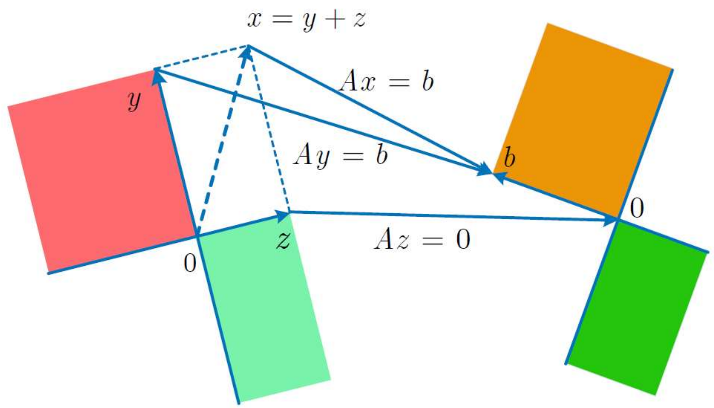
第十讲 矩阵论

2022.11.4



Recap

- 向量空间是具有线性结构的集合，即集合元素定义了加法和数乘运算、包含加法单位元(零)且满足通常的运算规则（交换律、结合律和分配律）。
- 向量空间中元素称为向量，可以是通常意义下欧氏空间中的向量，也可以是映射、函数或其它任何其它数学对象。
- R^S是一类重要的向量空间，其中的向量元素是集合S到实数轴R的所有函数的集合。特别地，$R^n = R^{\{1,2,\ldots,n\}}$由所有$\{1,2,\ldots,n\}$上的实函数构成。
- 保持向量空间线性结构的向量空间之间的映射称为线性变换。有限维向量空间之间的线性变换在特定基下表示为矩阵。不同基下的矩阵相似。

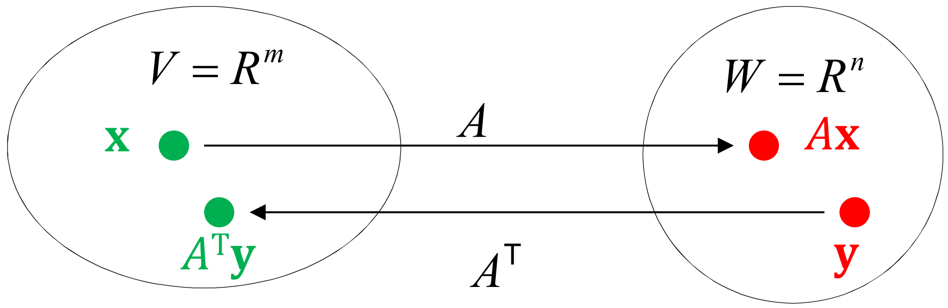
- 内积向量空间之间的线性变换$T \in L(V,W)$有唯一的伴随(共轭)变换$T^* \in L(V,W)$，使得

$$(T\mathbf{v},\mathbf{w}) = (\mathbf{v},T^*\mathbf{w}),\quad \forall \mathbf{v} \in V,\ \mathbf{w} \in W$$

- 对于有限维向量空间， 矩阵A的伴随矩阵是其转置矩阵A^{T}，满足

$$(A\mathbf{x},\mathbf{y}) = (\mathbf{x},A^{\mathrm{T}}\mathbf{y})，\text{即}(A\mathbf{x})^{\mathrm{T}}\mathbf{y} = \mathbf{x}^{\mathrm{T}}(A^{\mathrm{T}}\mathbf{y}),$$

注意两端的内积是在不同空间计算的，如下图所示：



内容

4. 矩阵的行与列

列空间
行空间

矩阵$A_{n\times m}=(\mathbf{a}_1,\dots,\mathbf{a}_m)$的所有列张成的线性空间$C(A)$（或$L(A)=\mathrm{Img}(A)$）:

$$C(A)=C(\mathbf{a}_1,\dots,\mathbf{a}_m)=\left\{\sum_{i=1}^{m}\mathbf{a}_i x_i \mid x_1,\dots,x_m\in R\right\}=\left\{A\mathbf{x}\mid \mathbf{x}\in R^m\right\}.$$

$A_{n\times m}$的行空间为A的所有行向量张成的空 间，即 A^{T}的列空间：

$$C(A^{\mathsf{T}})=\left\{A^{\mathsf{T}}\mathbf{y}\mid \mathbf{y}\in R^n\right\}$$

例. $A=\begin{pmatrix}3&1&2\\9&3&6\\6&2&4\end{pmatrix}$的行、列空间？

$$\begin{array}{c|ccc} & 3 & 1 & 2\\ \hline 1 & 3 & 1 & 2\\ 3 & 9 & 3 & 6\\ 2 & 6 & 2 & 4\end{array}$$

$$A=\begin{pmatrix}1\\3\\2\end{pmatrix}\begin{pmatrix}3&1&2\end{pmatrix}\overset{\Delta}{=}\mathbf{u}\mathbf{v}^{\mathsf{T}}\quad\Rightarrow\quad \begin{array}{l}C(A)=\left\{A\mathbf{x}=\mathbf{u}(\mathbf{v}^{\mathsf{T}}\mathbf{x})\right\}=C(\mathbf{u})\\ C(A^{\mathsf{T}})=\left\{A^{\mathsf{T}}\mathbf{x}=\mathbf{v}(\mathbf{u}^{\mathsf{T}}\mathbf{x})\right\}=C(\mathbf{v})\end{array}$$

我们将会看到，秩r矩阵可以分解为若干秩 1矩阵之和（SVD）:

$$A=\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}}+\dots+\mathbf{u}_r\mathbf{v}_r^{\mathsf{T}}$$，诸$\mathbf{u}_i$正交，诸$\mathbf{v}_j$正交.

列空间和行空间维数相同

列秩, 行秩

列秩 = 列空间的维数 = $\mathrm{rank}(A)$; 行秩 = 行空间的维数 = $\mathrm{rank}(A^{\mathsf{T}})$

下面的定理1表明，虽然行向量和列向量长度不同，但它们张成空间的维数是一样的, 这是线性代数中最为深刻的结论之一。

行秩等于列秩

定理2：对任一实数矩阵A，$\mathrm{rank}(A^{\mathsf{T}}) = \mathrm{rank}(A)$ (称为矩阵A的秩), 即行空间和列空间维数相同：$\dim(C(A^{\mathsf{T}})) = \dim(C(A))$.

引理1：（1）若$\mathbf{x}_1,\dots,\mathbf{x}_s$为行空间$C(A^{\mathsf{T}})$的一组线性无关向量,则$A\mathbf{x}_1,\dots, A\mathbf{x}_s$是列空间$C(A)$中的一组线性无关向量。（2）反之,列空间的一组线性无关向量组的A^{T}变换，在行空间也是线性无关的。

定理1的证明：引理 1(1)说明$\dim(C(A)) \ge \dim(C(A^{\mathsf{T}}))$, (2)说明
$\dim(C(A)) \le \dim(C(A^{\mathsf{T}}))$, 所以$\dim(C(A)) = \dim(C(A^{\mathsf{T}}))$, 即$\mathrm{rank}(A) = \mathrm{rank}(A^{\mathsf{T}})$.

引理1的证明：

假设$\mathbf{x}_1,\ldots,\mathbf{x}_s\in C(A^{\top})$是一个线性无关组,下面证明$A\mathbf{x}_1,\ldots,A\mathbf{x}_s\in C(A)$也一定是线性无关的。假设存在实数$c_1,..,c_s$,使得

$$\mathbf{0}=A\mathbf{x}_1c_1+\ldots+A\mathbf{x}_sc_s\triangleq A\mathbf{v},\ \text{其中}\ \mathbf{v}\triangleq\mathbf{x}_1c_1+\ldots+\mathbf{x}_sc_s\text{。}$$

$A\mathbf{v}=\mathbf{0}$表明$\mathbf{v}$与A的各行正交，即$\mathbf{v}\perp C(A^{\top})$.

但我们假设了$\mathbf{v}$属于A的行张成的空间$C(A^{\top})\Rightarrow\mathbf{v}\perp\mathbf{v}$，$\Rightarrow\mathbf{v}=\mathbf{x}_1c_1+\ldots+\mathbf{x}_sc_s=\mathbf{0}$，$\mathbf{x}_1,\ldots,\mathbf{x}_s$线性无关$\Rightarrow c_1=\ldots=c_s=0$，所以$A\mathbf{x}_1,\ldots,A\mathbf{x}_s$线性无关.

反之,若$\mathbf{y}_1,\ldots,\mathbf{y}_t\in C(A)$线性无关,则类似地可证明$A^{\top}\mathbf{y}_1,\ldots,A^{\top}\mathbf{y}_t$在行空间中是线性无关 的。至此我们证明了引 理1，也证明了定理1。

注：实际上我们也证明了：

$$C(A^{\top})=C(A^{\top}A)=C(A^{\top}AA^{\top})=\ldots,\quad C(A)=C(AA^{\top})=C(AA^{\top}A)=\ldots$$

这是因为：$\mathbf{x}_1,\ldots,\mathbf{x}_s\in C(A^{\top})$线性无关$\Rightarrow\mathbf{y}_1=A\mathbf{x}_1,\ldots,\mathbf{y}_s=A\mathbf{x}_s\in C(A)$线性无关

$\Rightarrow A^{\top}\mathbf{y}_1=A^{\top}A\mathbf{x}_1,\ \ldots,A^{\top}\mathbf{y}_s=A^{\top}A\mathbf{x}_s\in C(A^{\top})$线性无关$\Rightarrow\dim(C(A^{\top}A))\geq\dim(C(A^{\top}))$

但$C(A^{\top}A)\subset C(A^{\top})$，所以$C(A^{\top})=C(A^{\top}A)$。

零/核空间

矩阵$A_{n\times m}$的核空间(kernel space)或零空间(null space):

$$N(A)=\ker(A)=\{\mathbf{x}\in R^m \mid A\mathbf{x}=\mathbf{0}\},$$

$N(A)$为与A的各行正交的向量构成的子空间。

核空间的维数称为$\text{nullity}(A)=\dim(N(A))$。

同样$N(A^{\mathrm{T}})=\{\mathbf{y}\in R^n : A^{\mathrm{T}}\mathbf{y}=\mathbf{0}\}$是与$A$的列正交的核空间。

注1：若$\mathbf{x}$是$A\mathbf{x}=\mathbf{b}$的一个解，则$\mathbf{x}+\ker(A)$构成所有解.

注2：方阵A可逆$\Leftrightarrow A\mathbf{x}=\mathbf{0}$只有$\mathbf{x}=\mathbf{0}$解$\Leftrightarrow N(A)=\{\mathbf{0}\}\Leftrightarrow \text{nullity}(A)=0$

例1. $T: R^{\infty}\rightarrow R^{\infty}$, $T(x_1,x_2,x_3,\ldots)=(x_2,x_3,\ldots)$, $N(T)=(x_1,0,0,\ldots.)$.

例2. $T: R^3\rightarrow R^2$, $T(x,y,z)=(2x+y,\ y+z)$, $N(T)=\{(-\frac{a}{2},a,-a): a\in R\}$。

另外，T的矩阵表示为$A=\begin{pmatrix}2 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}$,

$$C(A)=\{(2x+y,\ y+z): x,y,z\in R\}=R^2,\quad N(A)=\{(-\frac{a}{2},a,-a): a\in R\},$$

A的各行与$N(A)$正交。

$T^*: R^2\rightarrow R^3, T^*(u,v)=(2u,u+v,v), C(A^T)=\{(2u,u+v,v): u,v\in R\}=N(A)^{\perp}$

$N(T^*)=N(A^T)=\{0\}=C(A)^{\perp}$.

四个基本空间

定理3. $N(A) = C(A^\top)^\perp, \quad N(A^\top) = C(A)^\perp$

$$\operatorname{rank}(A_{n\times m}) = m - \operatorname{nullity}(A) = n - \operatorname{nullity}(A^\top)$$

证明：设$\forall \mathbf{x} \in N(A), A\mathbf{x} = \mathbf{0}$. 对$\forall \mathbf{y} = A^\top \mathbf{b} \in C(A^\top)$,

$(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^\top \mathbf{b}) = (A\mathbf{x}, \mathbf{b}) = (0, \mathbf{b}) = 0$, 所以$\mathbf{x} \in C(A^\top)^\perp$，$N(A) \subset C(A^\top)^\perp$.

反之，若$\mathbf{x} \in C(A^\top)^\perp, \mathbf{x} \perp C(A^\top)$, 特别地正交于$A^\top$的每一列即$A$的每一行$\mathbf{a}_i$,

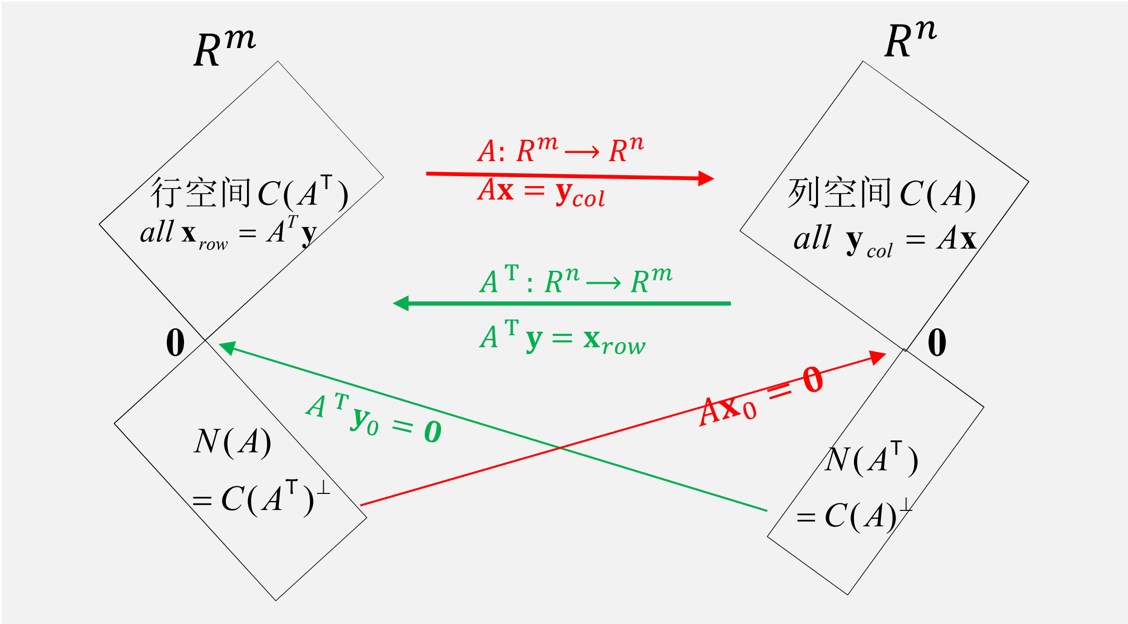
$$\text{所以} A\mathbf{x} = \begin{pmatrix} \mathbf{a}_1^\top \mathbf{x} \\ \\ \mathbf{a}_n^\top \mathbf{x} \end{pmatrix} = \mathbf{0}, \text{所以} \mathbf{x} \in N(A), \text{所以} C(A^\top)^\perp \subset N(A).$$

所以 $N(A) \oplus C(A^\top) = R^m$,

$$m = \dim(N(A)) + \dim(C(A^\top)) = \operatorname{nullity}(A) + \operatorname{rank}(A).$$

同样$N(A^\top) \oplus C(A) = R^n, C(A_{n\times m}) = N(A^\top)^\perp, \operatorname{rank}(A) + \operatorname{nullity}(A^\top) = n$

The big picture (Strang)



定理4. $C(A)=C(AA^T)$, $N(A^T)=N(AA^T)$
$C(A^T)=C(A^TA)$, $N(A)=N(A^TA)$

$$\boxed{A\mathbf{x}=0 \Leftrightarrow A^{\mathsf T}A\mathbf{x}=0}$$

证明1:

若$\mathbf{y}\in N(A^{\mathsf T})$, $A^{\mathsf T}\mathbf{y}=0 \Rightarrow AA^{\mathsf T}\mathbf{y}=0 \Rightarrow N(A^{\mathsf T})\subseteq N(AA^{\mathsf T})$

若$\mathbf{y}\in N(AA^{\mathsf T})$,即$AA^{\mathsf T}\mathbf{y}=\mathbf{0} \Rightarrow \mathbf{y}^{\mathsf T}AA^{\mathsf T}\mathbf{y}=\|A^{\mathsf T}\mathbf{y}\|^2=0 \Rightarrow A^{\mathsf T}\mathbf{y}=\mathbf{0}$

$\Rightarrow \mathbf{y}\in N(A^{\mathsf T}) \Rightarrow N(AA^{\mathsf T})\subseteq N(A^{\mathsf T})$, 所以$N(AA^{\mathsf T})=N(A^{\mathsf T})$,

进而$C(A)=N(A^{\mathsf T})^{\perp}=N(AA^{\mathsf T})^{\perp}=C(AA^{\mathsf T})$。

证明2: $C(A^{\mathsf T}A)\subset C(A^{\mathsf T}) \Rightarrow rank(A^{\mathsf T}A)\le rank(A^{\mathsf T})$.

由引理1, $\mathbf{x}_1,\ldots,\mathbf{x}_s\in C(A^T)$线性无关 $\Rightarrow A\mathbf{x}_1,\ldots,A\mathbf{x}_s\in C(A)$线性无关。

$\Rightarrow A^{\mathsf T}A\mathbf{x}_1,\ldots,A^{\mathsf T}A\mathbf{x}_s\in C(A^{\mathsf T}A)$线性无关, $\Rightarrow rank(A^{\mathsf T}A)=\dim(C(A^{\mathsf T}A))\ge rank(A^{\mathsf T})$

证明3: 显然$C(AA^{\mathsf T})\subset C(A)$。假设$\mathbf{y}_{col}\in C(A)$,则存在某个$\mathbf{x}\in R^m$, $\mathbf{y}_{col}=A\mathbf{x}$,

因为存在$\mathbf{x}_{row}\in C(A^{\mathsf T})$, $\mathbf{x}_0\in N(A)$, 使得$\mathbf{x}=\mathbf{x}_{row}+\mathbf{x}_0$,

所以$\mathbf{y}_{col}=A\mathbf{x}=A(\mathbf{x}_{row}+\mathbf{x}_0)=A\mathbf{x}_{row}$,

因为$\mathbf{x}_{row}\in C(A^{\mathsf T})$,存在某个$\mathbf{y}\in R^n$使得$\mathbf{x}_{row}=A^{\mathsf T}\mathbf{y}$,

所以$\mathbf{y}_{col}=A\mathbf{x}_{row}=AA^{\mathsf T}\mathbf{y}\in C(AA^{\mathsf T})\Rightarrow C(A)\subset C(AA^{\mathsf T})\Rightarrow C(A)=C(AA^{\mathsf T})$

5.方阵的谱分解

特征向量：方阵/算子变换下的不变量，列空间的刻画

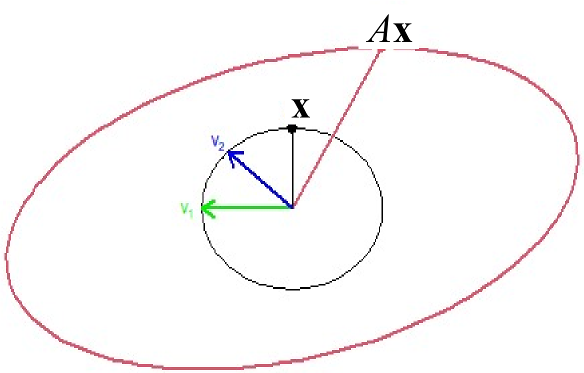
Brouwer不动点定理

特征根/向量的定义

定义：对于$n\times n$方阵A，若存在数λ和向量$\mathbf{v}, \mathbf{v}\neq\mathbf{0}$, 使得

$$A\mathbf{v}=\lambda\mathbf{v},$$

则$\mathbf{v}$和λ分别称为A的特征根和特征向量(注意不限制λ是实数以及$\mathbf{v}\in R^n$)。



例. 假设$A=\mathbf{u}\mathbf{v}^T, \mathbf{u}, \mathbf{v}\in R^n, C(A)=C(\mathbf{u})$。因为$A\mathbf{u}=\mathbf{u}(\mathbf{v}^T\mathbf{u})$, 所以$\mathbf{v}^T\mathbf{u}$是(唯一非$\mathbf{0}$)特征根，$\mathbf{u}$是对应的特征向量，它刻画了$A$的列。

对任何方阵A，假设所有特征根和特征向量为$\lambda_i, \mathbf{x}_i i=1,\ldots,n$，

记$V=(\mathbf{x}_1,\ldots,\mathbf{x}_n)$为所有特征向量组成的矩阵，则特征方程 $A\mathbf{x}_i=\mathbf{x}_i\lambda_i$，

$$i=1,\ldots,n \quad \Leftrightarrow A(\mathbf{x}_1,\ldots,\mathbf{x}_n)=(\mathbf{x}_1,\ldots,\mathbf{x}_n)\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix} \Leftrightarrow AV=V\Lambda,$$

若V可逆，则$A=V\Lambda V^{-1}$

若A是对称矩阵，则所有特征根为实数，且所有特征向量可取为正交的，因此有谱分解：

对称矩阵的谱分解

定理5. 任一$n\times n$对称阵A可表示为

$$A=O\Lambda O^{\mathsf{T}}$$

其中$\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$, $O=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$是正交矩阵, $O^{\mathsf{T}}O=OO^{\mathsf{T}}=I_n$,

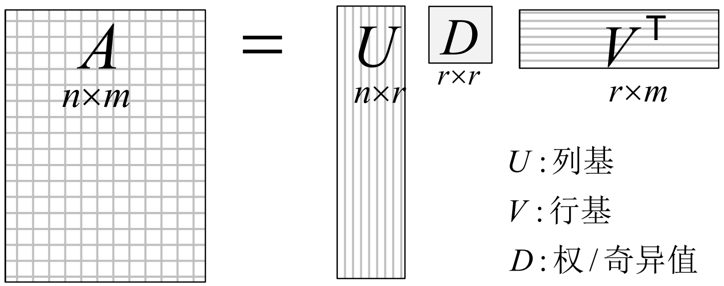
其中 λ_i是A的特征根，$\mathbf{v}_i$是对应的模为1的特征向量。

6. 矩阵的奇异值分解
（SVD Singular value decomposition ）

任一不是方阵的$n\times m$矩阵A没有特征根和特征向量，其行和列如何用类似于特征向量的不变 量进行刻画？

奇异值分解 (SVD : singular value decomposition):

$$SVD: A = UDV^{\mathsf{T}}$$



U:列基

V:行基

D:权/奇异值

key: 利用对称方阵AA^{T}以及$A^{\mathsf{T}}A$的特征向量分别刻画A和A^{T}的列。

（$C(A)=C(AA^{\mathsf{T}})$，$C(A^{\mathsf{T}})=C(A^{\mathsf{T}}A)$）。

$A^{\mathsf T}A, AA^{\mathsf T}$特征向量的对偶关系

引理2. 假设$\lambda \neq 0$是$A^{\mathsf T}A$或$AA^{\mathsf T}$的特征根。对应于λ，我们有如下结论:
(1) 若$\mathbf{v}$是$A^{\mathsf T}A$的单位特征向量, 则$\mathbf{u} = A\mathbf{v}/\sqrt{\lambda}$是$AA^{\mathsf T}$的单位特征向量;
(2) 若$\mathbf{u}$是$AA^{\mathsf T}$的单位特征向量, 则$\mathbf{v} = A^{\mathsf T}\mathbf{u}/\sqrt{\lambda}$是$A^{\mathsf T}A$的单位特征向量。
(3) 若非0特征根的特征向量$\mathbf{v}'s$相互正交，则对应的$\mathbf{u}'s$相互正交.
反之也对。

证明：若$\mathbf{v}$是$A^{\mathsf T}A$的特征根λ的单位特征向量, 则

$$A^{\mathsf T}A\mathbf{v} = \mathbf{v}\lambda \Rightarrow AA^{\mathsf T}(A\mathbf{v}) = (A\mathbf{v})\lambda,$$

所以$\mathbf{b} = A\mathbf{v}$是$AA^{\mathsf T}$的特征向量，因为$\mathbf{b}^{\mathsf T}\mathbf{b} = \mathbf{v}^{\mathsf T}A^{\mathsf T}A\mathbf{v} = \mathbf{v}^{\mathsf T}(\mathbf{v}\lambda) = \lambda,$

$\Rightarrow \mathbf{u} = \mathbf{b}/\sqrt{\lambda} = A\mathbf{v}/\sqrt{\lambda}$是$AA^{\mathsf T}$的单位特征向量.

同样，对于$AA^{\mathsf T}$的特征向量$\mathbf{u}$，$\mathbf{v} = A\mathbf{u}/\sqrt{\lambda}$是$A^{\mathsf T}A$的单位特征向量.

若$A^{\mathsf T}A$的非0特征根λ_1, λ_2对应的单位特征向量$\mathbf{v}_1, \mathbf{v}_2$正交($\lambda_1, \lambda_2$可能相同),

$\mathbf{u}_1 = A\mathbf{v}_1/\sqrt{\lambda_1}$,$\mathbf{u}_2 = A\mathbf{v}_2/\sqrt{\lambda_2}$,因为$A^{\mathsf T}A\mathbf{v}_2 = \mathbf{v}_2\lambda_2$

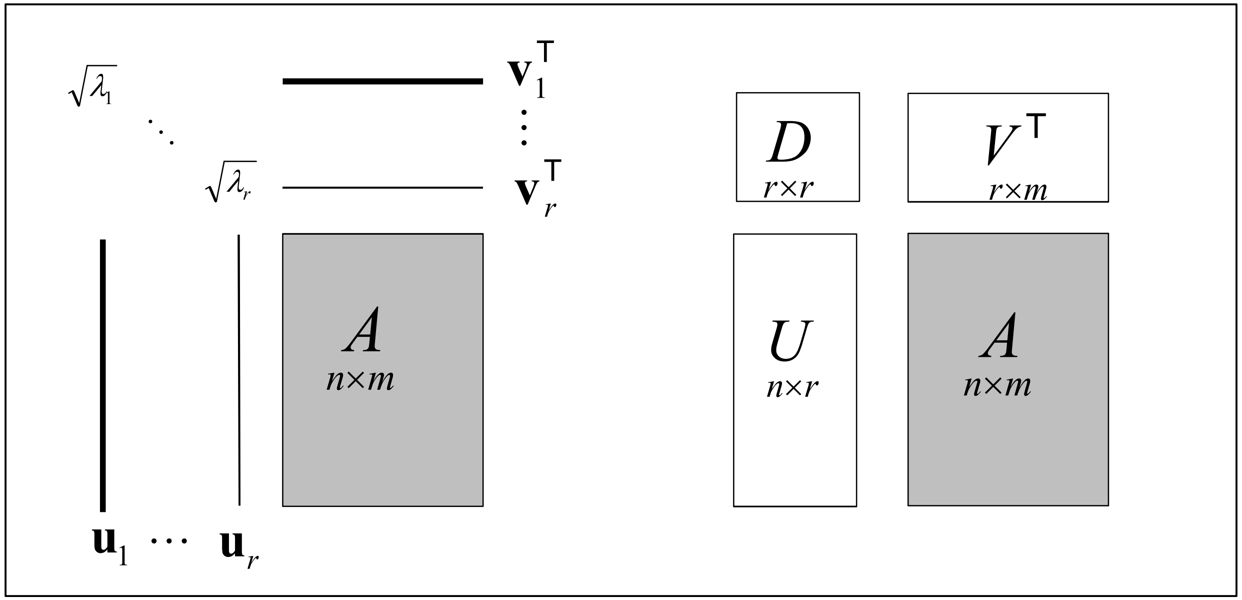
则$\mathbf{u}_1^{\mathsf T}\mathbf{u}_2 = \mathbf{v}_1^{\mathsf T}A^{\mathsf T}A\mathbf{v}_2/\sqrt{\lambda_1\lambda_2} = \mathbf{v}_1^{\mathsf T}\mathbf{v}_2\lambda_2/\sqrt{\lambda_1\lambda_2} = 0.$

奇异值分解
$A=UDV^{\mathrm{T}}$

定理6(SVD). 任一秩为r的$n\times m$矩阵A可表示为

$$A=U_{n\times r}D_{r\times r}V^{\mathrm{T}}{}_{r\times m}$$

其中$U^{\mathrm{T}}U=V^{\mathrm{T}}V=I_r$, $D=\mathrm{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_r})$的对角元称为奇异值,
其中$\lambda_1\ge\dots\ge\lambda_r>0$为$A^{\mathrm{T}}A$或$AA^{\mathrm{T}}$的$r$个正特征根， $U=(\mathbf{u}_1,\dots,\mathbf{u}_r)$的第$i$列$\mathbf{u}_i$为$AA^{\mathrm{T}}$的对应于特征根$\lambda_i$的特征向量， $V=(\mathbf{v}_1,\dots,\mathbf{v}_r)$的第$i$列$\mathbf{v}_i$为$A^{\mathrm{T}}A$的对应于特征根$\lambda_i$的特征向量。



证明：$\lambda_1 \geq \ldots \geq \lambda_r > 0$为$A^{\mathsf{T}}A$或$AA^{\mathsf{T}}$的$r$个正特征根.

由$A^{\mathsf{T}}A$的谱分解定理,我们可选择单位正交特 征向量$\mathbf{v}_1,\ldots,\mathbf{v}_r$,满足特征方程：

$$A^{\mathsf{T}}A\mathbf{v}_j = \mathbf{v}_j\lambda_j,\quad j = 1,\ldots.,r, \Leftrightarrow A^{\mathsf{T}}AV = VD^2，V = (\mathbf{v}_1,\ldots,\mathbf{v}_r)$$

两边同时左乘$A \Rightarrow AA^{\mathsf{T}}(AV) = (AV)D^2$,所以$AV$各列是$AA^{\mathsf{T}}$的特征向量。

下面把U的各列单位化，记$U = AVD^{-1} = (\mathbf{u}_1,\ldots,\mathbf{u}_r)$，则

$U^{\mathsf{T}}U = D^{-1}V^{\mathsf{T}}A^{\mathsf{T}}AVD^{-1} = I_r$, 且 $AA^{\mathsf{T}}U = UD^2$

所以U的各列是AA^{T}的相互正交的单位特征向量。

由$U = AVD^{-1} \Rightarrow \begin{cases} AV = UD \\ A^{\mathsf{T}}U = VD \end{cases} \Rightarrow AVV^{\mathsf{T}} = UDV^{\mathsf{T}}$

我们下面证明$A = AVV^{\mathsf{T}}$

注意：$V^{\mathsf{T}}V = I_r$并不蕴含$VV^{\mathsf{T}} = I_m$ 。

因为V是$C(A^{\mathsf{T}}A)$，也是$C(A^{\mathsf{T}})$的一组正交基$\Rightarrow$存在B使得$A^{\mathsf{T}} = VB$,

由$V^{\mathsf{T}}V = I_r \Rightarrow AVV^{\mathsf{T}} = B^{\mathsf{T}}V^{\mathsf{T}}VV^{\mathsf{T}} = B^{\mathsf{T}}V^{\mathsf{T}} = A$, 所以$A = UDV^{\mathsf{T}}$.

(事实上，$VV^{\mathsf{T}} = V(V^{\mathsf{T}}V)^{-1}V^{\mathsf{T}}$是$A$行空间的投影矩阵).

定理6*(SVD). 有时为了数学上的方便，将U,V扩展成正交方阵,
$\widetilde{U}=(U,*),\widetilde{V}=(V,*)$， $\widetilde{U}^{\mathsf{T}}\widetilde{U}=\widetilde{U}\widetilde{U}^{\mathsf{T}}=I_n,\widetilde{V}^{\mathsf{T}}\widetilde{V}=\widetilde{V}\widetilde{V}^{\mathsf{T}}=I_m$.
奇异值分解也表述为

$$A_{n\times m}=\widetilde{U}_{n\times n}\begin{pmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}_{n\times m}\widetilde{V}^{\mathsf{T}}{}_{m\times m}.$$

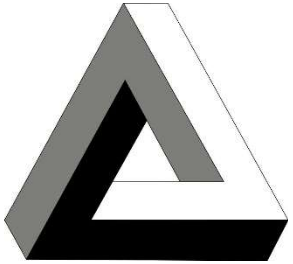
总结SVD： $A=UDV^{\mathsf{T}}$

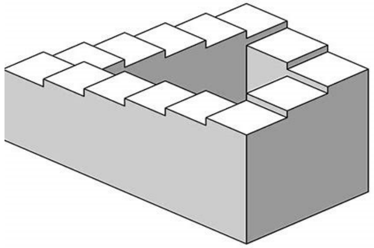
- 行特征(V是行空间$C(A^{\mathsf{T}})$的正交基)： $A^{\mathsf{T}}AV=VD^2$， $C(V)=C(A^{\mathsf{T}})$;
- 列特征(U是列空间$C(A)$的正交基)： $AA^{\mathsf{T}}U=UD^2$， $C(U)=C(A)$;
- 对偶 : $\begin{cases} AV=UD \\ A^{\mathsf{T}}U=VD \end{cases}$, $\begin{cases} A\mathbf{v}_j=\mathbf{u}_j\sqrt{\lambda_j} \\ A^{\mathsf{T}}\mathbf{u}_j=\mathbf{v}_j\sqrt{\lambda_j} \end{cases}$, $j=1,\ldots,r$
- 若A各列是中心化的,则UD的各列称为主成分 (第j列称为第j主成分，重要性或方差贡献为 λ_j)。V各列称为主成分方向。
- $A=UDV^{\mathsf{T}}=\sum_{j=1}^{r}\sqrt{\lambda_j}\mathbf{u}_j\mathbf{v}_j^{\mathsf{T}}$， 前$k$项 $\sum_{j=1}^{k}\sqrt{\lambda_j}\mathbf{u}_j\mathbf{v}_j^{\mathsf{T}}$是$A$的$k$阶(秩$k$)最佳逼近.

7.广义逆

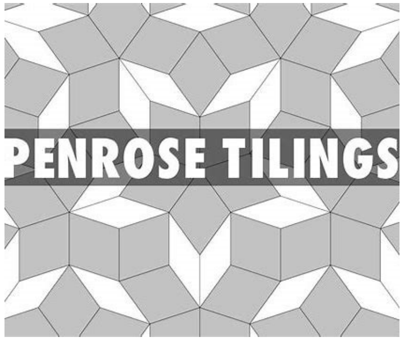
1955年，英国数学家Roger Penrose (彭罗斯，1931-) 发现, 通过使用广义逆可以将线性方程组通解简单地表示出来。2020年因其在黑洞，奇点的研究获得诺贝尔物理学奖。在娱乐数学方面，彭罗斯推广和发明了若干不可能图形，发现了除正三、四、六边形之外的其它形状地砖铺满地面但不重复的铺法。

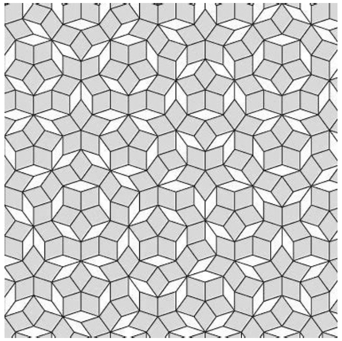
不可能的形状：彭罗斯三角和彭罗斯阶梯





彭罗斯地砖（几种图形填满地面，直线上永不重复）





广义逆参考：王松桂，线性模型引论，2004，科学出版社

我们知道，若$A\mathbf{x}=\mathbf{b}$有唯一解，解 $\mathbf{x}=A^{-1}\mathbf{b}$;

若$A\mathbf{x}=\mathbf{b}$解不唯一，甚至A不是方阵，能否将所有解表示为$\mathbf{x}=A^{inv}\mathbf{b}$?

A^{inv}是一类A的“逆”矩阵.

若$A\mathbf{x}=\mathbf{b}$有解，$\mathbf{b}\in C(A)\Rightarrow$ 存在$\mathbf{t}\in R^m$, $\mathbf{b}=A\mathbf{t}$.

若$\mathbf{x}=A^{inv}\mathbf{b}$是$A\mathbf{x}=\mathbf{b}$的解,则$A\mathbf{t}=\mathbf{b}=A(A^{inv}\mathbf{b})=A(A^{inv}A\mathbf{t})=AA^{inv}A\mathbf{t}$，

所以A^{inv}需要满足条件：$AA^{inv}A=A$，因此有下述定义：

广义逆A^-

任何 $A_{n\times m}$,若$X_{m\times n}$满足 $AXA=A$, X 称为A的广义逆,记为A^-

A^-存在性

设$A_{n\times m}=U_{n\times r}D_{r\times r}(V_{m\times r})^{\top}$是$A$的奇异值分解，则$VD^{-1}U^{\top}$是一个广义逆.

验证:因为$V^{\top}V=U^{\top}U=I_r$,所以$A(VD^{-1}U^{\top})A=UDV^{\top}VD^{-1}U^{\top}UDV^{\top}=UDV^{\top}=A$.

A^-的一般表达

命题1. 当A是满秩方阵的时候广义逆唯一，$A^- = A^{-1}$，否则不唯一，具体如下：

当$r = \mathrm{rank}(A) < \max(m,n)$时，若$A$有奇异值分解$A = \tilde{U}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\tilde{V}^{\mathsf{T}}$，

其中$\tilde{U}^{\mathsf{T}}\tilde{U} = \tilde{U}\tilde{U}^{\mathsf{T}} = I_n, \tilde{V}^{\mathsf{T}}\tilde{V} = \tilde{V}\tilde{V}^{\mathsf{T}} = I_m$，则任何广义逆具有形式：

$$A^- = \tilde{V}\begin{pmatrix} D^{-1} & * \\ * & * \end{pmatrix}\tilde{U}^{\mathsf{T}} \quad (*\text{表示任意}).$$

证：设X是任一广义逆，则

$AXA = A$

$$\Leftrightarrow \tilde{U}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\tilde{V}^{\mathsf{T}}X\tilde{U}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\tilde{V}^{\mathsf{T}} = \tilde{U}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\tilde{V}^{\mathsf{T}}$$

$$\Leftrightarrow \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{其中} Y = \tilde{V}^{\mathsf{T}}X\tilde{U} = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$$

则$Y_{11} = D^{-1}$，其它任意，所以$X = \tilde{V}Y\tilde{U}^{\mathsf{T}} = \tilde{V}\begin{pmatrix} D^{-1} & * \\ * & * \end{pmatrix}\tilde{U}^{\mathsf{T}}$。

Moore-Penrose 广义逆

任何 $A_{n\times m}$,若$X_{m\times n}$满足

$$AXA = A,\ \ XAX = X,\ \ (AX)^{\mathsf T} = AX,\ \ (XA)^{\mathsf T} = XA$$

X 称为A的Moore - Penrose广义逆,记为A^+.

命题2(A^+存在唯一). 若A有SVD分解 $A_{n\times m} = U_{n\times r}D_{r\times r}(V_{m\times r})^{\mathsf T}$,
其中U,V为列正交方阵: $U^{\mathsf T}U = I_r$, $V^{\mathsf T}V = I_r$, 则 $A^+ = VD^{-1}U^{\mathsf T}$.

证明:容易验证上述定义的A^+满足MP广义逆的四个条件。

A^+的唯一性证明如下:

若X,Y都是MP广义逆,下面证明$X = Y$.

$$X = XAX = X(AX)^{\mathsf T} = XX^{\mathsf T}A^{\mathsf T} = XX^{\mathsf T}(AYA)^{\mathsf T} = XX^{\mathsf T}A^{\mathsf T}Y^{\mathsf T}A^{\mathsf T}$$
$$= X(AX)^{\mathsf T}(AY)^{\mathsf T} = X(AX)(AY) = X(AXA)Y = XAY,$$

至此,我们把$X = XAX$中的第三项由X换作了Y。

类似的操作可把$XAX = XAY$中的第一项也换成Y:

$$X = XAY = (XA)^{\mathsf T}(YAY) = A^{\mathsf T}X^{\mathsf T}YAY = A^{\mathsf T}X^{\mathsf T}A^{\mathsf T}Y^{\mathsf T}Y = A^{\mathsf T}Y^{\mathsf T}Y$$
$$= (YA)^{\mathsf T}Y = YAY = Y.$$

线性方程解的一般形式

定理7. 若方程$A\mathbf{x}=\mathbf{b}$有解$(\mathbf{b}\neq\mathbf{0})$，则任何一个解都具有形式$\mathbf{x}=A^{-}\mathbf{b}$.

证明：假设$\mathbf{x}$是$A\mathbf{x}=\mathbf{b}$的任一解。对任何一个A的广义逆B,

$\mathbf{x}_0=B\mathbf{b}$是$A\mathbf{x}=\mathbf{b}$的解,所以$A(\mathbf{x}-\mathbf{x}_0)=0$, $\mathbf{v}=\mathbf{x}-\mathbf{x}_0\in N(A)$,

$$\mathbf{x}=\mathbf{x}_0+\mathbf{v}=B\mathbf{b}+\frac{\mathbf{v}\mathbf{b}^{\mathsf{T}}}{\|\mathbf{b}\|^2}\mathbf{b}=\left(B+\frac{\mathbf{v}\mathbf{b}^{\mathsf{T}}}{\|\mathbf{b}\|^2}\right)\mathbf{b}=C\mathbf{b}$$

其中$C=B+\frac{\mathbf{v}\mathbf{b}^{\mathsf{T}}}{\|\mathbf{b}\|^2}$，容易验证 $ACA=ABA+A\frac{\mathbf{v}\mathbf{b}^{\mathsf{T}}}{\|\mathbf{b}\|^2}A=ABA=A.$

所以$C=B+\frac{\mathbf{v}\mathbf{b}^{\mathsf{T}}}{\|\mathbf{b}\|^2}$是$A$的广义逆。

线性方程解的最优解

命题2. 若方程$A\mathbf{x}=\mathbf{b}$有解$(\mathbf{b}\neq\mathbf{0})$，则$\mathbf{x}=A^{+}\mathbf{b}$是所有解中长度最小的解。

证：设$\mathbf{x}$是$A\mathbf{x}=\mathbf{b}$的解,则$\mathbf{x}$可表示为$\mathbf{x}=A^{+}\mathbf{b}+\mathbf{v}$，

其中$A\mathbf{v}=0$，设A的SVD为$A=UDV^{\mathsf{T}},\Rightarrow UDV^{\mathsf{T}}\mathbf{v}=0\Rightarrow V^{\mathsf{T}}\mathbf{v}$

由$A^{+}=VD^{-1}U^{\mathsf{T}}\Rightarrow {A^{+}}^{\mathsf{T}}\mathbf{v}=0$

则$\|\mathbf{x}\|^2=\|A^{+}\mathbf{b}\|^2+\|\mathbf{v}\|^2\geq\|A^{+}\mathbf{b}\|^2$.

附录1：矩阵的数字特征及性质

秩rank$(A)=\dim(C(A))=\dim(C(A^{\mathrm{T}}))$:

$\mathrm{rank}(A^{\mathrm{T}})=\mathrm{rank}(A)=\mathrm{rank}(A^{\mathrm{T}}A)=\mathrm{rank}(AA^{\mathrm{T}})$

$\mathrm{rank}(A_{n\times m})\le\min(n,m)$

$\mathrm{rank}(A_{n\times p}B_{p\times m})\le\min(\mathrm{rank}(A)，\ \mathrm{rank}(B))$

$\mathrm{rank}(A_{n\times m}+B_{n\times m})\le\mathrm{rank}(A)+\mathrm{rank}(B)$

$\mathrm{rank}(CAB)=\mathrm{rank}(A)$, 如果$C,B$可逆.

$A_{n\times n}$可逆 $\Leftrightarrow \mathrm{rank}(A)=n$(满秩) $\Leftrightarrow |A|\neq 0$

设$A_{n\times n}$的所有特征根$\lambda_1,\ldots,\lambda_n$:

- $\det(A-\lambda I_n)=0$的n个解为A的所有特征根$\lambda_1,\ldots,\lambda_n$。
- $|A|=\det(A)=\prod_{k=1}^{n}\lambda_k$，$tr(A)=\sum_{k=1}^{n}\lambda_k$
- 对称阵的特征根为实数，不同特征根对应的特 征向量相互正交。
- 幂等阵的特征根为0或1
- 正交矩阵的实特征根为1或-1，虚数特征根模1。
- $C_{n\times p}B_{p\times n}$与$B_{p\times n}C_{n\times p}$有相同的非0特征根

迹trace：$\mathrm{tr}(A_{n\times n})=\sum_{i=1}^{n}a_{ii}$

$\mathrm{tr}(A^{\mathrm{T}})=\mathrm{tr}(A)$

$\mathrm{tr}(cA_{n\times n})=c\,\mathrm{tr}(A), c\in R$

$\mathrm{tr}(A_{n\times p}B_{p\times n})=\mathrm{tr}(BA)$

$\mathrm{tr}(A_{n\times n}+B_{n\times n})=\mathrm{tr}(A)+\mathrm{tr}(B)$

$\mathrm{tr}(A_{n\times n})=\sum\lambda_i$，$\lambda_i$，$i=1,..,n$为$A$的所有特征根

行列式determinant:

$\det(A_{n\times n})=|A|=\sum_{\tau}(-1)^{|\tau|}a_{1\tau(1)}\cdots a_{n\tau(n)}$，$\tau$为$\{1,2,\ldots,n\}$的所有置换，

其中若τ可由偶数次两两互换得到则$|\tau|=0$，否则$|\tau|=1$。

$|A^{\mathrm{T}}|=|A|$

$|cA_{n\times n}|=c^n|A|$, c常数

$|A_{n\times n}B_{n\times n}|=|A||B|$，特别$|A^{-1}|=|A|^{-1}$

$|A_{n\times n}|=\prod\lambda_i$

若A为正交矩阵，则$|A|=1$或-1

附录2：特殊方阵

定义　$A_{n\times n}=(a_{ij})$为方阵	特征根
对称阵 : $A^{\mathsf{T}}=A$	实数
对角阵 : $a_{ij}=0, i\neq j$	$a_{kk}, k=1,\dots,n$
正定矩阵$A>0$：A对称，对任何$\mathbf{x}\neq\mathbf{0}\in R^n, \mathbf{x}^{\mathsf{T}}A\mathbf{x}>0.$	正数
半正定矩阵$A\geq 0$：A对称，对任何$\mathbf{x}\in R^n, \mathbf{x}^{\mathsf{T}}A\mathbf{x}\geq 0.$	非负
正交矩阵(或标准正交矩阵)：$A^{\mathsf{T}}A=AA^{\mathsf{T}}=I_n$	模1的虚数 : $e^{i\theta}$
对称幂等阵/投影矩阵：$A^2=AA=A$	1或0

附录3：欧氏空间常用的线性变换：拉伸、旋转、镜像、投影

拉伸(stretch)、旋转(rotation)、镜像(reflection)、投影(projection)
参见Gilbert Strang: Linear algebra and its applications

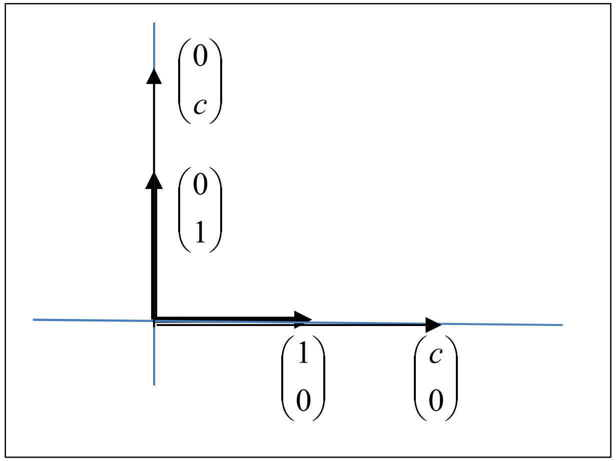
拉伸/对角阵

$$S=\begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_n \end{pmatrix}, S\mathbf{x}=\begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_n \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} c_1x_1 \\ \vdots \\ c_nx_n \end{pmatrix}$$

特列：

$S=cI_n$

$n=2$情形：$S=\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$将普通标准正交基$\mathbf{e}_1=\begin{pmatrix} 1 \\ 0 \end{pmatrix}\to\begin{pmatrix} c_1 \\ 0 \end{pmatrix}, \mathbf{e}_2=\begin{pmatrix} 0 \\ 1 \end{pmatrix}\to\begin{pmatrix} 0 \\ c_2 \end{pmatrix}$



$$\forall\mathbf{x},\ S\mathbf{x}=\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} c_1x_1 \\ c_2x_2 \end{pmatrix}$$

旋转/正交阵:
$\boldsymbol{O^T O = I_n}$

特列:

$\theta = 90^\circ$ 逆时针旋转

$O = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$

若n阶方阵O满足$O^\mathrm{T} O = I_n$,称为正交矩阵/正交旋转矩阵.

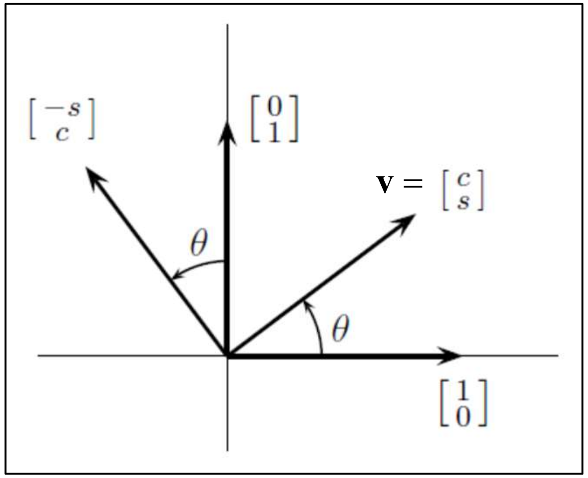
保持长度: 任何$\mathbf{x} \in R^n, \| O\mathbf{x} \|^2 = \mathbf{x}^\mathrm{T} O^\mathrm{T} O\mathbf{x} = \mathbf{x}^\mathrm{T}\mathbf{x} = \| \mathbf{x} \|^2$ 。

记$\mathbf{e}_i = (0,\ldots,1,\ldots0)$为普通标准正交基,$\boldsymbol{\omega}_i$为$O$的第$i$列

则 $\mathbf{e}_i \xrightarrow{O} \boldsymbol{\omega}_i \xrightarrow{O^\mathrm{T}} \mathbf{e}_i$, 即$O\mathbf{e}_i = \boldsymbol{\omega}_i, O^\mathrm{T}\boldsymbol{\omega}_i = O^\mathrm{T} O\mathbf{e}_i = \mathbf{e}_i$

$$n = 2\text{情形}: O = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}, c^2 + s^2 = 1, \ \text{通常写作} O = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$\text{旋转}\theta = \arccos(c), \text{它将普通标准正交基}\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \mathbf{v} = \begin{pmatrix} c \\ s \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} -s \\ c \end{pmatrix}.$$



$$\forall \mathbf{x}, \ O\mathbf{x} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} cx_1 - sx_2 \\ sx_1 + cx_2 \end{pmatrix}$$

方阵**O**的各列为标准正交基,则行向量也是

命题:若O是n阶方阵满足$O^{\mathrm{T}}O=I_n$,则$OO^{\mathrm{T}}=I_n$.

证明:记O的各列为$O=(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_n)$,$O^{\mathrm{T}}O=I_n$说明$\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_n$两两正交,它们一定线性无关,因而构成一组基。

对任何$\mathbf{x}$,存在$\mathbf{a}=(a_1,\ldots,a_n)^{\mathrm{T}}$使得$\mathbf{x}=O\mathbf{a}=\sum\boldsymbol{\omega}_i a_i$,

$$\Rightarrow OO^{\mathrm{T}}\mathbf{x}=OO^{\mathrm{T}}O\mathbf{a}=O(O^{\mathrm{T}}O)\mathbf{a}=O\mathbf{a}=\mathbf{x}\Rightarrow OO^{\mathrm{T}}=I_n.$$

$O^{\mathrm{T}}O=I_n \quad \Rightarrow \quad OO^{\mathrm{T}}O=O$,即($OO^{\mathrm{T}}$)$O=O$

上式说明(OO^{T})$(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_n)=(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_n)$,即$OO^{\mathrm{T}}\boldsymbol{\omega}_i=\boldsymbol{\omega}_i$,$i=1,\ldots,n$,

即标准正交基$\{\boldsymbol{\omega}_i, i=1,\ldots,\}$在线性变换$OO^{\mathrm{T}}$下保持不变,

所以OO^{T}必定是单位变换。

投影/对称幂等阵: $P^2=P$

特列：x轴投影

$P=\begin{pmatrix}1&0\\0&0\end{pmatrix}$

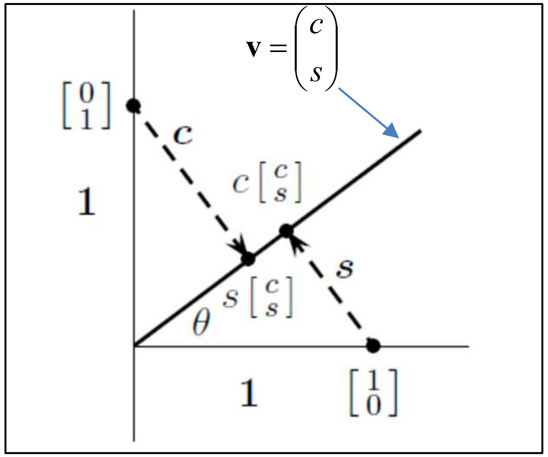
对称幂等矩阵/投影矩阵P: $P^{\mathrm{T}}=P, P^2=P$

$n=2$情形：对任何 $\mathbf{v}=\begin{pmatrix}c\\s\end{pmatrix}=\begin{pmatrix}\cos(\theta)\\\sin(\theta)\end{pmatrix}$,

投影阵$P=\mathbf{v}\mathbf{v}^{\mathrm{T}}=\begin{pmatrix}c^2&cs\\cs&s^2\end{pmatrix}=\begin{pmatrix}\cos^2(\theta)&\cos(\theta)\sin(\theta)\\\cos(\theta)\sin(\theta)&\sin^2(\theta)\end{pmatrix}$,

它将普通标准正交基 $\mathbf{e}_1=\begin{pmatrix}1\\0\end{pmatrix}\to\begin{pmatrix}c^2\\cs\end{pmatrix}=c\mathbf{v}$，$\mathbf{e}_2=\begin{pmatrix}0\\1\end{pmatrix}\to\begin{pmatrix}cs\\s^2\end{pmatrix}=s\mathbf{v}$，

即投影到$\mathbf{v}=\begin{pmatrix}c\\s\end{pmatrix}=\begin{pmatrix}\cos(\theta)\\\sin(\theta)\end{pmatrix}$张成的直线上



任何$\mathbf{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}$投影到$\mathbf{v}=\begin{pmatrix}c\\s\end{pmatrix}$上：

$$P\mathbf{x}=\begin{pmatrix}c(cx_1+sx_2)\\s(cx_1+sx_2)\end{pmatrix}=\begin{pmatrix}c\\s\end{pmatrix}(cx_1+sx_2)=\mathbf{v}(cx_1+sx_2)$$

反射或镜像/单位根矩阵: $\boldsymbol{R}^2 = \boldsymbol{I}_n$

反射(*reflection*)变换矩阵R：$R = I_n - 2P, P$为投影矩阵.

$\theta = 45^\circ$ 反射

$$R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

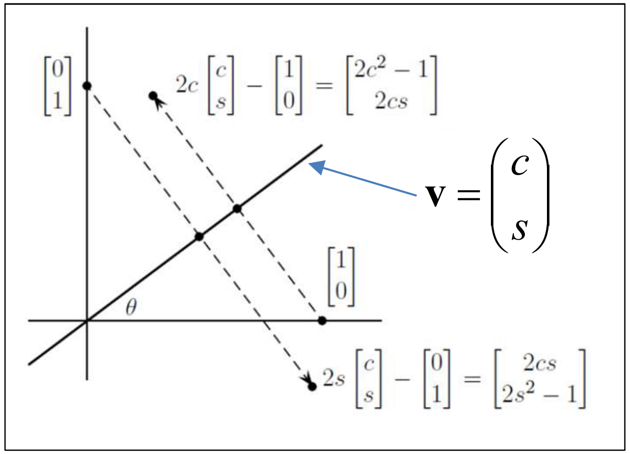
$\theta = 90^\circ$ 反射

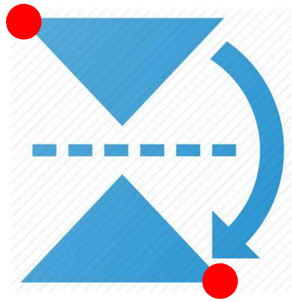
$$R = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

例. $n = 2$, $R = \begin{pmatrix} 2c^2 - 1 & 2cs \\ 2cs & 2s^2 - 1 \end{pmatrix} = 2\mathbf{v}\mathbf{v}^\mathsf{T} - I_n, \mathbf{v} = \begin{pmatrix} c \\ s \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$

$$R = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix},$$

它将普通标准正交基 $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} \sin(2\theta) \\ \cos(2\theta) \end{pmatrix}$.





Householder 反射变换：$H = I_n - 2\mathbf{v}\mathbf{v}^\mathsf{T}$, $\mathbf{v}^\mathsf{T}\mathbf{v} = 1$.

附录4：半正定矩阵的平方根

定义：

- 半正定：假设 A 对称，若对任何向量 $\mathbf{x}\neq\mathbf{0}$，$\mathbf{x}^{\top}A\mathbf{x}\geq 0$，则称 A 是半正定的（记作 $A\geq 0$）。
- 记号 / 偏序：若对称方阵 A，B 满足 $A-B\geq 0$（半正定），则记 $A\geq B$。
- 半正定矩阵的正平方根：若 $A_{n\times n}\geq 0$，则使得 $B^2=A$ 的半正定矩阵 B 称为 A 的正平方根，B 记作 $A^{1/2}$.

命题.

(1) 若 $A_{n\times n}\geq 0$，则对任何矩阵 $C_{n\times k}$，$CAC^{\top}\geq 0$.

（2）$A\geq 0$ 的平方根唯一。具体地，假设谱分解为 $A=O\Lambda O^{\top},\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$，$O$ 为正交矩阵，记 $\Delta=\mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n})$，则 $B=O\Delta O^{\top}$ 是 A 的<u>唯一</u>的正平方根。

(3) 若 A 对称，则 $A\leq I_n\Leftrightarrow\lambda(A)\leq 1$

(4) 若 $A\geq B>0$，则 $0<A^{-1}\leq B^{-1}$

(5) 若 $A^2\geq B^2>0$，则 $A\geq B>0$。

但反之未必成立（即 $A\geq B>0$ 未必蕴含 $A^2\geq B^2>0$）

证：（1）对任何 $\mathbf{x} \in R^k, \mathbf{x}^\top CAC^\top \mathbf{x} = \mathbf{y}^\top C\mathbf{y} \geq 0, \mathbf{y} = C^\top \mathbf{x} \in R^n$

（2）$A = O\Lambda O^\top$,由$B = O\Delta O^\top \Rightarrow B^2 = BB = O\Delta O^\top O\Delta O^\top = O\Delta^2 O^\top = O\Lambda O^\top = A.$

唯一性，假设 $C \geq 0$，$D \geq 0$满足$A = C^2 = D^2$, A的特征根为$\lambda_1,...,\lambda_n$,则C, D

的特征根为$\sqrt{\lambda_1},...,\sqrt{\lambda_n}$，由谱分解定理，存在 正交矩阵$P, Q$使得

$$C = P\Delta P^\top, D = Q\Delta Q^\top$$

由$C^2 = D^2 \Rightarrow P\Lambda P^\top = Q\Lambda Q^\top \Rightarrow \Lambda P^\top Q = P^\top Q\Lambda$, 令$R = P^\top Q = (r_{ij})$,

$\Lambda R = R\Lambda \Leftrightarrow$ 任何i, j，$\lambda_i r_{ij} = \lambda_j r_{ij} \Leftrightarrow \sqrt{\lambda_i} r_{ij} = \sqrt{\lambda_j} r_{ij} \Leftrightarrow \Delta R = R\Delta$

$\Leftrightarrow \Delta P^\top Q = P^\top Q\Delta \Leftrightarrow P\Delta P^\top = Q\Delta Q^\top$，所以$C = D = B.$

(3)设谱分解 $A = O\Lambda O^\top, \Lambda = diag(\lambda_1,...,\lambda_n)$,

$A \leq I_n \Rightarrow \Lambda = O^\top AO \leq O^\top O = I_n \Rightarrow \lambda_i \leq 1.$

反之，$\lambda_i \leq 1 \Rightarrow \Lambda \leq I_n \Rightarrow AO\Lambda O^\top \leq OO^\top = I_n$。(也可由引理1(特征根的表示）证明 ）。

(4)$A \geq B > 0 \Rightarrow B^{-1/2}AB^{-1/2} \geq I_n \Rightarrow \lambda(B^{-1/2}AB^{-1/2}) \geq 1$

$\Rightarrow \lambda(B^{1/2}A^{-1}B^{1/2}) \leq 1 \Rightarrow B^{1/2}A^{-1}B^{1/2} \leq I_n \Rightarrow A^{-1} \leq B^{-1}$

（5）参见倪国熙《常用的 矩阵理论和方法》，上 海科技出版社