第15章 奇异值分解

1 奇异值分解的定义与性质

矩阵特征值与特征向量

矩阵代表线性变换。特征向量为变化的属性之一

线性动力学系统：$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k(\boldsymbol{k} = \mathbf{0}，\mathbf{1}，\mathbf{2}，\cdots)$，则$\boldsymbol{x}_k = \boldsymbol{A}^n\boldsymbol{x}_0$

矩阵特征值与特征向量

- 特征值与特征向量$\boldsymbol{A}$为$\boldsymbol{n}\times\boldsymbol{n}$矩阵，如果数$\boldsymbol{\lambda}$和$\boldsymbol{n}$维非零向量$\boldsymbol{x}$使

$$\boldsymbol{Ax}=\boldsymbol{\lambda x}$$

- 则称 $\boldsymbol{\lambda}$ 是特征值， $\boldsymbol{x}$ 是对应于 $\boldsymbol{\lambda}$ 的特征向量

- 1) 矩阵与向量相乘$\boldsymbol{Ax}$是一种线性变换，把一个向量，变换成另一个向量
- 2) 矩阵$\boldsymbol{A}$的特征向量经过线性变换$\boldsymbol{A}$后方向与原方向保持一致

- 【重要应用：线性动力学系统】
  - 线性动力学系统中$\boldsymbol{x}_{\boldsymbol{k}+\boldsymbol{1}}=\boldsymbol{Ax}_{\boldsymbol{k}}(\boldsymbol{k}=\boldsymbol{0}，\boldsymbol{1}，\boldsymbol{2}，\cdots)$，则$\boldsymbol{x}_{\boldsymbol{k}}=\boldsymbol{A}^{\boldsymbol{n}}\boldsymbol{x}_{\boldsymbol{0}}$，计算$\boldsymbol{A}^{\boldsymbol{n}}$

对角矩阵

- 对角矩阵(diagonal matrix)，主对角线之外的元素皆为0
  - 记作$\mathbf{M} = \mathbf{diag}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \cdots, \boldsymbol{\lambda}_\mathbf{n}) = \begin{pmatrix} \boldsymbol{\lambda}_1 & & & \\ & \boldsymbol{\lambda}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\lambda}_\mathbf{n} \end{pmatrix}$
- 【用途】
  - $\mathbf{M}^k$比较容易计算
  - $\mathbf{M}$代表线性变换有明确的几何意义【缩放】

对角化矩阵

- $\mathbf{A}$为$\mathbf{n}\times\mathbf{n}$矩阵，如果$\mathbf{A}$有$\mathbf{n}$个特征值$\boldsymbol{\lambda_1,\lambda_2,\dots,\lambda_n}$，对应特征向量$\boldsymbol{V_1,V_2,\dots,V_n}$，以特征向量为列构造矩阵$\mathbf{P}$：$\mathbf{P}=[\boldsymbol{V_1\ V_2\ \cdots\ V_n}]$，则变换

$$\mathbf{AP}=\mathbf{A}[\boldsymbol{V_1\ V_2\ \cdots\ V_n}]=[\boldsymbol{\lambda_1V_1\ \lambda_2V_2\ \cdots\ \lambda_nV_n}]$$

$$\boldsymbol{P^{-1}AP}=\begin{bmatrix}\boldsymbol{V_1}^{-1}\\ \vdots\\ \boldsymbol{V_n}^{-1}\end{bmatrix}[\boldsymbol{\lambda_1V_1\ \cdots\ \lambda_nV_n}]=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$$

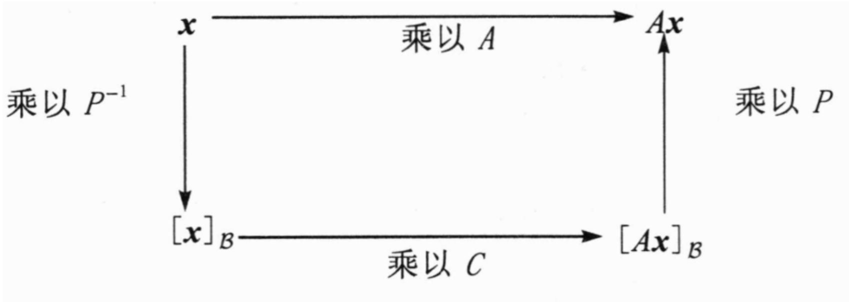
$$\boldsymbol{D}=\boldsymbol{P^{-1}}\mathbf{AP}=\boldsymbol{P^{-1}}\mathbf{A}[\boldsymbol{V_1\ V_2\ \cdots\ V_n}]=\boldsymbol{diag(\lambda_1,\lambda_2,\cdots,\lambda_n)}$$

- $\boldsymbol{A}$的对角阵分解：$\mathbf{A}=\boldsymbol{P^{-1}DP}$

- 【相似变换】
  - 把$\mathbf{A}$变成$\boldsymbol{P^{-1}}\mathbf{AP}$的变换
  - 若矩阵$\mathbf{A}$和$\mathbf{B}$相似，它们有相同的特征多项式和相同的特征值(和相同的重数)

相似变换的几何意义

- 相似变换：$\mathbf{A} = \mathbf{P}\mathbf{C}\mathbf{P}^{-1}$



  - 【几何意义】相似矩阵为线性变换在不同坐标系下的不同的矩阵表达形式
- A可对角化
  - 如果方阵A相似于对角矩阵，即存在可逆矩阵P和对角矩阵D，使得$A = PDP^{-1}$
  - 从分解式$A = PDP^{-1}$(其中D是对角矩阵)，能够了解到有关矩阵A的特征值和特征向量的信息。
- 【注】
  - 此处的变换P不一定是正交变换（旋转和镜像）

对角化矩阵

- 【定理】（列行展开)若A是$\mathbf{m}\times\mathbf{n}$矩阵，B是$\mathbf{n}\times\mathbf{p}$矩阵，则

$$\mathbf{AB}=\begin{bmatrix}col_1(\mathbf{A}) & col_2(\mathbf{A}) & \cdots & col_n(\mathbf{A})\end{bmatrix}\begin{bmatrix}row_1(\mathbf{B})\\ row_2(\mathbf{B})\\ \vdots\\ row_n(\mathbf{B})\end{bmatrix}$$

$$=col_1(\mathbf{A})row_1(\mathbf{B})+\cdots+col_n(\mathbf{A})row_n(\mathbf{B})$$

$$A=P^{-1}M\mathbf{P}$$

- 【外积展开形式】两组向量做外积。信息叠加， λ_i为权重，可用于代表信息
  - $A=\lambda_1u_1v_1^T+\lambda_2u_2v_2^T+\cdots+\lambda_nu_nv_n^T$
  - u_i , v_i分别为P^{-1} 和P^T的列向量
  - 当λ_i很小时可被忽略，则可用作数据压缩

实对称矩阵的对角化

- 若 $A^T = A$，则称A为对称矩阵
- 【Theorem】所有实对称矩阵都可分解为

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T, Q^{-1} = Q^T$$

其中Q由A的正交特征向量组成，A的特征值均为实数。

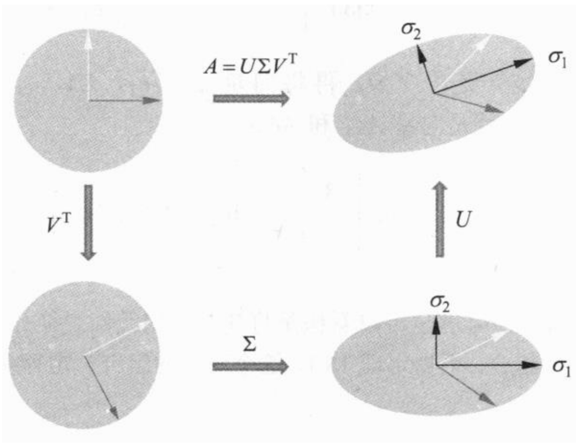
- 【注】
  - 正交矩阵可以看作是一种旋转或镜像变换，它将向量旋转或镜像到一个新的坐标系中。
  - 正交矩阵在计算机图形学中用于进行坐标变换，或在物理学中用于表示刚体的旋转和运动。

奇异值分解

奇异值分解

- 【**定义15.1 (奇异值分解)** 】指将非零的 $m\times n$ 实矩阵$A\in \mathbf{R}^{m\times n}$进行矩阵的因子分解: $A=U\Sigma V^{\mathrm{T}}$，其中U是m阶正交矩阵，V是n阶正交矩阵，Σ 是由降序排列的非负的对角线元素组成的$m\times n$ 对角矩阵

$$
\begin{gathered}
UU^{\mathrm{T}}=I\\
VV^{\mathrm{T}}=I\\
\Sigma=\operatorname{diag}(\sigma_1,\sigma_2,\cdots,\sigma_p)\\
\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_p\geqslant 0\\
p=\mathrm{m}\ \ (m,n)
\end{gathered}
$$



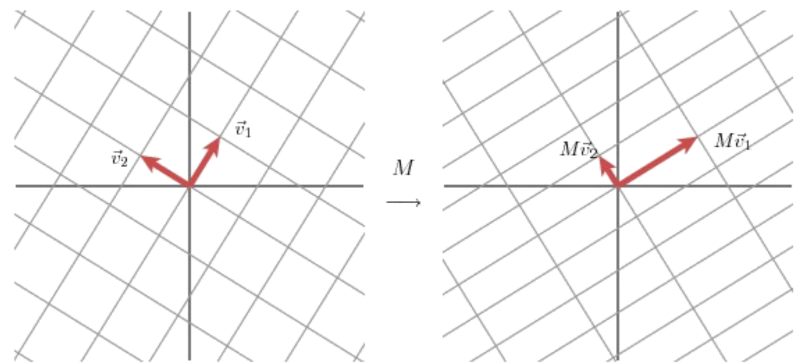
- 则
  - $U\Sigma V^{\mathrm{T}}$称为矩阵A的奇异值分解(singular value decomposition, SVD)
  - σ_i称为矩阵A的奇异值，U的列向量称为左奇异向量，V的列向量称为右奇异向量。
- 奇异值分解不要求矩阵A是方阵，可以看成方阵的对角化推广
- 【注】连续的变换等价于做一个复合变换（变换矩阵连乘代表的变换）。变换从右往左

奇异值分解的一种几何解释

将一组正交向量变换成另一组正向向量

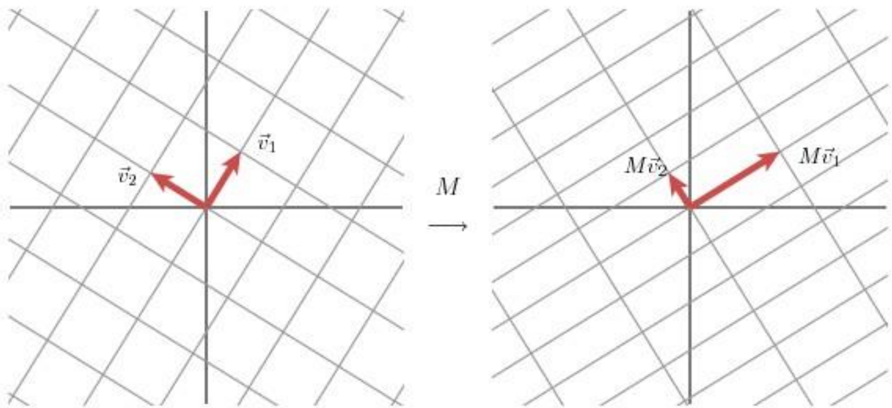
奇异值分解为该变换的变形形式

$AV = U\Sigma$，$A = U\Sigma V^T$



奇异值分解的一种几何解释

- 对于任意的2×2矩阵，通过SVD可以将一个相互垂直的网格(orthogonal grid)变换到另外一个相互垂直的网格。【目的，要找到变换前后都正交的向量】
  - 即，选择两个相互正交的单位向量v_1和v_2，使得向量Av_1和Av_2正交。
- u_1和u_2分别表示Av_1和Av_2的单位向量，$\sigma_1 * u_1 = Av_1$和$\sigma_2 * u_2 = Mv_2$，σ_1和σ_2表示向量Av_1和Av_2的模，称作为矩阵A的奇异值



奇异值分解的一种几何解释

$\begin{matrix} Av_1 = \sigma_1 u_1 \\ Av_2 = \sigma_2 u_2 \end{matrix}$，$A$线性变换，$v_1$和$v_2$是正交的单位向量，向量$x$投影展开

$$x = (v_1 \cdot x)v_1 + (v_2 \cdot x)v_2$$

两边同乘以A

$$Ax = (v_1 \cdot x)Av_1 + (v_2 \cdot x)Av_2$$
$$Ax = (v_1 \cdot x)\sigma_1 u_1 + (v_2 \cdot x)\sigma_2 u_2$$

$(v_1 \cdot x)= v_1^{\mathrm{T}} x$，上式变为

$$Ax = u_1\sigma_1 v_1^{\mathrm{T}} x + u_2\sigma_2 v_2^{\mathrm{T}} x$$
$$A = u_1\sigma_1 v_1^{T} + u_2\sigma_2 v_2^{\mathrm{T}}$$

上式实际上为矩阵$U\Sigma V^T$的外积展开形式，于是

$$A = U\Sigma V^T$$

其中，U的列向量分别是u_1，u_2，$\Sigma = diag(\sigma_1, \sigma_2)$对角矩阵，V的列向量分别是$v_1$，$v_2$

奇异值分解的另一种解释：外积展开形式

$$A = U\Sigma V^{\mathrm{T}} = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_n u_n v_n^{\mathrm{T}}$$

变换视角，把A看作一系列变换组合而成

数据视角，把A看作一系列（投影）分量组合而成

奇异值分解的另一种解释：外积展开形式

矩阵乘法的外积展开形式

$$U\Sigma \quad = [\sigma_1 u_1 \quad \sigma_2 u_2 \quad \cdots \quad \sigma_n u_n]$$

$$V^{\mathrm{T}} = \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix}$$

则

$$A = U\Sigma V^{\mathrm{T}} = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_n u_n v_n^{\mathrm{T}}$$

- 【注】
  - 变换视角，把A看作一系列变换组合而成
  - 数据视角，把A看作一系列分量组合而成

例15.1

给定一个5x4矩阵A

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$$

$$U = \begin{bmatrix} 0 & 0 & \sqrt{0.2} & 0 & \sqrt{0.8} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \sqrt{0.8} & 0 & -\sqrt{0.2} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad V^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

矩阵的奇异值分解不唯一

如果选择U为

$$U=\begin{bmatrix} 0 & 0 & \sqrt{0.2} & \sqrt{0.4} & -\sqrt{0.4} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{0.5} & \sqrt{0.5} \\ 0 & 0 & \sqrt{0.8} & -\sqrt{0.1} & \sqrt{0.1} \end{bmatrix}$$

而Σ和V不变，那么$U\Sigma V^{\mathrm{T}}$也是A的一个奇异值分解

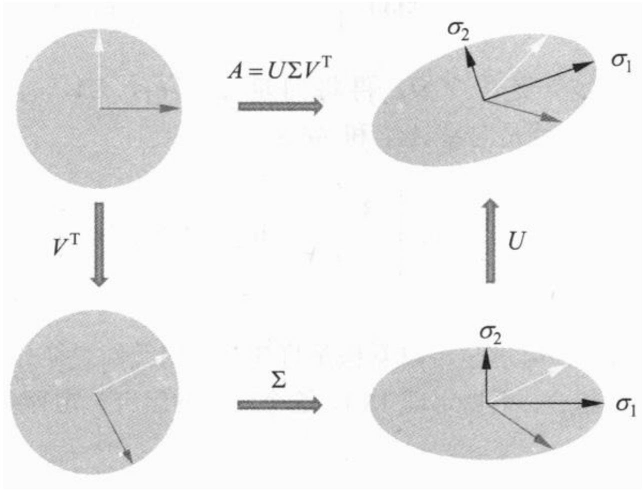
奇异值分解基本定理

【定理15.1(奇异值分解基本定理)】 若A为$\mathrm{m} \times \mathrm{n}$实矩阵, $\mathrm{A} \in \mathrm{R}^{\mathrm{m} \times \mathrm{n}}$ 则A的奇异值分解存在

$$A = U\Sigma V^{\mathrm{T}}$$

其中U是m阶正交矩阵，V是n阶正交矩阵，Σ为$\mathrm{m} \times \mathrm{n}$矩形对角矩阵，其对角线元素非负，且按降序排列。

【证明】对给定的矩阵A，构造出其奇异值分解的各个矩阵。设$m \geqslant n$($m < n$证明类似)



奇异值分解基本定理 - 构造性证明

【关键：A^TA 是对角矩阵】

(1) 确定V和Σ

首先，构造n阶正交实矩阵V和 $m \times n$矩形对角实矩阵Σ

矩阵A是 $m \times n$实矩阵，则矩阵A^TA是n阶实对称矩阵。因而A^TA的特征值都是实数，并且存在一个n阶正交实矩阵V实现A^TA的对角化，使得

$$A^TA = V^T\Lambda V$$

因此$V^T(A^TA)V = \Lambda$，其中Λ是n阶对角矩阵，其对角线元素由A^TA的特征值组成。而且，A^TA的特征值都是非负的【令λ是A^TA的一个特征值，x是对应的特征向量，$A^TAx = \lambda x$，

$|Ax|^2 = (Ax)^T(Ax) = x^TA^TAx = x^T(\lambda x) = x^T\lambda x = \lambda|x|^2$，于是$\lambda = \frac{|Ax|^2}{|x|^2} \geq 0$ 】

假设正交矩阵V的列，其排列使得对应的特征值形成降序排列

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

特征值的平方根(即为矩阵A的奇异值)$\sigma_j = \sqrt{\lambda_j}, j = 1,2,\cdots,n$

确定V和Σ

设矩阵A的秩是r，$rank(A)=r$，则矩阵A^TA的秩也是r。A^TA对称，秩等于正特征值个数，所以

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0, \qquad \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$$

对应地有 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$，$\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$

令$V_1 = [v_1\ v_2\ \ldots v_r]$，$V_2 = [v_{r+1}\ v_{r+2}\ \ldots v_n]$，其中

$v_1 \ldots v_r$为A^TA的正特征值对应的特征向量， $v_{r+1} \ldots v_n$为0特征值对应的特征向量

则$V = [V_1 V_2]$为矩阵A的奇异值分解中的n阶正交矩阵V

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

则r阶对角矩阵$\Sigma_1 = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$， $m \times n$矩形对角矩阵Σ为

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

关于V_2的性质

V_2的列向量是$A^{\mathrm{T}}A$对应于特征值为0的特征向量，因此

$$A^{\mathrm{T}}Av_j = 0, j = r + 1, \cdots, n$$

于是，V_2的列向量构成了$A^{\mathrm{T}}A$的零空间$N\left(A^{\mathrm{T}}A\right)$，而$N\left(A^{\mathrm{T}}A\right) = N(A)$

【所以】：V_2的列向量构成**A**的零空间的一组标准正交基

$$AV_2 = 0$$

由于V是正交矩阵，由$V = [V_1 V_2]$可得

$$I = VV^{\mathrm{T}} = V_1V_1^{\mathrm{T}} + V_2V_2^{\mathrm{T}}$$
$$A = AI = AV_1V_1^{\mathrm{T}} + AV_2V_2^{\mathrm{T}} = AV_1V_1^{\mathrm{T}}$$

15.11

确定U

(2) 确定U

接着构造m阶正交实矩阵,令

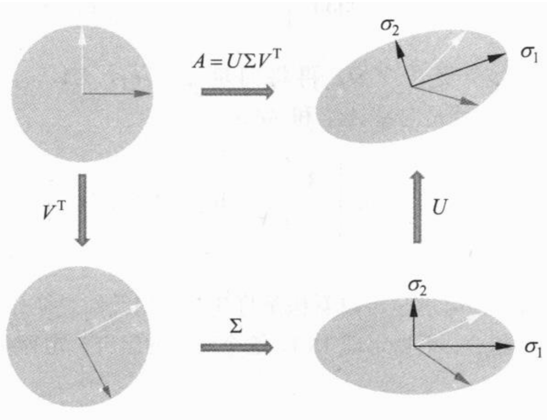
$$u_j = \frac{1}{\sigma_j} A v_j, \quad j = 1,2,\cdots,r$$

$$\mathrm{U}_1 = [u_1\ u_2\ \ldots\ u_r]$$

则有$AV_1 = U_1\Sigma_1$

【形式上$\mathrm{A} = U\Sigma V^T \Longrightarrow U = AV\Sigma^{-1}$，变换：右下-左下-左上=右下-右上】

U_1的列向量构成了一组标准正交集，维数为r【以下为证明】



$$u_i^{\mathrm{T}} u_j = \left(\frac{1}{\sigma_i} v_i^{\mathrm{T}} A^{\mathrm{T}}\right)\left(\frac{1}{\sigma_j} A v_j\right)$$

$$= \frac{1}{\sigma_i \sigma_j} v_i^{\mathrm{T}} (A^{\mathrm{T}} A v_j)$$

$$= \frac{\sigma_j}{\sigma_i} v_i^{\mathrm{T}} v_j$$

$$= \delta_{ij}, i = 1,2,\cdots,r; j = 1,2,\cdots,r$$

确定U_2和U

将A看成是从$\mathbf{R}^n$到$\mathbf{R}^m$的线性变换，则A的列空间和A的值域$R(A)$是相同的。因此 $u_1, u_2, \cdots, u_r$也是$R(A)$的一组标准正交基。

$R(A)^{\perp}$ 表示$R(A)$的正交补，则有$R(A)$的维数为r, $R(A)^{\perp}$ 的维数为m – r、

有$R(A)^{\perp} = N(A^T)$，

设$u_{\mathrm{r}+1}\ u_{\mathrm{r}+2}\ \ldots\ u_m$ 为$N(A^T)$的一组标准正交基【即U_2为$A^T U_2 = 0$的解】，令

$$U_2 = [u_{r+1}\ u_{r+2}\ \ldots\ u_m]$$
$$U = [U_1\ \ U_2]$$

则$u_1\ u_2\ \ldots\ u_m$构成了R^m的一组标准正交基

因此，U是m阶正交矩阵。为矩阵A的奇异值分解中的m阶正交矩阵。

紧/截断奇异值分解

紧奇异值分解与截断奇异值分解

- $A = U\Sigma V^{\mathrm{T}}$又称为矩阵的完全奇异值分解(full singular value decomposition)
- 常用方式
  - 紧奇异值分解，与原始矩阵等秩的奇异值分解
  - 截断奇异值分解，比原始矩阵低秩的奇异值分解

紧奇异值分解

定义15.2 $\mathrm{m} \times \mathrm{n}$实矩阵$A$，其秩为$rank(A) = r, r \leq \min(m, n)$，则称$U_r \Sigma_r V_r^{\mathrm{T}}$为$A$的紧奇异值分解(compact singular value decomposition)，即

$$\mathrm{A} = U_r \Sigma_r V_r^{\mathrm{T}}$$

U_r： $\mathrm{m} \times \mathrm{r}$ 矩阵， U的前r列

V_r： $\mathrm{n} \times \mathrm{r}$ 矩阵， V的前r列

Σ_r： r阶对角矩阵， Σ的前r个对角线元素，且秩相等

【注意】

紧的意思是，对焦矩阵信息表达效率最高，没有全为0的行/列

例15.2

矩阵A的秩r = 3

$$A=\begin{bmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{bmatrix}$$

$$U_r=\begin{bmatrix}0&0&\sqrt{0.2}\\1&0&0\\0&1&0\\0&0&0\\0&0&\sqrt{0.8}\end{bmatrix}\quad \Sigma_r=\begin{bmatrix}4&0&0\\0&3&0\\0&0&\sqrt{5}\end{bmatrix},\quad V_r^{\mathrm{T}}=\begin{bmatrix}0&0&0&1\\0&1&0&0\\1&0&0&0\end{bmatrix}$$

$$A=U_r\Sigma_rV_r^{\mathrm{T}}$$

截断奇异值分解

- 在矩阵的奇异值分解中，取最大的k个奇异值($k<r$, r为矩阵的秩)对应的部分，得到矩阵的截断奇异值分解

- 实际应用中，提到矩阵的奇异值分解，通常指截断奇异值分解

截断奇异值分解

（**定义15.3**）设A为$m \times n$实矩阵, 其秩$\operatorname{rank}(A) = r$, 且 $0 < k < r$, 则称 $U_k \Sigma_k V_k^{\mathrm{T}}$ 为矩阵 A 的截断奇异值分解 (truncated singular value decomposition)

$$A \approx U_k \Sigma_k V_k^{\mathrm{T}}$$

其中U_k是$m \times k$矩阵, V_k是$n \times k$矩阵, Σ_k是k阶对角矩阵; 矩阵U_k由完全奇异值分解中U的前k列、矩阵V_k由V的前k列、矩阵Σ_k由Σ的前k个对角线元素得到。对角矩阵Σ_k的秩比原始矩阵 A 的秩低。

例

例 **15.3** 矩阵$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$

的秩为 3 , 若取 $k = 2$ 则其截断奇异值分解是$A \approx A_2 = U_2 \Sigma_2 V_2^{\mathrm{T}}$

其中

$$U_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, V_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$A_2 = U_2 \Sigma_2 V_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

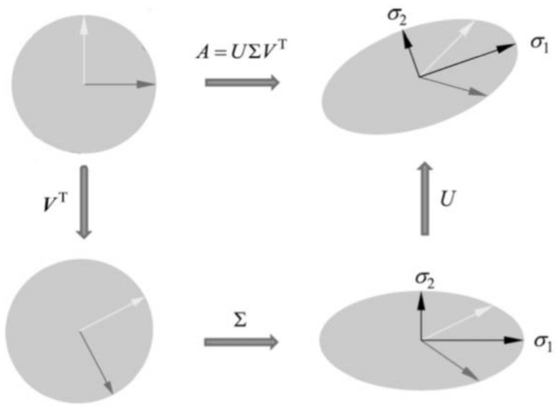
U_2, V_2 是例 15.1 的U和V的前2列, Σ_2是Σ的前2行前2列。 A_2与A比较, A的元素1和2在A_2中均变成 0

理解奇异值分解：几何角度

线性变换分解为连续的基础变换

从线性变换的角度理解奇异值分解

- $m \times n$矩阵A表示R^n到R^m的一个线性变换
  - T: $x \ \to A\,x$，$x \in R^n, Ax \in R^m$
- 线性变换T分解为三个简单的变换
  - 坐标系的旋转或反射变换V^{T}
    - V的列向量$v_1, \cdots, v_n$构成R^n的一组标准正交基，表示旋转或反射变换
  - 坐标轴的缩放变换Σ
    - Σ的对角元素$\sigma_1, \cdots, \sigma_n$ 非负实数，R^m中原始正交坐标系的缩放变换
  - 坐标系的旋转或反射变换U
    - U的列向量$u_1, \cdots, u_m$构成R^m一组标准正交基，表示旋转或反射变换
- 奇异值定理保证这种分解一定存在



几何解释

- 任意一个向量$x \in R^n$，经过基于$A = U\Sigma V^{\mathrm{T}}$的线性变换，等价于
  - 经过坐标系的旋转或反射变换V^{T}
  - 坐标轴的缩放变换Σ
  - 以及坐标系的旋转或反射变换U
  - 得到向量$Ax \in R^m$

主要性质

1) 设矩阵A的奇异值分解为$A = U\Sigma V^T$，则一下关系成立：

$$A^{\mathrm{T}}A = \left(U\Sigma V^{\mathrm{T}}\right)^{\mathrm{T}}\left(U\Sigma V^{\mathrm{T}}\right) = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = V\left(\Sigma^{\mathrm{T}}\Sigma\right)V^{\mathrm{T}}$$

$$AA^{\mathrm{T}} = \left(U\Sigma V^{\mathrm{T}}\right)\left(U\Sigma V^{\mathrm{T}}\right)^{\mathrm{T}} = U\left(\Sigma\Sigma^{\mathrm{T}}\right)U^{\mathrm{T}}$$

矩阵$A^{\mathrm{T}}A$和AA^{T}的特征分解存在，且可以由矩阵A的奇异值分解的矩阵表示

V的列向量是$A^{\mathrm{T}}A$的特征向量

U的列向量是AA^{T}的特征向量

Σ的奇异值是$A^{\mathrm{T}}A$和AA^{T}的特征值的平方根。

主要性质

2) 在矩阵A的奇异值分解中，奇异值、左奇异向量和右奇异向量之间存在对应关系

由$A = U\Sigma V^T$易知$AV = U\Sigma$，比较这一等式两端的第j列，得到

$$Av_j = \sigma_j u_j, j = 1,2,\cdots,n$$

矩阵A的右奇异向量和奇异值、 左奇异向量的关系

类似地，由$A^{\mathrm{T}}U = V\Sigma^{\mathrm{T}}$

得到

$$A^{\mathrm{T}}u_j = \sigma_j v_j, j = 1,2,\cdots,n$$

$$A^{\mathrm{T}}u_j = 0, j = n+1, n+2, \cdots, m$$

矩阵A的左奇异向量和奇异值、右奇异向量的关系。

主要性质

3) 矩阵A的奇异值分解中，奇异值$\sigma_1, \sigma_2, \cdots, \sigma_n$唯一，而矩阵U和V不唯一。

零空间的基不唯一

4) 矩阵A和Σ的秩相等，等于正奇异值σ_i的个数r(包含重复的奇异值)

5) 矩阵A的r个右奇异向量$v_1, v_2, \cdots, v_r$构成A^{T}的值域$R\left(A^{\mathrm{T}}\right)$的一组标准正交基

因为矩阵A^{T}是从$\mathbf{R}^m$映射到$\mathbf{R}^n$的线性变换，则A^{T}的值域$R\left(A^{\mathrm{T}}\right)$和AT的列空间是相同的，$v_1, v_2, \cdots, v_r$是$A^{\mathrm{T}}$的一组标准正交基，因而也是$R\left(A^{\mathrm{T}}\right)$的一组标准正交基。

矩阵A的$n-r$个右奇异向量$v_{r+1}, v_{r+2}, \cdots, v_n$构成$A$的零空间$N(A)$ 的一组 标准正交基。

矩阵A的r个左奇异向量$u_1, u_2, \cdots, u_r$构成值域$R(A)$的一组标准正交基。

矩阵A的$m-r$个左奇异向量$u_{r+1}, u_{r+2}, \cdots, u_m$构成$A^{\mathrm{T}}$的零空间$N\left(A^{\mathrm{T}}\right)$ 的 一组标准正交基

2 奇异值分解的计算

奇异值分解的计算

- 矩阵A的奇异值分解可以通过求对称矩阵$A^{\mathrm{T}}A$的特征值和特征向量得到。
- $A^{\mathrm{T}}A$的特征向量构成正交矩阵V的列
- $A^{\mathrm{T}}A$的特征值λ_j的平方根为奇异值σ_i，即$\sigma_j = \sqrt{\lambda_j}, j = 1,2,\cdots,n$

- 对其由大到小排列作为对角线元素，构成对角矩阵Σ
- 求正奇异值对应的左奇异向量，再求扩充的A^{T}的标准正交基，构成正交矩阵U的列
- 从而得到A的奇异值分解$A = U\Sigma V^{T}$

奇异值分解的计算

1) 求$A^{\mathrm{T}}A$的特征值和特征向量， $A^{\mathrm{T}}A$的特征向量构成正交矩阵V的列

计算对称矩阵$W=A^{\mathrm{T}}A$，求解特征方程$(W-\lambda I)x=0$

得到特征值λ_{i}，并将特征值由大到小排列：$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n\geq 0$

将特征值λ_{i} 代入特征方程求得对应的特征向量

2) 求n阶正交矩阵V

将特征向量单位化，得到单位特征向量$v_1,v_2,\cdots,v_n$ ，构成n阶正交矩阵V：$V=[v_1\ v_2\ldots v_n]$

3)求$\mathrm{m}\times\mathrm{n}$对角矩阵Σ

计算A的奇异值$\sigma_j=\sqrt{\lambda_j},j=1,2,\cdots,n$，构造 $\mathrm{m}\times\mathrm{n}$矩形对角矩阵Σ：$\Sigma=diag(\sigma_1,\sigma_2,\cdots,\sigma_n)$

4) 求m阶正交矩阵U

对A的前r个正奇异值，令$u_j=\frac{1}{\sigma_j}Av_j,j=1,2,\cdots,r$，得到$U_1=[u_1\ u_2\ \cdots\ u_r]$

求A^{T}的零空间【 $A^{\mathrm{T}}X=0$ 的解】的一组标准正交基$\{u_{r+1},\cdots,u_m\}$，令$U_2=[u_{r+1}\cdots\ u_m]$，令$U=[U_1\ U_2]$

5)得到奇异值分解$A=U\Sigma V^T$

3 奇异值分解与矩阵近似

矩阵的最优近

矩阵代表信息量，矩阵在给定度量下的近似表达

弗罗贝尼乌斯范数

奇异值分解也是一种矩阵近似的方法（在弗罗贝尼乌斯范数(Frobenius norm)意义下的近似）

【即，矩阵A经过奇异值分解-截断近似-变换回原空间得矩阵X， A和X的弗罗贝尼乌斯范数接近】

【此时的矩阵不再只是看作是变换，也可以代表数据本身】

矩阵的弗罗贝尼乌斯范数是向量的L_2范数的直接推广，对应着机器学习中的平方损失函数

定义15.4 (弗罗贝尼乌斯范数) 矩阵$A \in \mathbf{R}^{m\times n}, A = \left[a_{ij}\right]_{m\times n}$, 定义矩阵$A$的弗罗贝尼乌斯范数为

$$\| A \|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \left(a_{ij}\right)^2\right)^{\frac{1}{2}}$$

引理15.1 设矩阵 $A \in \mathbf{R}^{m\times n}, A$ 的奇异值分解为 $U\Sigma V^{\mathrm{T}}$, 其中 $\Sigma = \operatorname{diag}\left(\sigma_1, \sigma_2, \cdots, \sigma_n\right)$, 则

$$\| A \|_F = \left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)^{\frac{1}{2}}$$

弗罗贝尼乌斯范数

设矩阵$A \in R^{m\times n}$, A的奇异值分解为$U\Sigma V^T$, 其中$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$，则

$$|A|_F = \left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)^{\frac{1}{2}}$$

【定理15.2 】设矩阵$A \in R^{m\times n}$, 矩阵的秩$rank(A) = r$, 并设M 为$R^{m\times n}$ 中所有秩不超过k 的矩阵集合，$0 < k < r$, 则存在一个秩为k 的矩阵$X \in \mathcal{M}$, 使得

$$|A - X|_F = \min_{S\in\mathcal{M}} |A - S|_F$$

称矩阵X 为矩阵A 在弗罗贝尼乌斯范数意义下的最优近似。

定理15.3

【定理15.3】设矩阵$A \in R^{m\times n}$, 矩阵的秩$rank(A) = r$, A的奇异值分解为$U\Sigma V^T$,并设M 为$R^{m\times n}$ 中所有秩不超过k 的矩阵集合，$0 < k < r$, 若秩为k 的矩阵$X \in \mathcal{M}$满足$|A - X|_F = \min_{S\in\mathcal{M}} |A - S|_F$，则

$$|A - X|_F = \left(\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2\right)^{\frac{1}{2}}$$

特别地，若$A' = U\Sigma' V^T$, 其中

$$\Sigma' = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_k & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix}$$

则

$$|A - A'|_F = \left(\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2\right)^{\frac{1}{2}} = \min_{S\in\mathcal{M}} |A - S|_F$$

矩阵的最优近似

- 在秩不超过k的$m \times n$矩阵的集合中，存在矩阵A的弗罗贝尼乌斯范数意义下的最优近似矩阵X， $A^{'} = U\Sigma^{'}V^{T}$是达到最优值的一个矩阵

- 紧奇异值分解是在弗罗贝尼乌斯范数意义下的无损压缩
- 截断奇异值分解是有损压缩
- 截断奇异值分解得到的矩阵的秩为k，通常远小于原始矩阵的秩r，所以是由低秩矩阵实现了对原始矩阵的压缩

矩阵的外积展开式

矩阵看成变换，矩阵由一系列外积形式的变化组合而成

矩阵的外积展开式

矩阵A的奇异值分解$A = U\Sigma V^T$也可以由外积形式表示

若将A的奇异值分解看成矩阵$U\Sigma$和V^T的乘积，将$U\Sigma$按列向量分块，将V^T按行向量分块，即得

$$U\Sigma = [\sigma_1 u_1 \quad \sigma_2 u_2 \quad \cdots \quad \sigma_n u_n],\ V^{\mathrm{T}} = \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix}$$

则【直观上，每一个$A_{i,j}$都是$U\Sigma$ 第i行和V^{T}第j列内积而来，即为下式的i,j的值】

$$A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_n u_n v_n^{\mathrm{T}}$$

其中$u_i v_j^{\mathrm{T}}$是矩阵，$u_i v_j^{\mathrm{T}} = \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{mi} \end{bmatrix} [u_{1j}\ v_{2j}\ \cdots\ v_{nj}] = \begin{bmatrix} u_{1i}v_{1j} & u_{1i}v_{2j} & \cdots & u_{1i}v_{nj} \\ u_{2i}v_{1j} & u_{2i}v_{2j} & \cdots & u_{2i}v_{nj} \\ \vdots & \vdots & & \vdots \\ u_{mi}v_{1j} & u_{mi}v_{2j} & \cdots & u_{mi}v_{nj} \end{bmatrix}$

A的外积展开式也可写为 $A = \sum_{k=1}^{n} A_k = \sum_{k=1}^{n} \sigma_k u_k v_k^{\mathrm{T}}$，为矩阵的有序加权

矩阵的外积展开式

由矩阵A的外积展开式知，若A的秩为n，则

$$A=\sigma_1 u_1 v_1^{\mathrm{T}}+\sigma_2 u_2 v_2^{\mathrm{T}}+\cdots+\sigma_n u_n v_n^{\mathrm{T}}$$

设矩阵

$$A_k=\sigma_1 u_1 v_1^{\mathrm{T}}+\sigma_2 u_2 v_2^{\mathrm{T}}+\cdots+\sigma_k u_k v_k^{\mathrm{T}}$$

则A_k的秩为k，并且A_k是秩为k矩阵在弗罗贝尼乌斯范数意义A的最优近似矩阵，矩阵A_k就是A的截断奇异值分解

由于通常奇异值σ_{i}递减很快，所以k取很小值时，A_k也可以对A有很好的近似。