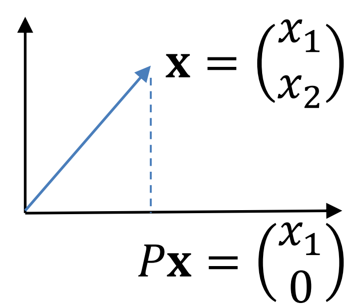
第九讲 欧氏空间中的投影

2023.11.17



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

$$P\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$

主要内容

- 向量空间
  - 函数空间: 欧氏R^n, 实函数L^2, 随机变量$L^2(\mathcal{F})$
- 内积向量空间的投影
- 欧氏空间：矩阵预备知识
  - 四个基本空间：列空间、行空间、核空间
  - 常用线性变换方阵
  - 奇异值分解（SVD）、广义逆 A^-
- 欧氏空间的投影

上次课

Recap：列空间与行空间

矩阵$A_{n\times m}$的列空间

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

$A_{n\times m}$的行空间即A^{T}的列空间：

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

定理3：对任一实数矩阵A，$\mathrm{rank}(A^{\mathsf{T}})=\dim(\mathrm{C}(A^{\mathsf{T}}))=\dim(C(A))=\mathrm{rank}(A)$.

定理4. $C(A)=C(AA^{\mathsf{T}})$，$C(A^{\mathsf{T}})=C(A^{\mathsf{T}}A)$，
特别地, $\mathrm{rank}(AA^{\mathsf{T}})=\mathrm{rank}(A)$

定理5. $C(A^{\mathsf{T}})\oplus N(A)=R^m$，$C(A)\oplus N(A^{\mathsf{T}})=R^n$

$$\mathrm{rank}(A_{n\times m})=m-\mathrm{nullity}(A)=n-\mathrm{nullity}(A^{\mathsf{T}})$$

常用线性变换方阵：拉伸、旋转、镜像、投影

参见Gilbert Strang: Linear algebra and its applications （p146）

理解矩阵变换 - 观察矩阵将标准坐标轴变换到什么向量：

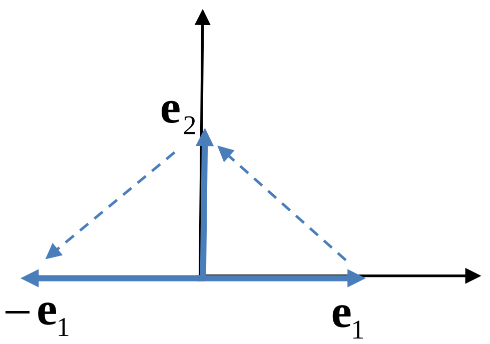
记$A_{n\times m}=(\mathbf{a}_1,\ldots,\mathbf{a}_m):R^m\to R^n$，矩阵$A$的列向量为$\mathbf{a}_1,\ldots,\mathbf{a}_m\in R^n$

R^m中，记$\mathbf{e}_k=$第k个坐标轴$=I_m$的第k列$=(0,\ldots,1,\ldots 0)^{\mathrm{T}}$，则

$A\mathbf{e}_k=\mathbf{a}_k$， $k=1,\ldots,m$， 即$AI_m=A(\mathbf{e}_1,\ldots,\mathbf{e}_m)=A=(\mathbf{a}_1,\ldots,\mathbf{a}_m)$

例如$A=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$，它将标准正交基$\mathbf{e}_1=\begin{pmatrix}1\\0\end{pmatrix}$变换为$\mathbf{e}_2=\begin{pmatrix}0\\1\end{pmatrix}$，

将$\mathbf{e}_2=\begin{pmatrix}0\\1\end{pmatrix}$变换为$-\mathbf{e}_1=\begin{pmatrix}-1\\0\end{pmatrix}$， A是逆时针90°旋转变换。



下面考虑常用的几个2阶变换方阵

$\begin{pmatrix}1&0\\0&2\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&0\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$
拉伸	90°逆时针旋转	x轴投影	45°反射

拉伸 $\leftrightarrow$ 数乘

$S=\begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}$: $S\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} c_1x_1 \\ c_2x_2 \end{pmatrix}$

矩阵的数乘$cA \overset{\text{定义为}}{=} (ca_{ij})$，应理解为拉伸：$diag(c)A=\begin{pmatrix} c & & \\ & \ddots & \\ & & c \end{pmatrix}A$

注1: cA定义是明确的，但不符合矩阵乘积的要求，有时会引起混乱。

比如若$c=\mathbf{x}^{\mathrm{T}}\mathbf{x}\in R$，$\mathbf{x}$是向量，则$cA=\mathbf{x}^{\mathrm{T}}\mathbf{x}A$是三个矩阵/向量的乘积，按照结合律$cA=\mathbf{x}^{\mathrm{T}}\mathbf{x}A=\mathbf{x}^{\mathrm{T}}(\mathbf{x}A)$，括号中出现了问题。

注2: 向量$\mathbf{a}\in R^n$与实数$c\in R$的数乘，最好写成$\mathbf{a}c$，而不是$c\mathbf{a}$。

这是因为$\mathbf{a}c$是$n\times 1$向量与1×1的数的乘积，符合矩阵向量乘积法则，而$c\mathbf{a}$为1×1的数与$n\times 1$向量的乘积，不符合乘积法则。

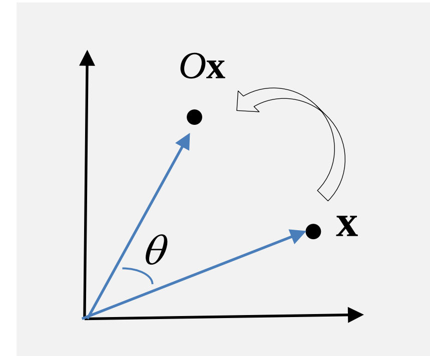
比如若$c=\mathbf{x}^{\mathrm{T}}\mathbf{x}\in R$，则常数在前的写法$c\mathbf{a}=\mathbf{x}^{\mathrm{T}}\mathbf{x}\mathbf{a}=\mathbf{x}^{\mathrm{T}}(\mathbf{x}\mathbf{a})$出现了矛盾。

旋转 $\leftrightarrow e^{i\theta}$

$$O=\begin{pmatrix} c & -s \\ s & c \end{pmatrix}=\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},\quad c^2+s^2=1,$$

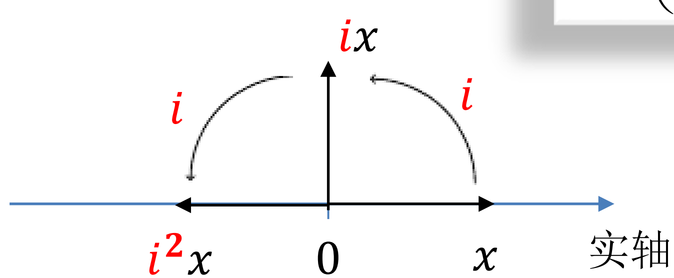
$$O \overset{对应}{\leftrightarrow} e^{i\theta}=\cos(\theta)+i\sin(\theta)$$

$$OO^{\mathrm{T}}=I \overset{对应}{\leftrightarrow} e^{i\theta}(\overline{e^{i\theta}})=1$$



矩阵O和复数$e^{i\theta}$都实现了平面上的旋转，但后者只需初等代数运算

特别：当$\theta=90°$时(90°逆时针旋转)，$\Omega=\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \overset{对应}{\leftrightarrow} i$：

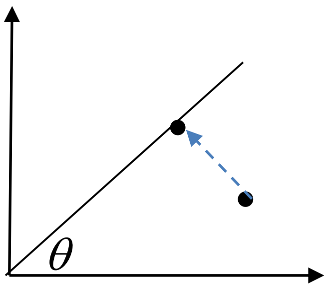


	共轭	模长平方	平方	形式记号	变换
Ω	$\Omega^{\mathrm{T}}=-\Omega$	$\Omega\Omega^{\mathrm{T}}=I_2$	$\Omega^2=-I_2$	$\Omega=(-I_2)^{1/2}$	$\Omega\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} -b \\ a \end{pmatrix}$
i	$\bar{i}=-i$	$i\bar{i}=1$	$i^2=-1$	$i=\sqrt{-1}$	$i(a+bi)=ia-b$

投影$P^2 = P$
$P \leftrightarrow 0 \; or \; 1$

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

$P\mathbf{e}_1 = \begin{pmatrix} c^2 \\ cs \end{pmatrix} = c\mathbf{v}$，$P\mathbf{e}_2 = \begin{pmatrix} cs \\ s^2 \end{pmatrix} = s\mathbf{v}$



$\theta = 0$时，$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$，向横轴投影：$P\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$

- 一般$n \times n$投影矩阵（对称幂等）：$P^{\mathsf{T}} = P, P^2 = P.$
- P特征根为0或1，存在正交矩阵O,使得

$$P = O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^{\mathsf{T}}, r = rank(P)$$

- 二次型可简化为平方和：

$$\mathbf{x}^{\mathsf{T}}P\mathbf{x} = \| P\mathbf{x} \|^2 = \mathbf{x}^{\mathsf{T}}O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^{\mathsf{T}}\mathbf{x} = \mathbf{y}^{\mathsf{T}}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\mathbf{y} = y_1^2 + \ldots + y_r^2$$

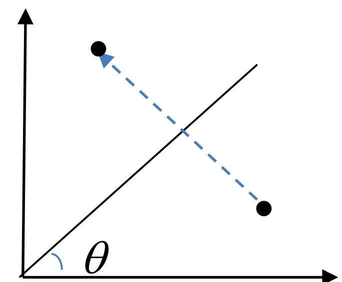
其中$\mathbf{y} = O^{\mathsf{T}}\mathbf{x}$。

反射$R^2 = I_n$
$R \leftrightarrow \pm 1$

$$\mathbf{v} = \begin{pmatrix} c \\ s \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix},$$ 反射变换（reflection）

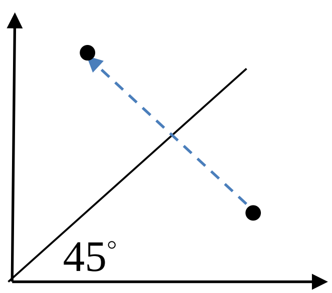
$$R = 2\mathbf{v}\mathbf{v}^{\mathrm{T}} - I_n = \begin{pmatrix} 2c^2 - 1 & 2cs \\ 2cs & 2s^2 - 1 \end{pmatrix} = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix},$$

$$R\mathbf{e}_1 = \begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}, \quad R\mathbf{e}_2 = \begin{pmatrix} \sin(2\theta) \\ \cos(2\theta) \end{pmatrix}.$$



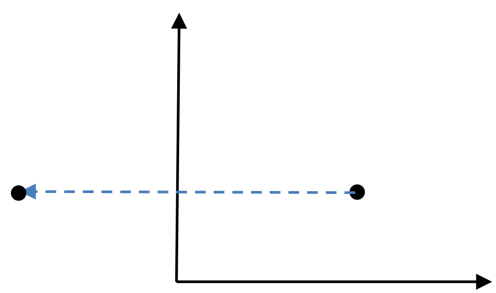
$\theta = 45^\circ$ 反射

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



$\theta = 90^\circ$ 反射

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



反射变换矩阵R：$R = I_n - 2P$或$2P - I_n$，其中P为投影矩阵.

$R^2 = I_n$,我们可称之为“平方根”矩阵。

正交矩阵
$O^{\top}O = I_n$

若n阶方阵O满足$O^{\top}O = I_n$(不必要求$OO^{\top} = I_n$),称为正交矩阵.
正交 = 旋转 + 反射

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

证明:$O^{\top}O = I_n$,两边同时左乘O,得

$$OO^{\top}O = O, \quad (OO^{\top})O = O$$

记O的各列为$O = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)$,上式说明

$$(OO^{\top})(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n) = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n), \text{ 即 } OO^{\top}\boldsymbol{\omega}_i = \boldsymbol{\omega}_i, \quad i = 1, \ldots, n,$$

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

所以$OO^{\top}$必定是单位变换,验证如下:

任何$\mathbf{x}$可用正交基$\{\boldsymbol{\omega}_i, \ i = 1, \ldots,\}$表示:$\mathbf{x} = O\mathbf{a} = \sum \boldsymbol{\omega}_i a_i$

$\Rightarrow OO^{\top}\mathbf{x} = OO^{\top}O\mathbf{a} = O(O^{\top}O)\mathbf{a} = O\mathbf{a} = \mathbf{x} \Rightarrow OO^{\top} = I_n$。

行列空间的特征刻画

特征根特征向量

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

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

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$\mathbf{v}\in R^n$)。

特征向量：线性变换的不变量，列空间的刻画。

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，则特征方程$A\mathbf{x}_i=\mathbf{x}_i\lambda_i,1\le i\le n\Leftrightarrow$

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

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：

对称矩阵的谱分解

定理6. 任一$n\times n$对称阵A可表示为

$$A=V\Lambda V^{\mathrm{T}}$$

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

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。

当特征根、特征向量$\mathbf{v}$为实数时，$A\mathbf{v}$与$\mathbf{v}$方向重合。复特征根情形，如何理解复特征向量？

复特征根

假设$A_{n\times n}$是一个实矩阵，则下述事实等价(其中$a,b\in R,\mathbf{u},\mathbf{v}\in R^n$):

(1) $A(\mathbf{u}+\mathbf{v}i)=(\mathbf{u}+\mathbf{v}i)(a+bi)$, 即$a+bi$是$A$的复特征根,特征向量$\mathbf{u}+\mathbf{v}i$;

(2) $A(\mathbf{u}-\mathbf{v}i)=(\mathbf{u}-\mathbf{v}i)(a-bi)$,即$a-bi$是$A$的特征根，特征向量$\mathbf{u}-\mathbf{v}i$;

(3) $A(\mathbf{u},\mathbf{v})=(\mathbf{u},\mathbf{v})\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$。

证明：$A\mathbf{x}=\mathbf{x}\lambda \Leftrightarrow A(\mathbf{u}+\mathbf{v}i)=(\mathbf{u}+\mathbf{v}i)(a+bi)=a\mathbf{u}-b\mathbf{v}+(b\mathbf{u}+a\mathbf{v})i$

$\Leftrightarrow A\mathbf{u}=a\mathbf{u}-b\mathbf{v}, A\mathbf{v}=b\mathbf{u}+a\mathbf{v}$

$\Leftrightarrow A(\mathbf{u}-\mathbf{v}i)=a\mathbf{u}-b\mathbf{v}-(b\mathbf{u}+a\mathbf{v})i=(\mathbf{u}-\mathbf{v}i)(a-bi)$，即$A\bar{\mathbf{x}}=\bar{\mathbf{x}}\bar{\lambda}$。

注：$C(\mathbf{u},\mathbf{v})$是A变换的2维不变子空间：对任何$\mathbf{w}\in C(\mathbf{u},\mathbf{v})$，设其坐标为$x_1$，$x_2$，

即$\mathbf{w}=\mathbf{u}x_1+\mathbf{v}x_2=(\mathbf{u},\mathbf{v})\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$，则$A\mathbf{w}=A(\mathbf{u},\mathbf{v})\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=(\mathbf{u},\mathbf{v})\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$,

其坐标为$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\propto\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$的正交旋转。

例5. $\Omega = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$是平面上的反转（逆时针）90°的变换矩阵:

$$R\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}, \quad R^2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}$$

Ω的特征根为 $i, -i$, 特征向量$\begin{pmatrix} i \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ i \end{pmatrix}$, 特征方程

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} i \\ 1 \end{pmatrix}=i\begin{pmatrix} i \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ i \end{pmatrix}=-i\begin{pmatrix} 1 \\ i \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, B = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

特征方程 $\Leftrightarrow$ 实数空间中$\Omega\mathbf{u} = -\mathbf{v}, \Omega\mathbf{v} = \mathbf{u} \Leftrightarrow \Omega(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, \mathbf{v})B = (-\mathbf{v}, \mathbf{u})$,

$\Rightarrow \Omega = (-\mathbf{v}, \mathbf{u})(\mathbf{u}, \mathbf{v})^{-1} = (-\mathbf{v}, \mathbf{u})(\mathbf{u}, \mathbf{v})$

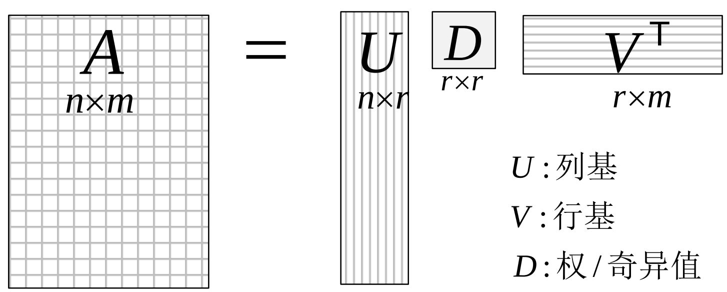
即 $\underset{\text{反转}90^\circ}{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}} = \underset{90^\circ\text{反射}}{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}} \underset{45^\circ\text{反射}}{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}$,

SVD: 非方阵的特征

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

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。

奇异值分解SVD

定理7(SVD). 任一秩为r的$n\times m$矩阵A可表示为

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

其中$U^{\mathsf{T}}U=V^{\mathsf{T}}V=I_r$, $D=\mathrm{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_r})$的对角元称为奇异值,

其中$\lambda_1\ge\dots\ge\lambda_r>0$为$A^{\mathsf{T}}A$或$AA^{\mathsf{T}}$的$r$个正特征根，$U_{n\times r}=(\mathbf{u}_1,\dots,\mathbf{u}_r)$的第$i$列$\mathbf{u}_i$为$AA^{\mathsf{T}}$的对应于特征根$\lambda_i$的特征向量，$V_{m\times r}=(\mathbf{v}_1,\dots,\mathbf{v}_r)$的第$i$列$\mathbf{v}_i$为$A^{\mathsf{T}}A$的对应于特征根$\lambda_i$的特征向量。

SVD的发现本质上利用了简单事实:

若$\mathbf{v}$是$A^{\mathsf{T}}A$的特征向量，则$A\mathbf{v}$是AA^{T}的特征向量:

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

证明：$\lambda_1\ge\dots\ge\lambda_r>0$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$\mathbf{v}_1,\dots,\mathbf{v}_r$,满足特征方程:

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

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

$Y^{\top}Y = V^{\top}A^{\top}AV = V^{\top}VD^2 = D^2$,所以$Y$的各列相互正交，但不是单位长度，
只需令$U = YD^{-1} = AVD^{-1}$,则$U^{\top}U = D^{-1}V^{\top}A^{\top}AVD^{-1} = I_r$,
即U的各列是$AA^{\top}$的单位长且相互正交的特征向量。

由$U = AVD^{-1} \Rightarrow AV = UD \Rightarrow AVV^{\top} = UDV^{\top}$
只需说明$A = AVV^{\top}$。

注意：V是$m \times r$矩阵，
$V^{\top}V = I_r$并不蕴含$VV^{\top} = I_m$,
事实上，$VV^{\top}$的秩为$r \le m$

因为V是$C(A^{\top}A)$，也是$C(A^{\top})$的一组正交基$\Rightarrow$存在B使得$A^{\top} = VB$,
由$V^{\top}V = I_r \Rightarrow AVV^{\top} = B^{\top}V^{\top}VV^{\top} = B^{\top}V^{\top} = A$, 所以$A = UDV^{\top}$。

$A^{\top} = (\mathbf{a}_1,\dots,\mathbf{a}_n)$，因为$V$的各列是$C(A^{\top})$的一组基，存在$r \times 1$系数$\mathbf{b}_i$使得$\mathbf{a}_i = V\mathbf{b}_i$
$\Rightarrow A^{\top} = (V\mathbf{b}_1,\dots,V\mathbf{b}_n) = V(\mathbf{b}_1,\dots,\mathbf{b}_n) = VB$

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

定理7*(*SVD*). 有时为了数学上的方便，将U,V扩展成正交方阵，

$\tilde U=(U,*),\tilde V=(V,*)$，$\tilde U^{\mathsf T}\tilde U=\tilde U\tilde U^{\mathsf T}=I_n,\tilde V^{\mathsf T}\tilde V=\tilde V\tilde V^{\mathsf T}=I_m$.

奇异值分解也表述为

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

总结SVD：$A=UDV^{\mathsf T}$

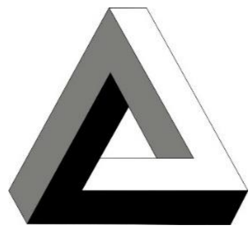
- 行特征(V是行空间$C(A^{\mathsf T})$的正交基)：$A^{\mathsf T}AV=VD^2$，$C(V)=C(A^{\mathsf T})$;
- 列特征(U是列空间$C(A)$的正交基)：$AA^{\mathsf T}U=UD^2$，$C(U)=C(A\ )$;
- 对偶: $\begin{cases} A\,V=UD\\ A^{\mathsf T}U=VD\end{cases}$, $\begin{cases} A\,\mathbf v_j=\mathbf u_j\sqrt{\lambda_j}\\ A^{\mathsf T}\mathbf u_j=\mathbf v_j\sqrt{\lambda_j}\end{cases}$, $j=1,\dots,r$
- 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.

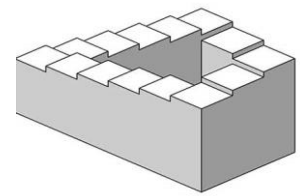
广义逆

广义逆参考：王松桂，线性模型引论，2004，科学出版社

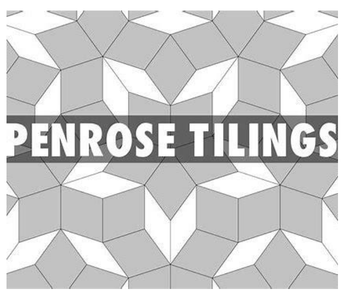
1955年，英国数学家Roger Penrose (彭罗斯，1931-) 发现, 通过使用广义逆可以将线性方程组通解简单地表示出来。2020年因其在黑洞，奇点的研究获得诺贝尔物理学奖。在娱乐数学方面，彭罗斯推广和发明了若干不可能图形，发现了除正三、四、六边形之外的其它形状地砖铺满地面但不重复的铺法。

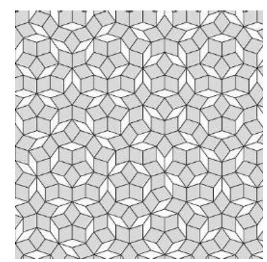
不可能的形状：彭罗斯三角和彭罗斯阶梯





彭罗斯地砖（几种图形填满地面，直线上永不重复）





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，即$\mathbf{b}\in C(A)\Rightarrow$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^-

定义：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^-

存在性：

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

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

A^-的一般表达

命题7. 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$r = \operatorname{rank}(A) < \max(m, n)$时，若$A$有奇异值分解$A = \tilde{U}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\tilde{V}^{\top}$,

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

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

证：设X是任一广义逆，则

$AXA = A$

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

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

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

Moore-Penrose 广义逆

任何 $A_{n\times m}$,若$X_{m\times n}$满足

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

X 称为A的Moore - Penrose广义逆,记为A^{+}.

命题8(A^{+}存在唯一). 若A有SVD分解 $A_{n\times m} = U_{n\times r} D_{r\times r} (V_{m\times r})^{\mathrm{T}}$,
其中U,V为列正交方阵:$U^{\mathrm{T}}U = I_r$,$V^{\mathrm{T}}V = I_r$,则 $A^{+} = VD^{-1}U^{\mathrm{T}}$.

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$X = Y$.

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

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:

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

线性方程解的一般形式

定理8. 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$A(\mathbf{x}-\mathbf{x}_0)=0$, $\mathbf{v}=\mathbf{x}-\mathbf{x}_0\in N(A)$,

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

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

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

线性方程解的最优解

命题9. 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$A=UDV^{\mathrm{T}}$,$\Rightarrow UDV^{\mathrm{T}}\mathbf{v}=0\Rightarrow V^{\mathrm{T}}\mathbf{v}$

由$A^{+}=VD^{-1}U^{\mathrm{T}}\Rightarrow {A^{+}}^{\mathrm{T}}\mathbf{v}=0\Rightarrow\|\mathbf{x}\|^2=\|A^{+}\mathbf{b}\|^2+\|\mathbf{v}\|^2\geq\|A^{+}\mathbf{b}\|^2$.

欧氏空间中的正交投影

内积： $\mathbf{u},\mathbf{v}\in R^n,\ \langle\mathbf{u},\mathbf{v}\rangle=\mathbf{u}^{\mathrm{T}}\mathbf{v}=\sum u_i\,v_i$

模长： $\|\mathbf{u}\|=\sqrt{\langle\mathbf{u},\mathbf{u}\rangle}=\sqrt{\sum u_i^2}$

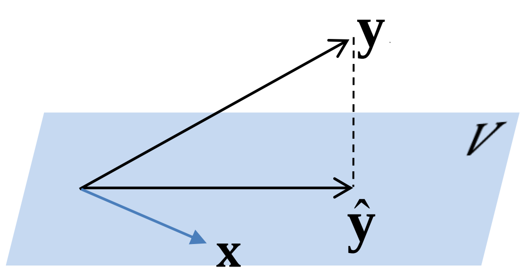
其它形式内积也可以，比如，$\langle\mathbf{u},\mathbf{v}\rangle=\mathbf{u}^{\mathrm{T}}G\mathbf{v},G>0$

投影

设$\mathbf{y}\neq\mathbf{0}\in R^n$，$\mathbf{y}$在线性子空间$V\subset R^n$空间上的正交投影$\hat{\mathbf{y}}$满足：

(1) $\hat{\mathbf{y}}\in V$

(2) $\mathbf{y}-\hat{\mathbf{y}}\perp V$，即对$\forall\mathbf{x}\in V$，$<\mathbf{x},\mathbf{y}-\hat{\mathbf{y}}>=0\Leftrightarrow<\mathbf{x},\mathbf{y}>=<\mathbf{x},\hat{\mathbf{y}}>$.



记$\mathbf{y}^{\perp}=\mathbf{y}-\hat{\mathbf{y}}$，

正交分解： $\mathbf{y}=\hat{\mathbf{y}}+\mathbf{y}^{\perp},\ \hat{\mathbf{y}}\perp\mathbf{y}^{\perp}$;

平方和分解： $\|\mathbf{y}\|^2=\|\hat{\mathbf{y}}\|^2+\|\mathbf{y}^{\perp}\|^2$

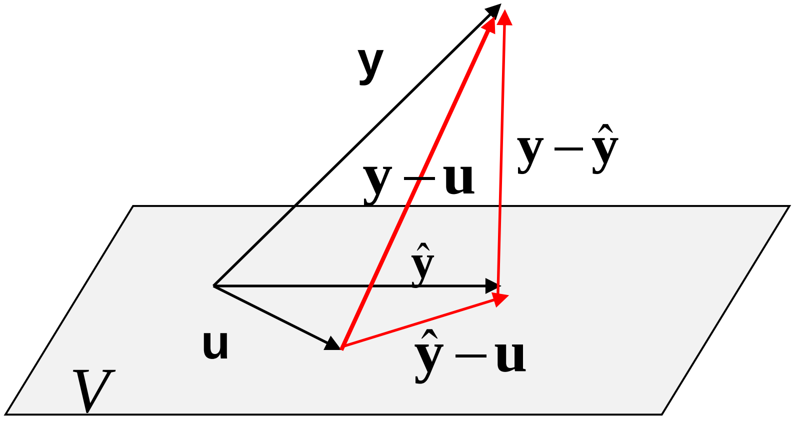
投影参考： James H Stapleton (1995) Linear statistical models. Wiley
（下载： /books/5.pdf）

投影的最优性：最小二乘

定理2（最小二乘,欧氏情形）：$\mathbf{y}\in R^n$，记$\hat{\mathbf{y}}=P(\mathbf{y}\,|\,V)$为$\mathbf{y}$在子空间$V\subset R^n$上的投影，则$\|\mathbf{y}-\hat{\mathbf{y}}\|^2=\min\limits_{\mathbf{u}\in V}\|\mathbf{y}-\mathbf{u}\|^2$

证：对任何$\mathbf{u}\in V$，因为$\hat{\mathbf{y}}-\mathbf{u}\in V$，故$\mathbf{y}-\hat{\mathbf{y}}\perp\hat{\mathbf{y}}-\mathbf{u}$.

所以$\|\mathbf{y}-\mathbf{u}\|^2=\|\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\mathbf{u}\|^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\mathbf{u}\|^2\geq\|\mathbf{y}-\hat{\mathbf{y}}\|^2$.



投影矩阵

命题10. 设$V=C(A_{n\times m})\subset R^n$，$\mathbf{y}\in R^n$，则$\mathbf{y}$在$V$上的投影为

$$\hat{\mathbf{y}}=P_A\mathbf{y}=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}\mathbf{y},$$

其中A对应的投影矩阵为$P_A=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}$ (n阶方阵).

证明1: 因为$\mathbf{y}^{\perp}=\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-A\boldsymbol{\beta}\perp C(A)$,所以对任何$\mathbf{t}\in R^m$，$\mathbf{u}=A\mathbf{t}\in C(A)$,

$$0=<\mathbf{u},\mathbf{y}^{\perp}>=<A\mathbf{t},\mathbf{y}^{\perp}>\overset{\text{adjoint}}{=}<\mathbf{t},A^{\mathrm{T}}\mathbf{y}^{\perp}>$$

即$A^{\mathrm{T}}\mathbf{y}^{\perp}\in R^m$与任何$\mathbf{t}\in R^m$都正交，所以$A^{\mathrm{T}}\mathbf{y}^{\perp}=\mathbf{0}$，即正则方程:

$$A^{\mathrm{T}}\mathbf{y}=A^{\mathrm{T}}A\boldsymbol{\beta}$$

因为$A^{\mathrm{T}}\mathbf{y}\in C(A^{\mathrm{T}})=C(A^{\mathrm{T}}A)$,方程有解,所有解$\hat{\boldsymbol{\beta}}=(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}\mathbf{y}$

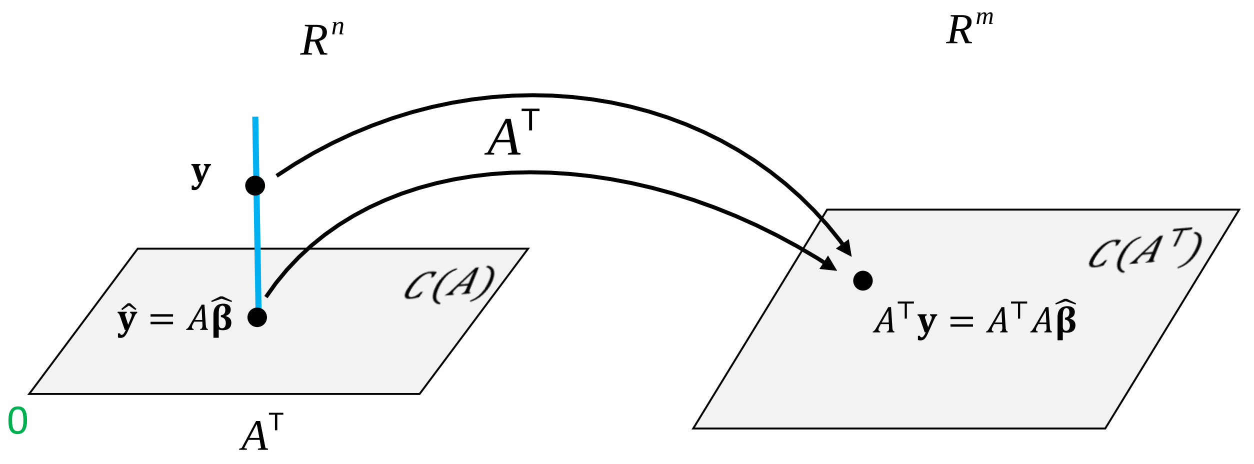
$\Rightarrow\hat{\mathbf{y}}=A\hat{\boldsymbol{\beta}}=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}\mathbf{y}$，所以$P_A=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}$.

证明2: 因为$\hat{\mathbf{y}}\in V$,存在$\boldsymbol{\beta}_{m\times 1}$,使得$\hat{\mathbf{y}}=A\boldsymbol{\beta}$。

因为$\mathbf{y}^{\perp}=\mathbf{y}-\hat{\mathbf{y}}\perp C(A)$,特别地，$\mathbf{y}^{\perp}$与$A$的每列都正交

$$\Rightarrow\quad A^{\mathrm{T}}(\mathbf{y}-\hat{\mathbf{y}})=A^{\mathrm{T}}(\mathbf{y}-A\boldsymbol{\beta})=\mathbf{0}\iff A^{\mathrm{T}}A\boldsymbol{\beta}=A^{\mathrm{T}}\mathbf{y}$$

$$\Rightarrow\hat{\boldsymbol{\beta}}=(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}\mathbf{y}\Rightarrow\hat{\mathbf{y}}=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}\mathbf{y},P_A=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}$$



R^n中解近似方程$\mathbf{y} \approx A\widehat{\boldsymbol{\beta}}$ $\xrightarrow{A^\top}$ R^m中解方程$A^\top\mathbf{y} = A^\top A\widehat{\boldsymbol{\beta}}$

（求解$\mathbf{y}$的最佳逼近$\mathbf{y} = A\widehat{\boldsymbol{\beta}}$）

$A^\top\mathbf{y}^\perp = A^\top(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{0}$:

蓝线$\mathbf{y} + C(A)^\perp$上的点在$C(A)$上的投影都相同

$\Leftrightarrow$ $A^\top$将它们映射为R^m中的同一点$A^\top\hat{\mathbf{y}}$。

例6. 设$\mathbf{x} \neq \mathbf{0} \in R^n$, $\mathbf{x}$张成的空间$U = C(\mathbf{x}) = \{\mathbf{x}c \mid c \in R\}$，投影矩阵

$$P_{\mathbf{x}} = \mathbf{x}(\mathbf{x}^\top\mathbf{x})^{-1}\mathbf{x}^\top$$

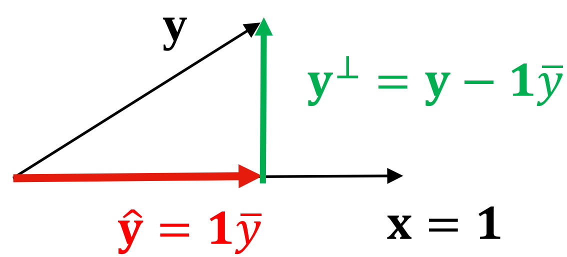
$\mathbf{y} \in R^n$在U上的投影 $\hat{\mathbf{y}} = P_{\mathbf{x}}\mathbf{y} = \mathbf{x}(\mathbf{x}^\top\mathbf{x})^{-1}\mathbf{x}^\top\mathbf{y} = \mathbf{x}\lambda, \lambda = \mathbf{x}^\top\mathbf{y}/\mathbf{x}^\top\mathbf{x}$。

特别地，如果$\mathbf{x} = \mathbf{1} = (1,\ldots,1)^\top \in R^n, P_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$,

$$\hat{\mathbf{y}} = P_{\mathbf{1}}\mathbf{y} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top\mathbf{y} = \mathbf{1}\bar{y},$$

$$\mathbf{y}^\perp = \mathbf{y} - P_{\mathbf{1}}\mathbf{y} = \mathbf{y} - \mathbf{1}\bar{y} = (y_1 - \bar{y},\ldots,y_n - \bar{y})^\top,$$

$\mathbf{y}^\perp$称为$\mathbf{y}$的中心化向量。



中心化向量

$$\mathbf{y}^\perp = (I_n - P_{\mathbf{1}})\mathbf{y} = P_{\mathbf{1}^\perp}\mathbf{y}$$

为$\mathbf{y}$在$\mathbf{1}$的正交补空间上的投影

投影矩阵的性质

命题11. 设A是$n\times m$矩阵，则

(1) $P_A=A(AA^{\mathrm T})^{-}A^{\mathrm T}$唯一，与广义逆$(AA^{\mathrm T})^{-}$的具体选择无关。

(2) 投影矩阵等价于对称幂等阵：$P_A=A(A^{\mathrm T}A)^{-}A^{\mathrm T}$是对称幂等矩阵，反之，任一对称幂等阵是投影阵。

(3) 任一投影矩阵$P_{n\times n}$有谱分解$P=O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^{\mathrm T}$，其中$r=rank(P)$，

$OO^{\mathrm T}=O^{\mathrm T}O=I_n$，

(4) $P_AA=A$.

(5) P_A是$C(A)$对应的投影阵，I_n-P_A是$C(A)^{\perp}$对应的投影阵，两者正交：

$$P_A(I_n-P_A)=0$$

(6) 线性子空间V对应的投影矩阵与空间的基的选取无关。

(7) 若$A=(\mathbf{a}_1,\ldots,\mathbf{a}_m)$,即$A^{\mathrm T}A$为对角阵，则$P_A=\sum_{k=1}^{m}P_{\mathbf{a}_k}$.

(8) 按列划分$A=(A_1,A_2)$,若$A_1^{\mathrm T}A_2=0$,则 $P_A=P_{A_1}+P_{A_2}$.

多数结论是第8讲命题6的重述

证明:(1) 这是命题6的特殊情形，因为投影唯一，投影阵P_A必定唯一,
与广义逆$(A^{\mathsf T}A)^-$的选择无关。但这里用矩阵方法再次证明：

因为$C(A^{\mathsf T})=C(A^{\mathsf T}A)$,所以存在$B$使得$A^{\mathsf T}=A^{\mathsf T}AB$,
所以$P_A=A(A^{\mathsf T}A)^-A^{\mathsf T}=B^{\mathsf T}A^{\mathsf T}A(A^{\mathsf T}A)^-A^{\mathsf T}AB=B^{\mathsf T}A^{\mathsf T}AB$与$(A^{\mathsf T}A)^-$无关，
且$P_A=B^{\mathsf T}A^{\mathsf T}AB$是对称的。

（2）$A^{\mathsf T}=A^{\mathsf T}AB\Rightarrow P_A{}^2=A(A^{\mathsf T}A)^-A^{\mathsf T}A(A^{\mathsf T}A)^-A^{\mathsf T}$
$=A(A^{\mathsf T}A)^-A^{\mathsf T}A(A^{\mathsf T}A)^-A^{\mathsf T}AB=A(A^{\mathsf T}A)^-A^{\mathsf T}AB=A(A^{\mathsf T}A)^-A^{\mathsf T}=P_A$。

反之，设P是$n\times n$对称幂等矩阵，$r=rank(P)$,则有谱分解$P=O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^{\mathsf T}$,

其中$OO^{\mathsf T}=O^{\mathsf T}O=I_n$,划分$O=(O_1,O_2)$,则$O_1{}^{\mathsf T}O_1=I_r$, $P=O_1O_1{}^{\mathsf T}=O_1\left(O_1{}^{\mathsf T}O_1\right)^{-1}O_1{}^{\mathsf T}$,
所以P具有投影矩阵的形式，所以任何对称幂等矩阵是投影阵。
（3）$-(7)$略。

命题12. 若$A_{n\times m}$有奇异值分解$A=U_{n\times r}D_{r\times r}\left(V_{m\times r}\right)^{\top}$,其中$U^{\top}U=V^{\top}V=I_r$,

则$P_A=A(A^{\top}A)^{-}A^{\top}=UU^{\top}$.

证明：$C(U)=C(A)$,所以$P_A=P_U=U(U^{\top}U)^{-1}U^{\top}=UU^{\top}$.

注：命题12提供了投影阵更容易理解的形式：

设U的各列为$U=(\mathbf{u}_1,...,\mathbf{u}_r)$,则任何$\mathbf{y}\in R^n$在$C(A)$上的投影

$$P_A\mathbf{y}=UU^{\top}\mathbf{y}=\mathbf{u}_1(\mathbf{u}_1{}^{\top}\mathbf{y})+...+\mathbf{u}_r(\mathbf{u}_r{}^{\top}\mathbf{y})$$

正是我们对正交投影的理解方式：$\mathbf{u}_1,...,\mathbf{u}_r$为基（坐标轴），

$\mathbf{u}_1{}^{\top}\mathbf{y},...,\mathbf{u}_r{}^{\top}\mathbf{y}$ 为$\mathbf{y}$在这些坐标轴上的投影坐标。