型和经济型 SVD，分解结果中唯一不变的就是矩阵 $\boldsymbol{V}$，它一直保持方阵形态。

从向量空间角度来讲，$\boldsymbol{U}_{n\times D}$ 和 $\boldsymbol{U}_{n\times(n-D)}$ 有怎样的差异和联系？这是本书第 23 章要回答的问题。

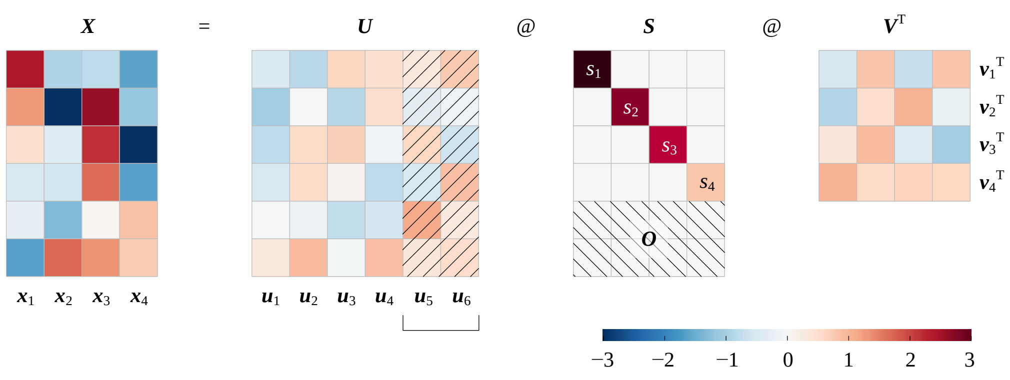


图 2. 数据 $\boldsymbol{X}$ 完全型 SVD 分解分块热图

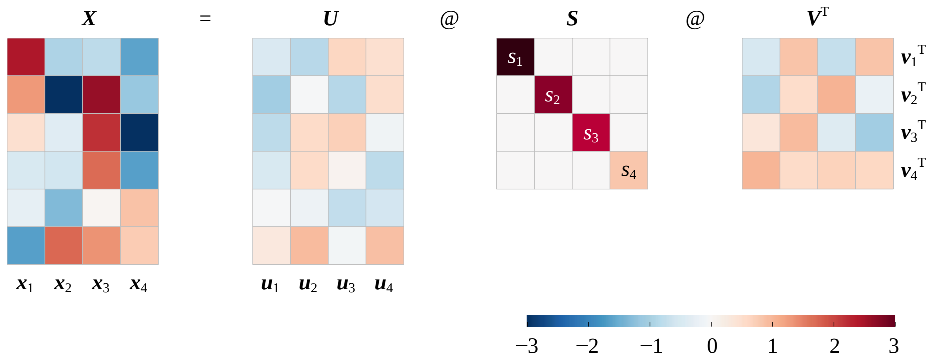


图 3. 数据 $\boldsymbol{X}$ 经济型 SVD 分解热图

Bk4_Ch16_01.py 中 Bk4_Ch16_01_B 部分绘制图 3。

16.3 紧凑型：非满秩

特别地，如果 $\mathrm{rank}(\boldsymbol{X}) = r < D$，奇异值 s_j 满足：

$$s_1 \geq s_2 \geq \cdots s_r > 0, \quad s_{r+1} = s_{r+2} = \cdots s_D = 0 \tag{4}$$

这种条件下，奇异值矩阵 $\boldsymbol{S}$ 可以分成四个子块：

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{r\times r} & \boldsymbol{O}_{r\times(D-r)} \\ \boldsymbol{O}_{(D-r)\times r} & \boldsymbol{O}_{(D-r)\times(D-r)} \end{bmatrix} \tag{5}$$

上式中，矩阵 $\boldsymbol{S}_{r\times r}$ 对角线元素奇异值均大于 0。

将 (5) 代入完全型 SVD 分解 (3)，整理得到：

$$\begin{aligned} \boldsymbol{X}_{n\times D} &= \begin{bmatrix} \boldsymbol{U}_{n\times r} & \boldsymbol{U}_{n\times(D-r)} \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_{r\times r} & \boldsymbol{O}_{r\times(D-r)} \\ \boldsymbol{O}_{(D-r)\times r} & \boldsymbol{O}_{(D-r)\times(D-r)} \end{bmatrix} \begin{bmatrix} \left(\boldsymbol{V}_{D\times r}\right)^{\mathrm{T}} \\ \left(\boldsymbol{V}_{D\times(D-r)}\right)^{\mathrm{T}} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{U}_{n\times r}\boldsymbol{S}_{r\times r} & \boldsymbol{O}_{n\times(D-r)} \end{bmatrix} \begin{bmatrix} \left(\boldsymbol{V}_{D\times r}\right)^{\mathrm{T}} \\ \left(\boldsymbol{V}_{D\times(D-r)}\right)^{\mathrm{T}} \end{bmatrix} \\ &= \boldsymbol{U}_{n\times r}\boldsymbol{S}_{r\times r}\left(\boldsymbol{V}_{D\times r}\right)^{\mathrm{T}} \end{aligned} \tag{6}$$

大家特别注意 (6) 中，矩阵 $\boldsymbol{V}$ 先分块后再转置。

图 4 和图 5 比较经济型和紧凑型 SVD 分解，图 4 阴影部分为消去子块。为了展示紧凑型 SVD 分解，我们用 $\boldsymbol{X}$ 第一、二列数据之和替代 $\boldsymbol{X}$ 矩阵第四列，即 $\boldsymbol{x}_4 = \boldsymbol{x}_1 + \boldsymbol{x}_2$。这样 $\boldsymbol{X}$ 矩阵列向量线性相关，$\text{rank}(\boldsymbol{X}) = 3$，而 $s_4 = 0$。再次强调，只有 $\boldsymbol{X}$ 为非满秩情况下，才存在紧缩型 SVD 分解。紧缩型 SVD 分解中，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都不是方阵。

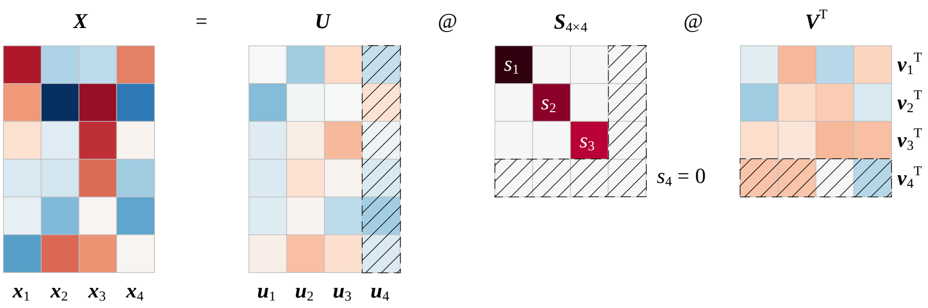


图 4. 数据 $\boldsymbol{X}$ 经济型 SVD 分解热图

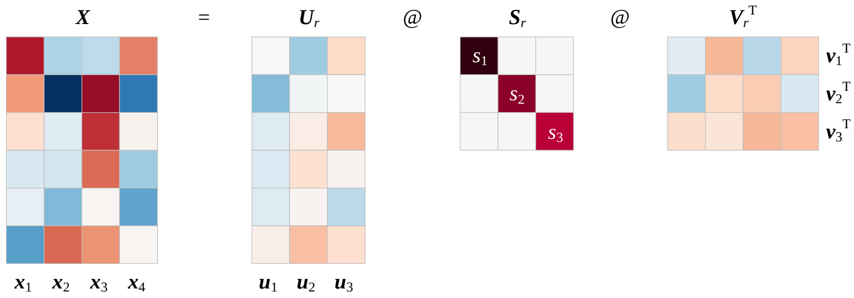


图 5. 数据 $\boldsymbol{X}$ 紧凑型 SVD 分解热图

Bk4_Ch16_01.py 中 Bk4_Ch16_01_C 部分绘制图 4。

16.4 截断型：近似

如果 $\text{rank}(\boldsymbol{X}) = r \leqslant D$，取紧缩型奇异值分解中前 p 个奇异值 ($p < r$) 对应的 $\boldsymbol{U}$、$\boldsymbol{S}$、$\boldsymbol{V}$ 矩阵成分，用它们还原原始数据就是截断型奇异值分解：

$$\boldsymbol{X}_{n\times D} \approx \hat{\boldsymbol{X}}_{n\times D} = \boldsymbol{U}_{n\times p}\boldsymbol{S}_{p\times p}\left(\boldsymbol{V}_{D\times p}\right)^{\mathrm{T}} \tag{7}$$

请大家自行补足上式中矩阵分块和对应的乘法运算。

(7) 不是等号，也就是截断型奇异值分解不能完全还原原始数据。换句话，截断型奇异值分解是对原矩阵 $\boldsymbol{X}$ 的一种近似。图 6 所示为 SVD 截断型分解热图，可以发现 $\boldsymbol{X}_{n\times D}$ 和 $\hat{\boldsymbol{X}}_{n\times D}$ 两幅热图存在一定误差。

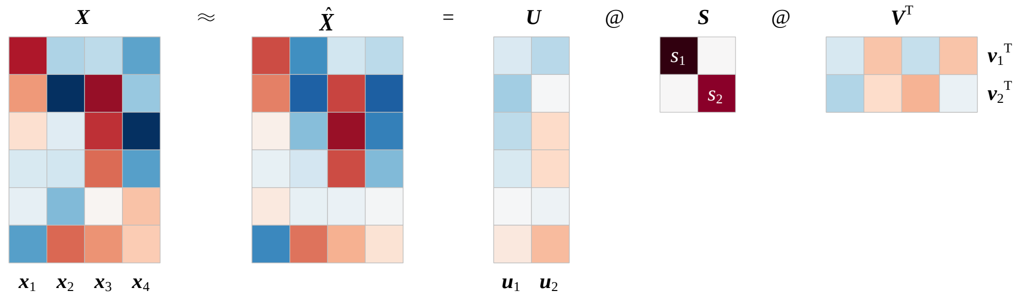


图 6. 采用截断型 SVD 分解还原数据运算热图

Bk4_Ch16_01.py 中 Bk4_Ch16_01_D 绘制图 6。

16.5 数据还原：层层叠加

上一章介绍过，经济型 SVD 分解可以展开写作：

$$
\begin{aligned}
\boldsymbol{X}_{n\times D} &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}} \end{bmatrix} \\
&= \underbrace{s_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}}_{\boldsymbol{X}_1} + \underbrace{s_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}}_{\boldsymbol{X}_2} + \cdots + \underbrace{s_D \boldsymbol{u}_D \boldsymbol{v}_D^{\mathrm{T}}}_{\boldsymbol{X}_D}
\end{aligned}
\tag{8}
$$

上式中奇异值从大到小排列，即 $s_1 \geq s_2 \geq \ldots s_D$。图 7 所示上述运算热图，$D = 4$。

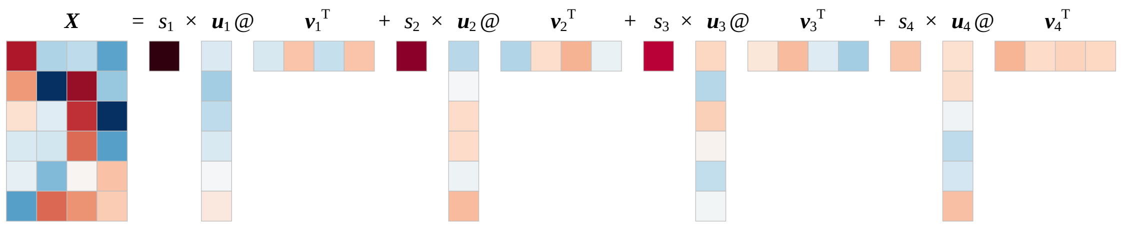


图 7. SVD 分解展开计算热图

组成部分

定义矩阵 $\boldsymbol{X}_j$ 为：

$$\boldsymbol{X}_j = s_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathrm{T}} \tag{9}$$

矩阵 $\boldsymbol{X}_j$ 形状和 $\boldsymbol{X}$ 相同。图 8 所示为矩阵 $\boldsymbol{X}_j$ (j = 1, 2, 3, 4) 计算过程热图。

观察图 8 每幅矩阵 $\boldsymbol{X}_j$ 热图不难发现，矩阵 $\boldsymbol{X}_j$ 自身列向量之间存在倍数关系。也就是说，矩阵 $\boldsymbol{X}_j$ 的秩为 1，即 $\mathrm{rank}(\boldsymbol{X}_j) = 1$。

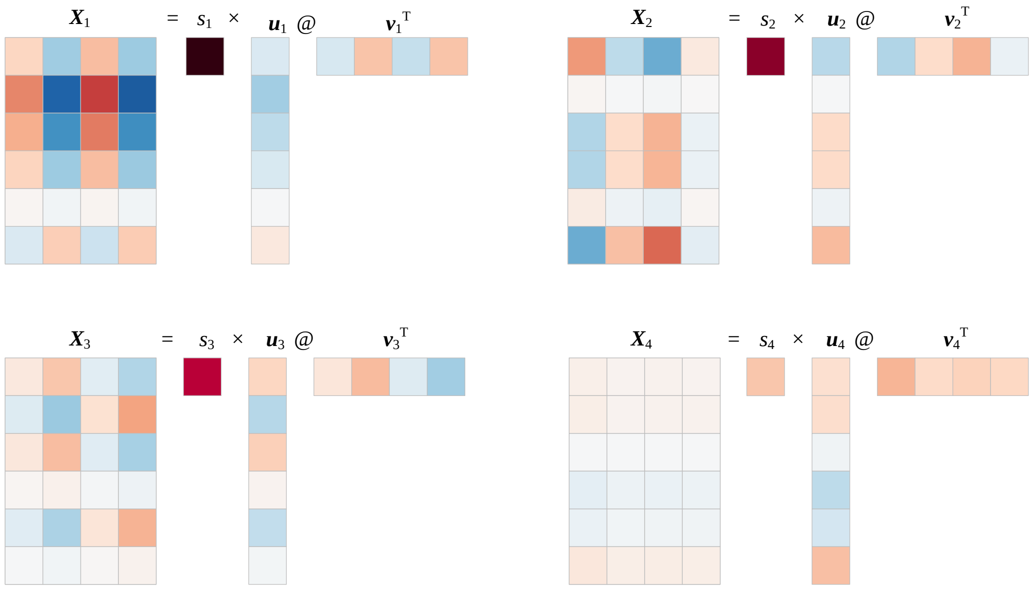


图 8. 还原数据的叠加成分

还原

(9) 代入 (8) 得到：

$$\boldsymbol{X}_{n \times D} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_D \tag{10}$$

当 j = 1 ~ D 时，将 $\boldsymbol{X}_j$ 一层层叠加最后还原原始数据矩阵 $\boldsymbol{X}$，如图 9 所示。

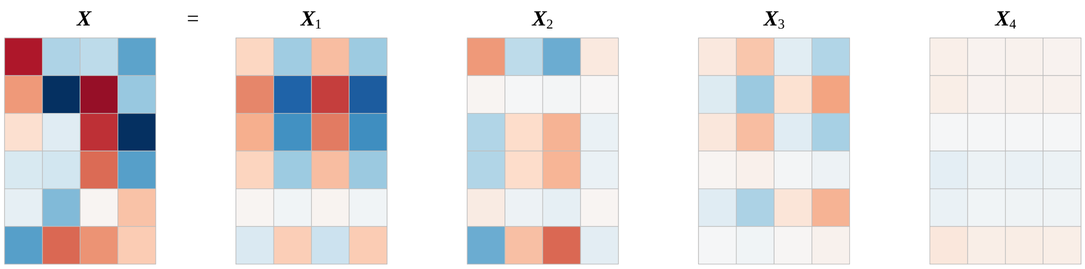


图 9. 还原原始数据

张量积

利用向量张量积，(8) 可以写成：

$$\boldsymbol{X} = \underbrace{s_1 \boldsymbol{u}_1 \otimes \boldsymbol{v}_1}_{\hat{\boldsymbol{X}}_1} + \underbrace{s_2 \boldsymbol{u}_2 \otimes \boldsymbol{v}_2}_{\hat{\boldsymbol{X}}_2} + \cdots + \underbrace{s_D \boldsymbol{u}_D \otimes \boldsymbol{v}_D}_{\hat{\boldsymbol{X}}_D} = \sum_{j=1}^{D} s_j \boldsymbol{u}_j \otimes \boldsymbol{v}_j \tag{11}$$

图 10 所示为张量积 $\boldsymbol{u}_j \otimes \boldsymbol{v}_j$ 计算热图，可以发现热图色差并不明显。这说明 $\boldsymbol{u}_j \otimes \boldsymbol{v}_j$ 本身并不能区分 $\boldsymbol{X}_j$，这是因为 $\boldsymbol{u}_j$ 和 $\boldsymbol{v}_j$ 都是单位向量。

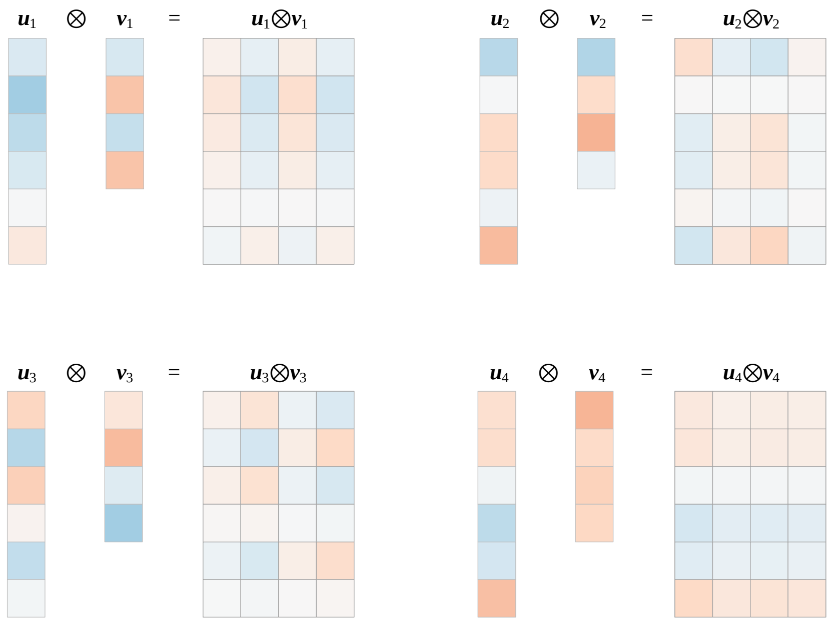


图 10. 向量张量热图

然后再用奇异值 s_j 乘以对应张量积 $\boldsymbol{u}_j \otimes \boldsymbol{v}_j$ 得到 $\boldsymbol{X}_j$，具体如图 11 所示。可以发现 $\boldsymbol{X}_1$ 热图色差最明显。也就是说，奇异值 s_j 的大小决定了成分的重要性，而 $\boldsymbol{u}_j$ 和 $\boldsymbol{v}_j$ 决定了投影方向。

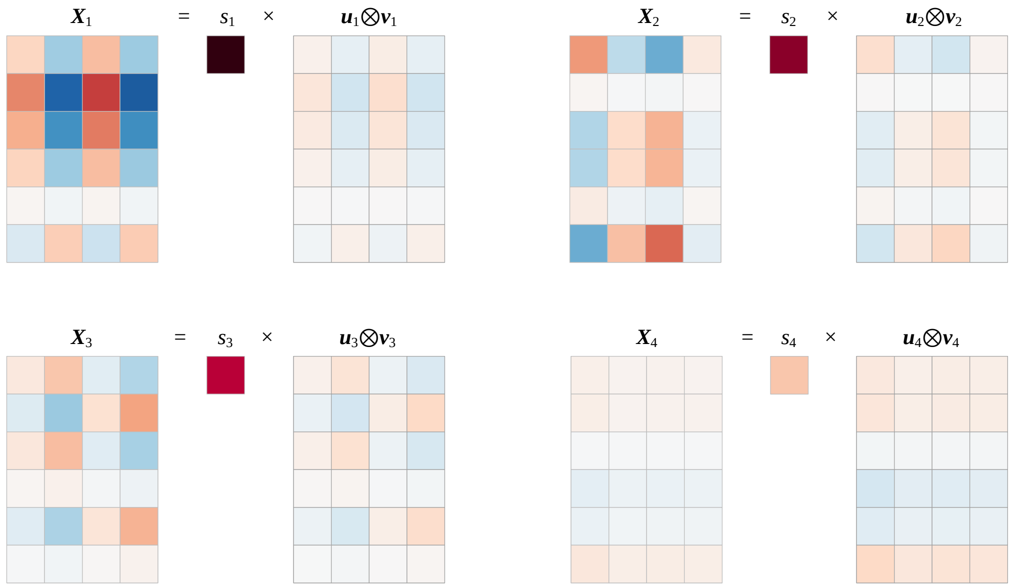


图 11. 奇异值标量乘张量积结果

正交投影

上一章指出 $\boldsymbol{v}_j$ 和 $\boldsymbol{u}_j$ 存在如下关系：

$$\boldsymbol{X}\boldsymbol{v}_j = s_j \boldsymbol{u}_j \tag{12}$$

将 (12) 代入 (11)，就得到：

$$\begin{aligned}\boldsymbol{X} &= \underbrace{\boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1}_{\hat{\boldsymbol{x}}_1} + \underbrace{\boldsymbol{X}\boldsymbol{v}_2 \otimes \boldsymbol{v}_2}_{\hat{\boldsymbol{x}}_2} + \cdots + \underbrace{\boldsymbol{X}\boldsymbol{v}_D \otimes \boldsymbol{v}_D}_{\hat{\boldsymbol{x}}_D} \\ &= \boldsymbol{X}\left(\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_D \otimes \boldsymbol{v}_D\right)\end{aligned} \tag{13}$$

这就是本书第 9、10 章反复提到的“二次投影 + 层层叠加”。以 $\boldsymbol{v}_1$ 为例，数据 $\boldsymbol{X}$ 在 span($\boldsymbol{v}_1$) 中投影在 $\mathbb{R}^D$ 中的像就是 $\boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$。span($\boldsymbol{v}_1$) 是 $\mathbb{R}^D$ 的子空间，维度为 1。这就意味着 $\boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$ 的秩为 1，即 $\operatorname{rank}\left(\boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1\right) = 1$。

之所以选择 $\boldsymbol{v}_1$ 做第一投影方向，就是因为在所有的一维方向中，$\boldsymbol{v}_1$ 方向对应的奇异值 s_1 最大。大家可能又会好奇，几何视角下，奇异值 s_1 到底是什么？卖个关子，这个问题在本书第 18 章回答。

Bk4_Ch16_01.py 中 Bk4_Ch16_01_E 计算张量积并绘制热图。

16.6 估计与误差：截断型 SVD

把数据矩阵 $\boldsymbol{X}$ 对应的热图看做一幅图像，本节介绍如何采用较少数据尽可能还原原始图像，并准确知道误差是多少。

两层叠加

奇异值按大小排列，选取 s_1 和 s_2 还原原始数据，其中 s_1 最大，s_2 其次。

根据上一节讨论，从图像还原角度，s_1 对应 $\boldsymbol{X}_1$，$\boldsymbol{X}_1$ 还原了 $\boldsymbol{X}$ 图像大部分特征；s_2 对应 $\boldsymbol{X}_2$，$\boldsymbol{X}_2$ 在 $\boldsymbol{X}_1$ 基础上进一步还原 $\boldsymbol{X}$。

$\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 叠加得到 $\hat{\boldsymbol{X}}$。如图 12 所示，$\boldsymbol{X}$ 和 $\hat{\boldsymbol{X}}$ 热图的相似度已经很高：

$$\boldsymbol{X}_{n\times D} \approx \hat{\boldsymbol{X}}_{n\times D} = \boldsymbol{X}_1 + \boldsymbol{X}_2 \tag{14}$$

$\boldsymbol{X}$ 和 $\hat{\boldsymbol{X}}$ 热图误差矩阵为：

$$\boldsymbol{E}_\varepsilon = \boldsymbol{X}_{n\times D} - \hat{\boldsymbol{X}}_{n\times D} \tag{15}$$

我们给 $\boldsymbol{E}_\varepsilon$ 加了个下角标，以便和标准正交基 $\boldsymbol{E}$ 区分。

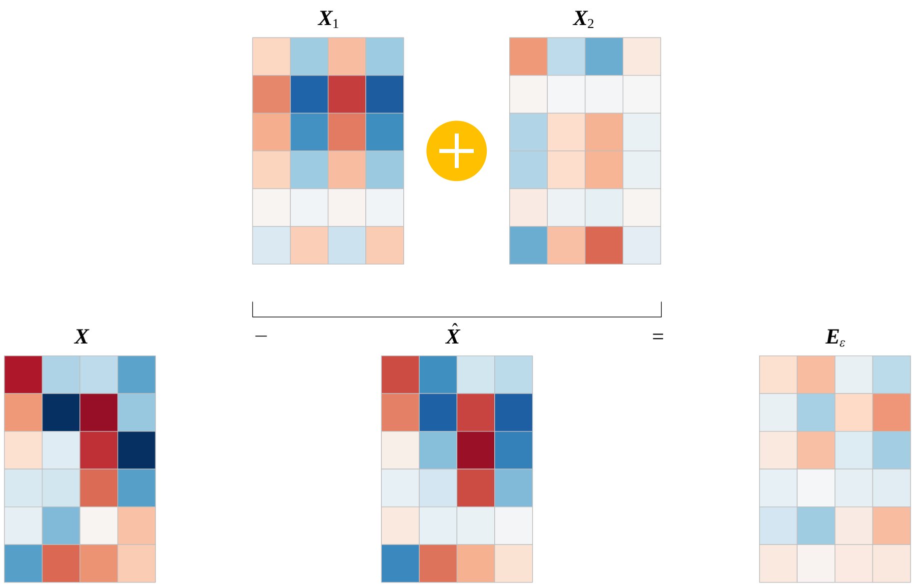


图 12. 利用前两个奇异值对应的矩阵还原原数据

将 (14) 展开写成：

$$\boldsymbol{X} \approx \hat{\boldsymbol{X}} = s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2\end{bmatrix}\begin{bmatrix} s_1 & \\ & s_2\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}}\end{bmatrix} \tag{16}$$

上式实际上就是主成分分析中，用前两个主元还原原始数据对应的计算，具体热图如图 13 所示。

本系列丛书《概率统计》一册将从中心化数据、z 分数、协方差矩阵和相关性系数矩阵等角度讲解主成分分析的不同技术途径，而《数据科学》一册将从数据应用角度再谈主成分分析。

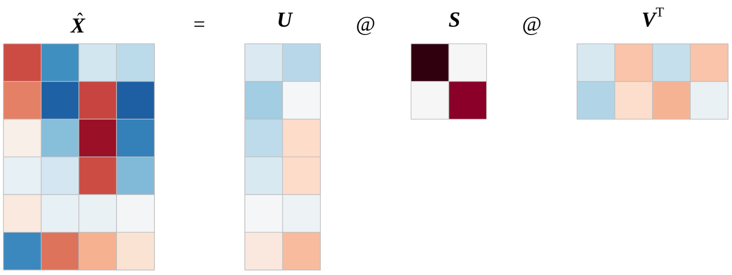


图 13. 用前两个主元还原原始数据

采用 s_1 和 s_2 还原原始数据时，误差 E_ε 具体为：

$$
\begin{aligned}
X - \hat{X} &= s_3 u_3 v_3^{\mathrm{T}} + s_4 u_4 v_4^{\mathrm{T}} + \cdots + s_D u_D v_D^{\mathrm{T}} \\
&= X\left(v_3 \otimes v_3 + v_4 \otimes v_4 + \cdots + v_D \otimes v_D\right) \\
&= X\sum_{j=3}^{D} v_j \otimes v_j
\end{aligned}
\tag{17}
$$

三层叠加

图 14 所示为利用前三个奇异值对应矩阵还原数据，可以发现 X 和 $\hat{X}$ 热图误差进一步缩小。

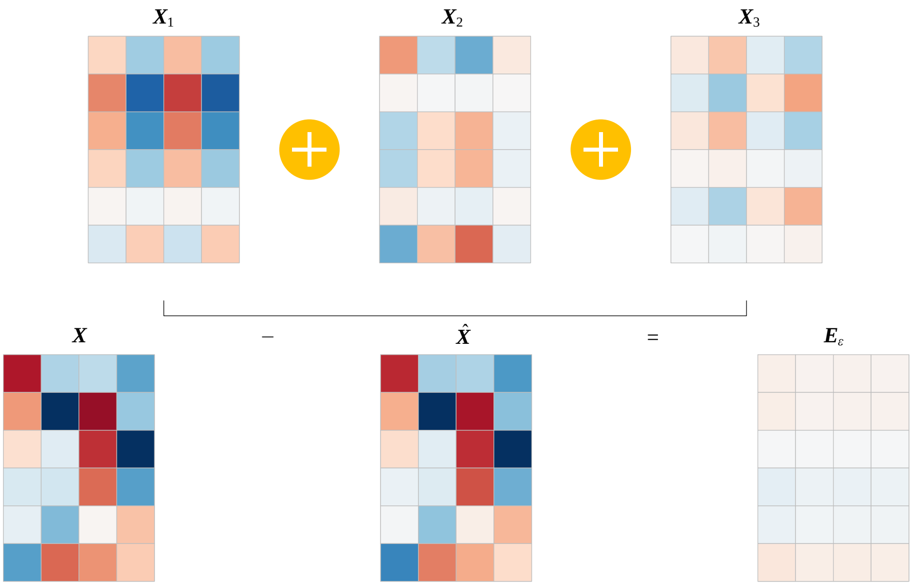


图 14. 利用前三个奇异值对应的矩阵还原数据

当 $D = 4$ 时，采用 s_1、s_2、s_3 还原原始数据时，误差 $\boldsymbol{E}_\varepsilon$ 只剩一个成分：

$$\boldsymbol{X} - \hat{\boldsymbol{X}} = s_4 \boldsymbol{u}_4 \boldsymbol{v}_4^{\mathrm{T}} = \boldsymbol{X}\boldsymbol{v}_4 \otimes \boldsymbol{v}_4 \tag{18}$$

如果采用全部成分还原原始数据，请大家自行计算误差矩阵是否为 $\boldsymbol{O}$ 矩阵。

Bk4_Ch16_01.py 中 Bk4_Ch16_01_F 绘制本节数据还原和误差热图。

鸢尾花照片

我们在本书第 1 章见过图 15 (a) 这幅鸢尾花照片，这张黑白照片本身就是数据矩阵。对这个数据矩阵进行奇异值分解，并依照本节介绍的数据还原方法用不同主成分 (Principal Component, PC) 还原原始图片。

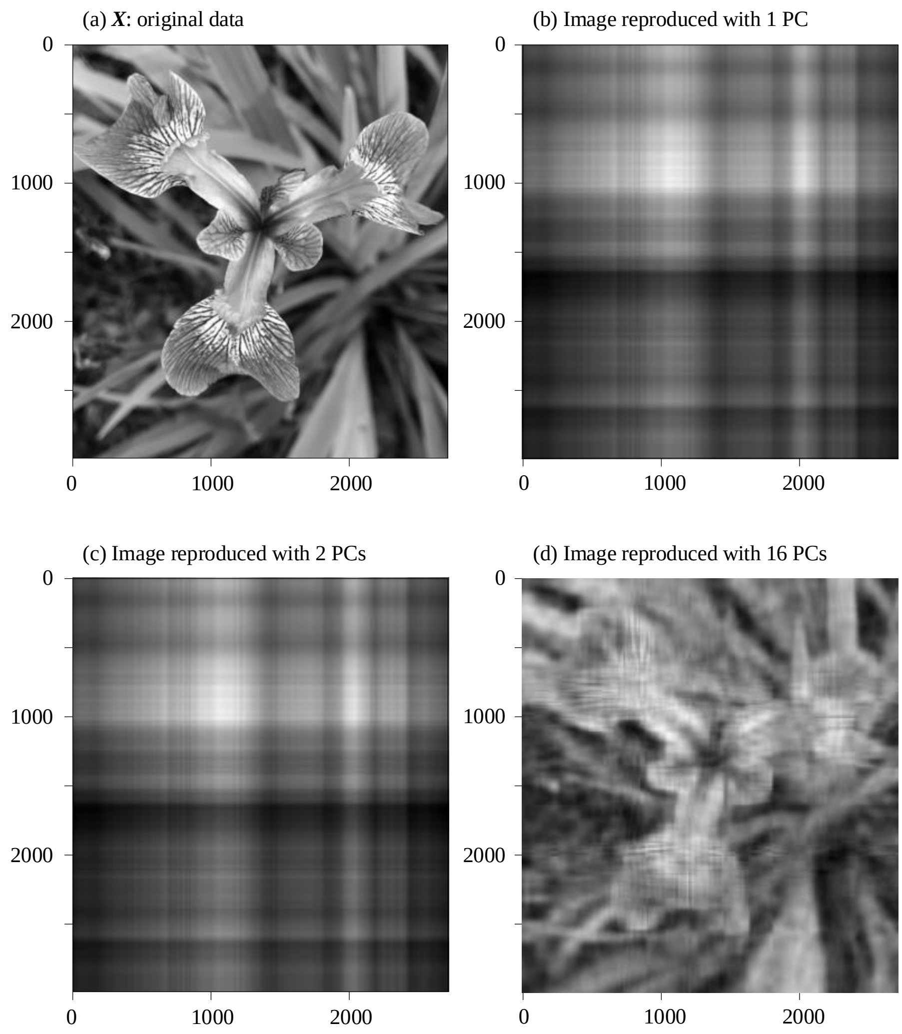


图 15. 还原原始图片

这个主成分对应的投影方向就是本节规范正交基向量 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3$ 等等。图 15 (b) 和 (c) 所示为分别采用一个和两个主成分还原原始图片，我们还很难从图片中看到鸢尾花的踪影。从向量空间角度来说，图 15 (b) 图片的数据的秩为 1，维度也是 1；图 15 (c) 图片的数据的秩为 2，维度也是 2。图 15 (d) 则是采用前 16 个主成分还原原始图片，图片本身已经明显看到鸢尾花原始照片的特征，而这幅图片的数据量却小于原图像的 1%。

本系列丛书《数据科学》还会采用图 15 这个例子深入探讨主成分分析。

16.7 正交投影：数据正交化

本书之前第 10 章介绍过，下式相当于数据矩阵 $\boldsymbol{X}$ 向规范正交基 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_D]$ 构成的 D 维空间投影：

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V} \tag{19}$$

乘积结果 $\boldsymbol{Z}$ 代表 $\boldsymbol{X}$ 在新的规范正交基 $[\boldsymbol{v}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_D]$ 下的坐标。本章介绍的 SVD 分解恰好帮我们找到了一个规范正交基 $\boldsymbol{V}$。本节聊聊投影结果 $\boldsymbol{Z}$ 的性质。

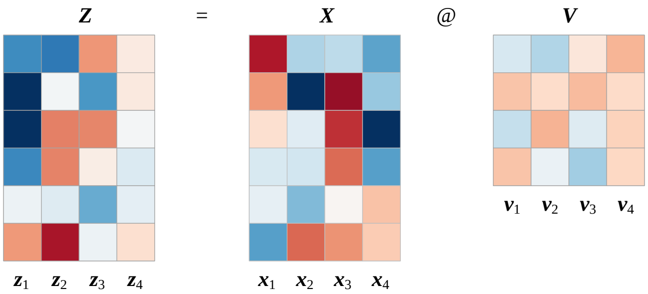


图 16. $\boldsymbol{X}$ 向规范正交基 $\boldsymbol{V}$ 投影

由于 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}$，代入 (19) 得到：

$$\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} = \begin{bmatrix} s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D \end{bmatrix} \tag{20}$$

即，

$$\underbrace{\begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D \end{bmatrix}}_{\boldsymbol{Z}} = \underbrace{\begin{bmatrix} s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D \end{bmatrix}}_{\boldsymbol{US}} \tag{21}$$

如图 17 所示，上式给了我们计算 $\boldsymbol{Z}$ 的第二条路径。换句话说，$\boldsymbol{u}_j$ 实际上就是“单位化”的投影坐标，s_j 是 $\boldsymbol{z}_j$ 向量的模，即 $\|\boldsymbol{X}\boldsymbol{v}_j\| = \|\boldsymbol{z}_j\| = s_j$。

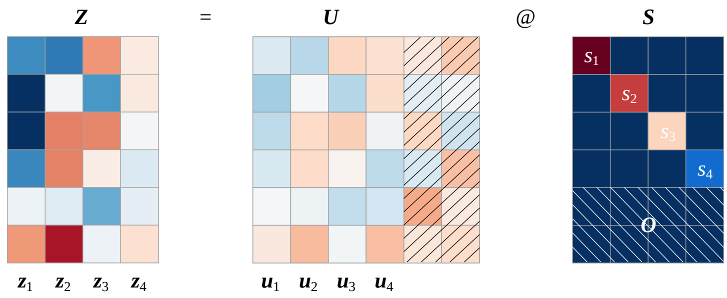


图 17. 第二条计算 $\boldsymbol{Z}$ 的路径

格拉姆矩阵

对 $\boldsymbol{Z}$ 求格拉姆矩阵：

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1^{\mathrm{T}} \\ \boldsymbol{z}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{z}_D^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 & \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_2 & \cdots & \boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_D \\ \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_1 & \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_2 & \cdots & \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{z}_D^{\mathrm{T}}\boldsymbol{z}_1 & \boldsymbol{z}_D^{\mathrm{T}}\boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D^{\mathrm{T}}\boldsymbol{z}_D \end{bmatrix} \tag{22}$$

请大家将上式写成向量内积形式。

将 (20) 代入得到 (22)：

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} = (\boldsymbol{U}\boldsymbol{S})^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S} = \boldsymbol{S}^{\mathrm{T}}\underbrace{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}}_{\boldsymbol{I}}\boldsymbol{S} = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2 \end{bmatrix} \tag{23}$$

如图 18 所示，发现 $\boldsymbol{Z}$ 的格拉姆矩阵为对角阵，也就是说 $\boldsymbol{Z}$ 的列向量两两正交，即：

$$\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{z}_j = \boldsymbol{z}_j^{\mathrm{T}}\boldsymbol{z}_i = \boldsymbol{z}_i \cdot \boldsymbol{z}_j = \boldsymbol{z}_j \cdot \boldsymbol{z}_i = \langle \boldsymbol{z}_i, \boldsymbol{z}_j \rangle = \langle \boldsymbol{z}_j, \boldsymbol{z}_i \rangle = 0, \quad i \neq j \tag{24}$$

回看图 16，$\boldsymbol{X} \to \boldsymbol{Z}$ 的过程就是正交化 (orthogonalization)。

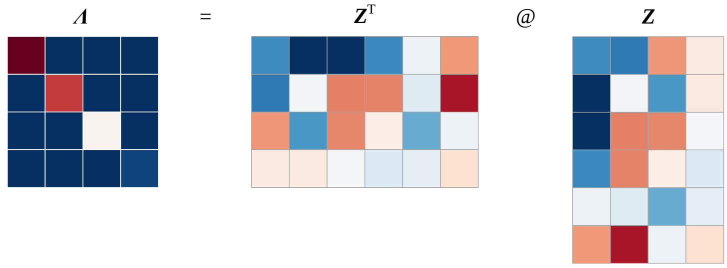


图 18. **Z** 的格拉姆矩阵

如下四幅图最能概括本章的核心内容。奇异值分解的四种不同类型都有特殊意义，都有不同应用场合。

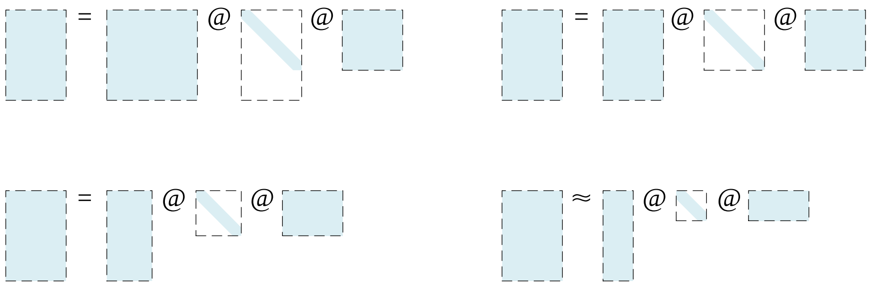


图 19. 总结本章重要内容的四副图

再次强调，矩阵分解的内核还是矩阵乘法。相信大家已经在本章奇异值分解中看到矩阵乘法的不同视角、分块矩阵乘法等数学工具。

此外，张量积和正交投影这两个工具在解释奇异值分解上有立竿见影的效果。

本章留了个悬念，奇异值分解中的奇异值的几何内涵到底是什么？我们将在本书第 18 章回答这个问题。在那里，大家会用优化视角一睹奇异值的几何本质。

本章虽然是矩阵分解板块的最后一章，但是本书有关矩阵分解的故事远没有结束。本书后续会从优化角度、数据角度、空间角度、应用角度一次次回顾这些线性代数的有力武器。