# Linear Algebra

**--Taught By CCC**

# 三、向量空间及四个子空间及延申

## （一）、向量空间及子空间的定义

对于一个向量空间而言，需要满足两个最基本的运算，分别是向量加法与数乘，分别为：

$$\begin{bmatrix}a_1\\a_2\end{bmatrix}+\begin{bmatrix}b_1\\b_2\end{bmatrix}=\begin{bmatrix}a_1+b_1\\a_2+b_2\end{bmatrix}$$

$$c\begin{bmatrix}a_1\\a_2\end{bmatrix}=\begin{bmatrix}ca_1\\ca_2\end{bmatrix}$$

除此之外，向量空间需要保证其运算的封闭性，也就是说，要保证向量空间中的两个向量在完成了以上两个运算过后的结果还在向量空间中。为此，向量空间还需要满足8个性质：

（1）、交换律：

$$\alpha+\beta=\beta+\alpha$$

（2）、结合律：

$$\alpha+\beta+\gamma=(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$$

（3）、对于向量空间中存在唯一的零向量，使得：

$$\mathbf{x}+\mathbf{0}=\mathbf{x}$$

（4）、对于向量空间中任意的向量x，存在唯一的-x，使得：

$$\mathbf{x}+(-\mathbf{x})=\mathbf{0}$$

（5）、对于任意的标量c，和向量x与y，有：

$$c(x+y)=cx+cy$$

（6）、对于任意的标量a和b，与向量x，有：

$$（a+b）x=ax+bx$$

（7）、对于任意的标量a，b与向量x，有：

$$abx=a(bx)=b(ax)$$

（8）、对于向量空间中的任意向量，有：

$$1\cdot x=x$$

最小的向量空间是{0}，即只包含一个零向量。

对于每一个向量空间而言，均存在向量子空间。由于向量子空间是向量空间的子集，因此，如果一个集合是向量空间，那么它的子空间一定满足上述八条性质。所以，如果要验证一个集合是否是子空间，只需要验证向量加法和标量乘法这两种运算即可。

## （二）、四个典型的子空间

在本课程中，讨论的都是有限维的、元素为实数（最后会拓展到复数）的向量空间。因此，会从矩阵的角度，来考察这个向量空间的四个子空间--列空间、零空间、行空间和左零空间。

主要还是从$Ax=b$的角度来看待子空间这一问题。

## 1、列空间

我们从矩阵的角度来看列空间。

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

所谓列空间$C(A)$就是若干个列向量进行线性组合后得到的子空间。在$m\times n$矩阵中，列空间就是$n$个列向量张成的空间。因为列向量的维数最高是$m$，所以说，列空间是$\mathbb{R}^m$的子空间。

如果$Ax=b$要有解，那么，$b$一定是矩阵A列空间中的一个向量，否则，它就一定无解。

## 2、零空间

要知道什么是零空间，首先要引入$Ax=0$这一式子。所谓零空间，就是能够使得$Ax=0$的解的集合。

$x$的维数跟矩阵$A$的列数有关，所以说，零空间是$\mathbb{R}^n$的子空间。

在用高斯消去法解$Ax=0$的过程中，会出RRE(Row Reduced Echelon Form)下阶梯矩阵，会出现$r$个主元以及$n-r$个自由元，这$n-r$个自由元可以用$r$个主元来表示。这一视角具有莫大的意义，由于$n-r$个自由元都可以用$r$个主元线性表示，那么，列空间可以只需要考虑$r$个主元列即可，而不需要去考虑剩下的$n-r$个自由元列。

下面是快速得到零空间解系的方法：

对于理想的情况而言，通过高斯消去法，可以将矩阵化为这个形式：

$$\begin{bmatrix} & \\ I & F\\ 0\dots0 & 0\dots0\end{bmatrix}$$

其中，$I$是$r\times r$的，而$F$是$r\times(n-r)$。可以将解矩阵写成以下形式：

$$\begin{bmatrix}-F\\I\end{bmatrix}$$

如果说主元在矩阵中并不能写成一个单位矩阵的形式，只需要将其中的某两行进行互换即可。

**题外话：关于$Ax=b$的问题**

在熟悉了上面两个子空间以后，可以回来再思考$Ax=b$这个问题。

经过高斯消去法以后，可以将$x$表示成$x=x_p+x_n$的形式，其中，$x_p$是特解，$x_n$是零空间的一组基。而$Ax=b$的解一定能够写成$x=x_p+x_n$的形式。

考虑一个特殊的$m\times n$的矩阵，如果说这个矩阵满足$m=n=r$的话，则该方程的$x_n=0,x_p=A^{-1}b$，则$x=x_n+x_p=A^{-1}b$。

$$A^{-1}\begin{bmatrix}A & b\end{bmatrix}=>\begin{bmatrix}I & A^{-1}b\end{bmatrix}$$

对于另一种$m\times n$的矩阵，如果这个矩阵是$m>n$的，那存在两种情况：

（1）$m>n=r$(列满秩)。那么这个$Ax=b$是否有解，取决于最后的$m-r$行是否会全部被消成$0$。如果全被消成0了，则有唯一解。反之，没有解。$C(A)$是$r$维的，$N(A)$是$\{0\}$。这种情况有三种特征：1）所有列都是pivot column；2）没有自由元和special solutions；3）$N(A)$是$\{0\}$

（2）$m>n>r$。同上面的情况，没全消成0，那么直接无解；反之有无穷解，$x=x_p+x_n$。

还有一种情况，就是$m<n$这种情况，分类讨论一下就是：

（1）n > m = r(行满秩)。一定有解，并且一定有$m-r$个free variable。所以说$C(A)$是$\mathbb{R}^m$,$N(A)$的维度是$m-r$。

（2）$n>m>r$。同上，不一定有解，主要看最后几行能不能被消成全0。

**向量的独立性、基和维数**

**定义**：如果$x_1\mathbf{v}_1+x_2\mathbf{v}_2+\ldots+x_n\mathbf{v}_n=0$只存在唯一解$x_1=x_2=x_3=\ldots=x_n=0$，则称为$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$线性无关；反之，则称这组向量线性相关。

为了确定一组向量是否是线性无关的，可以使用Gauss Elimination，如果说$m=n=r$或者$m>n=r$时，则说明这组向量是线性无关的；或者换一个角度考虑，如果说一组向量，他们的零空间为$\{0\}$时，就一定是线性无关的。

上面用于求解零空间向量的方法能够保证那些个向量全是线性无关的。

所以说，可以把零空间以及线性无关这两个概念给联系起来。进一步来看，只要你能够找到一个非零的解，那就一定说明存在无数组系数，使得向量的线性组合为0。

当$n>m$的时候，这组向量一定是线性相关的

## 3、行空间

定义：由矩阵行向量的线性组合构成的空间就是行空间，由于行向量的维数是$n$，所以说，行空间是$\mathbb{R}^n$的子空间，记作$C(A^T)$，也可以视作$A^T$的列空间。

——新概念：基（Basis）

**定义**：对于向量空间而言，一组基是满足以下两条性质的一系列向量：

（1）、这组向量是线性无关的

（2）、这组向量能够张成向量空间

向量空间中的每一个向量都能够写成一组基的线性组合。

一个简单的**例子**：

$$\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\text{能 够 张 成 }\mathbb{R}^2$$

另一个简单的**例子**：

$$\begin{bmatrix}1\\0\\0\\\vdots\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\\vdots\\0\\0\end{bmatrix},\cdots,\begin{bmatrix}0\\0\\0\\\vdots\\0\\1\end{bmatrix}\text{能 够 张 成 }\mathbb{R}^n$$

**性质**：对于任意$n$个线性无关的$n$维向量是$\mathbb{R}^n$的一组基。

**性质**：对于一组向量$v_1,v_2,\ldots,v_n$而言，如果它们要构成一组基，那么就得要求他们是一个非奇异/可逆矩阵的列向量。

**注意**：$\mathbb{R}^n$有无穷多组基。

**性质**：对于向量空间中的任意一个向量，在任意一组基下，有且仅有一种方式（线性组合）来表示这个向量。

**证明**：

假设在向量空间的同一组基$\{v_1,v_2,\ldots,v_n\}$下，存在两种不同的方式（线性组合）来表示同一个向量，记作$\beta=a_1v_1+a_2v_2+\ldots+a_nv_n=b_1v_1+b_2v_2+\ldots b_nv_n$。

那么，有

$$\begin{aligned}\beta-\beta&=(a_1v_1+a_2v_2+\ldots+a_nv_n)-(b_1v_1+b_2v_2+\ldots b_nv_n)\\&=(a_1-b_1)v_1+(a_2-b_2)v_2+\ldots+(a_n-b_n)v_n\\&=0\end{aligned}$$

由于$\{v_1,v_2,\ldots,v_n\}$是一组基，即这组向量是线性无关的，这也就意味着不存在一组非零系数使得这组向量的线性组合为0。进而可以得出$a_1-b_1=a_2-b_2=\ldots=a_n-b_n=0$，即$a_1=b_1,a_2=b_2,\ldots,a_n=b_n$。故在同一组基下，不存在两种不同的方式来表示同一个向量，即在同一基下，每个向量的表示方法唯一。□

**——另一个新概念：维度(Dimension)**

**性质**：若$\mathbb{V} = \{v_1, v_2, \ldots, v_n\}$和$\mathbb{W} = \{w_1, w_2, \ldots, w_m\}$是同一个向量空间的两组基，那么有$n = m$。

**证明**：

不失一般性，假设$n > m$。由于基的性质要求这两组向量是线性无关的，因此对于向量组$\mathbb{V}$和$\mathbb{W}$而言，他们的秩分别就是$n$和$m$。由于向量空间中的每一个向量都可以用一组基来唯一表示，那么这两组基中的每一个向量也一定能用另一组基中的向量唯一表示。由于$rank(\mathbb{V}) > rank(\mathbb{W})$，那么，$\mathbb{V}$中一定存在部分向量无法用$\mathbb{W}$中的向量线性表示。也就是说，$\mathbb{W}$中的向量无法线性表示向量空间中的所有向量，这也就导出了矛盾。

因此，如果两组向量同为某一向量空间的一组基，那么他俩的向量个数一定是相等的，也就是说$|\mathbb{V}| = |\mathbb{W}|, n = m$。$\square$

**定义**：向量空间的**维数**就是向量空间的基包含的向量个数。

一个简单的例子：

$$
\begin{gathered}
dim(\mathbb{R}^2) = 2 \\
dim(\mathbb{R}^n) = n \\
dim(2 \times 2\ matrix) = 4 \\
dim(n \times n\ matrix) = n^2 \\
dim(n \times n\ diagonal\ matrix) = n
\end{gathered}
$$

### 4、左零空间

**定义**：$A^T$的零空间就是$A$的左零空间。

$$
\begin{aligned}
N(A^T) &= \{y : A^T y = 0\} \\
&= \{y : y^T A = 0^T\}
\end{aligned}
$$

经过高斯消元后，矩阵会变成RRE Form，也就是上$r$行会变成带有主元的行，只需要令$y$的左$r$

个元素为0，剩下$m-r$个元素分配为：$(1, 0, \ldots 0), (0, 1, 0, \ldots 0), \ldots, (0, 0, \ldots, 1)$即可。

当然，$R$与$A$的左零空间是不一样的，所以接下来要求矩阵$A$的左零空间：

$$EA = R$$

R的最下面$m-r$行都是0，也就是说，E的下面$m-r$行分别是左零空间的$m-r$个基。（因为$E$是可逆的，所以说下面$m-r$行都是线性无关的）

**四大空间的维数**

首先我们有$m \times n$的矩阵$A$一个，现在来考察他的四大空间维数。

行空间为$C(A^T)$是$\mathbb{R}^n$的子空间

列空间为$C(A)$是$\mathbb{R}^m$的子空间

零空间为$N(A)$是$\mathbb{R}^n$的子空间

左零空间为$N(A^T)$是$\mathbb{R}^m$的子空间

**注意**：两组向量$A, B$可以互相线性表示 $\iff$ A的向量可以用B表示，B的向量可以用A表示

RRE Form的行空间和A的行空间一样，且维数一致，是矩阵的$rank$

RRE Form的列空间与A的列空间维数一样，但并不能表示同一个空间$rank$

RRE Form的零空间和A的零空间一致，且维数一样，是$n - rank(A)$

RRE Form的左零空间和A的左零空间不一致，但维数一样，是$m - rank(A)$

## （三）、正交性

### 1、线性代数基本定理第一部分

列空间和行空间的维数均为$r$，零空间和左零空间的维数分别为$m-r$和$n-r$。

接下来是本章正篇：

**点积（内积）**：$<\underline{v},\underline{w}>=\underline{v}\cdot\underline{w}=\underline{v}^T\underline{w}=\begin{bmatrix}v_1 & v_2 & \cdots & v_n\end{bmatrix}\begin{bmatrix}w_1\\w_2\\\vdots\\w_n\end{bmatrix}=v_1w_1+v_2w_2+\cdots+v_nw_n$

另外，一个向量自己同自己做内积就是：$\underline{v}^T\underline{v}=v_1^2+v_2^2+\ldots+v_n^2=||\underline{v}||^2$

而一个向量的长度则为：$||\underline{v}||=(\underline{v}^T\underline{v})^{1/2}$

**定义**：如果两个向量$\underline{v}$和$\underline{w}$的内积$<\underline{v},\underline{w}>=0$，则称这两个向量正交。

**性质**：$<\underline{v},\underline{w}>=0 \iff ||\underline{v}||^2+||\underline{w}||^2=||\underline{v}+\underline{w}||^2$

**证明**：

$$\begin{aligned}||\underline{v}+\underline{w}||^2&=<\underline{v}+\underline{w},\underline{v}+\underline{w}>\\&=(\underline{v}+\underline{w})^T(\underline{v}+\underline{w})\\&=\underline{v}^T\underline{v}+\underline{v}^T\underline{w}+\underline{w}^T\underline{v}+\underline{w}^T\underline{w}\{<\underline{v},\underline{w}>=0\}\\&=||\underline{v}||^2+||\underline{w}||^2\square\end{aligned}$$

**定义**：两个子空间$\mathbb{V}$和$\mathbb{W}$正交，对任意$v\in\mathbb{V},w\in\mathbb{W}$，均有$v^Tw=0$。

从内积的视角看矩阵乘法：

$$Ax=0=>\begin{bmatrix}row_1\\row_2\\\vdots\\row_m\end{bmatrix}\begin{bmatrix}\underline{x}\end{bmatrix}=0$$

$$A^Ty=0=>\begin{bmatrix}column_1^T\\column_2^T\\\vdots\\column_m^T\end{bmatrix}\begin{bmatrix}\underline{y}\end{bmatrix}=0$$

也就是说，每一个行向量和零空间中的向量做内积，得到的都是0。因此，可以认为**行空间和零空间正交**，记为：$\mathbf{C(A^T)\perp N(A)}$，称$C(A^T)$和$N(A)$是$\mathbb{R}^n$的正交子空间。

类似的，$A^Ty=0$，可以得到**列空间和左零空间正交**，记为：$\mathbf{C(A)\perp N(A^T)}$，称$C(A)$和$N(A^T)$是$\mathbb{R}^m$的正交子空间。

**定义**：向量子空间$\mathbb{V}$的正交补$\mathbb{V}^\perp$包含了每一个正交于$\mathbb{V}$的向量，用集合可以描述为$\mathbb{V}^\perp=\{v^\perp|\forall v\in\mathbb{V},v\perp v^\perp\}$

**性质**：$V^\perp$也是一个向量子空间

**性质**(Almost definition)：

$$(C(A^T))^\perp=N(A)$$
$$(C(A))^\perp=N(A^T)$$

**性质**：

$$(N(A))^\perp=C(A^T)$$

**证明**：

要证明这个性质，只需要证明$<\underline{v},\underline{w}>=0,\forall v\in N(A),w\in C(A^T)$

假设$\exists\underline{w}\notin C(A^T)$，则我们可以把行向量$\underline{w}$加到A里面去，记作：

$$A^{'}=\begin{bmatrix}A\\\underline{v}\end{bmatrix}$$

而$N(A^{'})=N(A)$，然而，$dim(C(A^T))=r$，并且，$dim(C(A^{'})^T)=r+1$，然而$dim(C(A^{'})^T)+dim(N(A))=n+1$，与题设矛盾，因此$(N(A))^\perp=C(A^T)$。$\square$

下面！是重要的时刻！

## 2、线性代数基本定理第二部分

$$
\begin{aligned}
(C(A^T))^{\perp} &= N(A)(in\ \mathbb{R}^m) \\
(C(A))^{\perp} &= N(A^T)(in\ \mathbb{R}^n) \\
(N(A))^{\perp} &= C(A^T)(in\ \mathbb{R}^m) \\
(N(A^T))^{\perp} &= C(A)(in\ \mathbb{R}^n)
\end{aligned}
$$

**性质**：$v_1, v_2, \ldots, v_n$是$\mathbb{R}^n$中线性无关的一组向量，当且仅当这组向量张成$\mathbb{R}^n$。

**证明**：令

$$
A_{n\times n} = \begin{bmatrix} \underline{v_1} & \underline{v_2} & \cdots \underline{v_n} \end{bmatrix}
$$

'⇒': 如果$v_1, v_2, \ldots, v_n$是$\mathbb{R}^n$中线性无关的一组向量，那么根据定义，这组向量就是$C(A)$的一组基，则$rank(A) = n$。因此，$A$是一个可逆矩阵。则对于$Ax = b$而言，每给出一个$b \in \mathbb{R}^n$，都能够得到解。则，$C(A)$是$\mathbb{R}^n$，因此，$v_1, v_2, \ldots, v_n$张成了$\mathbb{R}^n$。

'⇐'：如果说$v_1, v_2, \ldots, v_n$张成了$\mathbb{R}^n$，则对于$b \in \mathbb{R}^n$，方程$Ax = b$均有解。那么，可以得到$C(A) = \mathbb{R}^n$，也因此，$v_1, v_2, \ldots, v_n$为线性无关的一组解。□

**性质**：n个线性无关的n维向量可以张成$\mathbb{R}^n$。可以张成$\mathbb{R}^n$的一组向量是线性无关的。

**性质**：如果说$V \perp W$，那么$V \bigcap W = \{\underline{0}\}$。

**证明**：假设$V \bigcap W \neq \{\underline{0}\}$，也就是说，$\underline{\alpha} \neq \underline{0}, \exists \underline{\alpha} \in V \wedge \underline{\alpha} \in V$。根据定义，一定会有$\alpha \perp \alpha$，即$< \alpha, \alpha >= 0$。然而，$< \underline{\alpha}, \underline{\alpha} >= ||\underline{\alpha}||^2 \neq 0$，导出矛盾。因此，不$\exists \underline{\alpha} \neq 0$，使得$V \bigcap W = \underline{\alpha}$，也就是$V \bigcap W = \{\underline{0}\}$。□

**性质**：对于$\forall \underline{x} \in V$，可以有$x = x_r + x_n$，其中，$x_r \in C(A^T), x_n \in N(A)$。

**证明**：令$v_1, v_2, \ldots, v_r$是$C(A^T)$的一组基，$w_1, w_2, \ldots, w_{n-r}$是$N(A)$的一组基。

令$a_1v_1 + a_2v_2 + \ldots + a_rv_r + b_1w_1 + b_2w_2 + \ldots + b_{n-r}w_{n-r} = 0$

**证法1**：使用$v_1, v_2, \ldots, v_r, w_1, w_2, \ldots, w_{n-r}$分别左乘上式子，可以得到：

$$
\begin{cases}
a_1v_1^Tv_1 & & & & = 0 \\
& a_2v_2^Tv_2 & & & = 0 \\
& & \vdots & & \\
& & & a_nv_n^Tv_n & = 0 \\
b_1w_1^Tw_1 & & & & = 0 \\
& b_2w_2^Tw_2 & & & = 0 \\
& & \vdots & & \\
& & & b_{n-r}w_{n-r}^Tw_{n-r} & = 0
\end{cases}
$$

由于基向量不可能是零向量，因此，$v_i^Tv_i \neq 0$，所以说系数$a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_{n-r}$必须全为0。故，这组向量线性无关。

**证法2**：令$\underline{u} = a_1v_1 + a_2v_2 + \ldots + a_rv_r = -(b_1w_1 + b_2w_2 + \ldots + b_{n-r}w_{n-r})$

由于$v_i \in C(A^T), w_k \in N(A)$，且$C(A^T) \perp N(A)$，故$\underline{u} = \underline{0}$，又因为基向量一定是线性无关的，所以说，系数只能够全为0。因此，这两组向量组合起来也是线性无关的。

因此，$v_1, v_2, \ldots, v_r, w_1, w_2, \ldots, w_{n-r}$是互相独立的，这组向量是$\mathbb{R}^n$的一组基。

$$
\begin{aligned}
x &= (a_1v_1 + a_2v_2 + \ldots + a_rv_r) + (b_1w_1 + b_2w_2 + \ldots + b_{n-r}w_{n-r}) \\
&= x_r + x_n(for\ v_1, v_2, \cdots, v_r\ are\ in\ C(A^T), and\ w_1, w_2, \ldots, w_{n-r}\ are\ in\ n(A))
\end{aligned}
\square
$$

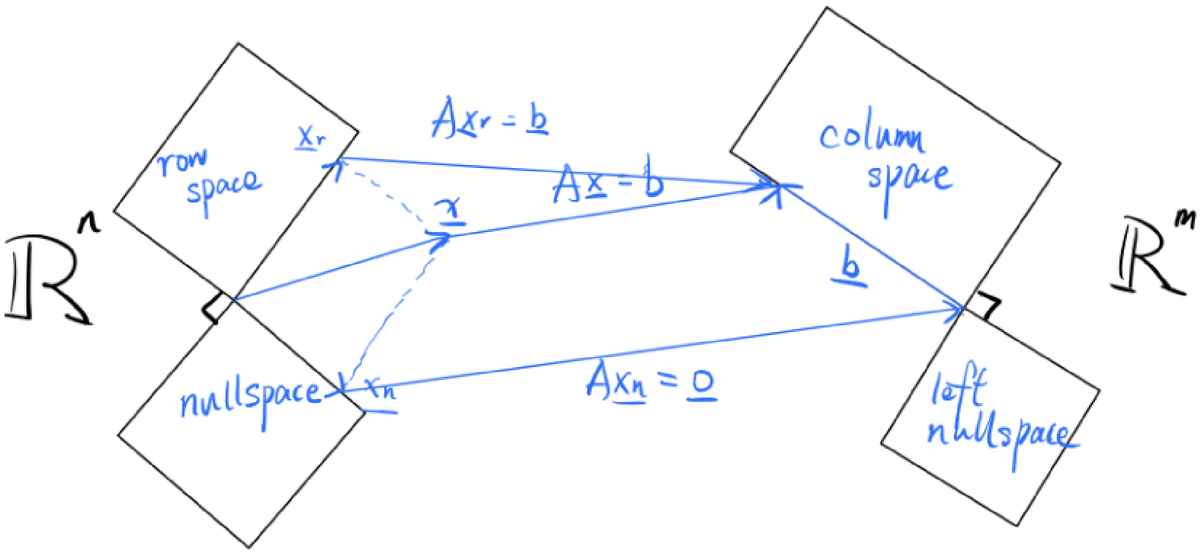
**性质**：给定矩阵$A$，$x = x_r + x_n$这样的分解是唯一的。（因为上面已经证明了他们是线性无关的）

**性质**：给定矩阵A，如果对他取转置得到$A^T$，那么可以得到$x = x_c + x_{ln}$，其中$x_c \in C(A), x_{ln} \in N(A^T)$

那么，对于$\underline{x} \in \mathbb{R}^n$，$Ax = A(x_r + x_n) = Ax_r + 0 = Ax_r$

也就是，通过矩阵$A$，$x_n$就变成了$0$，$Ax_n = 0$；而$x_r$通过A变换到了列空间里，即$Ax_r = b$

这也就组成了封面图！




**性质**：对于列空间中的每一个向量$\underline{b}$，存在且仅存在唯一的$x_r \in C(A)$与之对应。

**证明**：假设存在$b = Ax_r = Ax_r^{'} \in C(A)$，其中$x_r, x_r^{'} \in C(A^T)$

$$\begin{aligned}&\Rightarrow A(x_r - x_r^{'}) = 0\\&\Rightarrow x_r - x_r^{'} \in N(A)\\&Because,\ linear\ combination\ of\ vectors\\&in\ column\ space\ will\ not\ appears\ in\ nullspace\\&except\ \underline{0}.\ Therefore,\ x_r - x_r^{'}\ will\ only\ be\ \underline{0},\\&yielding\ x_r = x_r^{'}.\square\end{aligned}$$

**一个简单的例子**：

$$A = \begin{bmatrix}1 & 2\\3 & 6\end{bmatrix}, x = \begin{bmatrix}4\\3\end{bmatrix},\ 有\ x = a\begin{bmatrix}1\\2\end{bmatrix} + b\begin{bmatrix}2\\-1\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix} + \begin{bmatrix}2\\-1\end{bmatrix}$$

## 3、投影

首先，举一个简单的例子看一下什么是投影。

向量$x = \begin{bmatrix}1\\2\\5\end{bmatrix}$投影到 $z$ 轴和 $xy$ 平面上的投影分别是什么？$(1). p_1 = \begin{bmatrix}0\\0\\5\end{bmatrix}; (2). p_2 = \begin{bmatrix}1\\2\\0\end{bmatrix}$

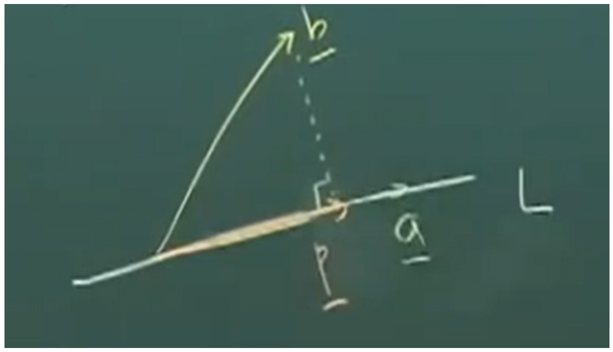
如果要把算投影的这个过程变换成一个矩阵，那么，上面两个向量的运算过程可以改成：

$$p_1 = \mathbb{P}_1 x = \begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\2\\5\end{bmatrix} = \begin{bmatrix}0\\0\\5\end{bmatrix}$$
$$p_2 = \mathbb{P}_2 x = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\2\\5\end{bmatrix} = \begin{bmatrix}1\\2\\0\end{bmatrix}$$

会发现一个比较简单的事实，$p_1 + p_2 = x$。其原因是，投影到的两个子空间互为正交补，它们的并集是$\mathbb{R}^3$。而他们的投影矩阵来说，有$\mathbb{P}_1 + \mathbb{P}_2 = I$。

**——投影到一条直线上**

**问题定义**：我们要找到向量$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$投影到方向$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$的投影是什么?




可以令$\underline{p} = \hat{x}\underline{a}$，并且令$\underline{e} = \underline{b} - \underline{p} = \underline{b} - \hat{x}\underline{a}$ 和 $\underline{a}$应该是正交的。

有

$$\begin{aligned} &\underline{a}^T\underline{e} = \underline{a}^T(\underline{b} - \hat{x}\underline{a}) = \underline{a}^T\underline{b} - \underline{a}^T\hat{x}\underline{a} = 0 \\ &\Rightarrow \hat{x}\underline{a}^T\underline{a} = \underline{a}^T\underline{b} \\ &\Rightarrow \hat{x} = \frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}} \end{aligned}$$

故

$$\underline{p} = \hat{x}\underline{a} = \frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}}\underline{a}$$

如果$\underline{b} = \underline{a}$，那么，$\underline{x} = 1, \underline{p} = \underline{a}$

如果$\underline{b} \perp \underline{a}$，那么$\hat{x} = 0$，$\underline{p} = \underline{0}$

由于投影的算式可以写成$\underline{p} = \mathbb{P}\underline{b} = \frac{\underline{a}\underline{a}^T}{\underline{a}^T\underline{a}}\underline{a}$，其中$\mathbb{P} = \frac{\underline{a}\underline{a}^T}{\underline{a}^T\underline{a}}(m \times m\ matrix)$[这边不用考虑投影矩阵的实际意义，我可以直接理解为它解出来就应该是这个样子]

进一步，$\underline{e} = \underline{b} - \underline{p} = \underline{b} - \mathbb{P}\underline{b} = (I - \mathbb{P})\underline{b}$

所以说，这里就诞生了两个视角来看到投影这个问题，第一个，投影是算出某一个比例，这个倍数乘以向量$\underline{a}$就能够得到向量$b$在$a$方向上的投影；第二个视角，就是直接算出投影矩阵，然后用投影矩阵乘上向量，最终就能够得到投影后的向量。

老师这里给了一个例子，投影的方向是：$\underline{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$。

那么，根据算式，投影矩阵$\mathbb{P} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix}$

这时，我突然心血来潮，计算了一下$\mathbb{P}^2 = \frac{1}{9*9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} = \mathbb{P}$

从投影这个动作的角度来考虑，可以理解成$\mathbb{P}^2\underline{b} = \mathbb{P}(\mathbb{P}\underline{b})$，投影矩阵对一个已经在直线上的向量做投影，得到的还会是其本身，即$\mathbb{P}\underline{b}$。

而从计算的角度来考虑$\mathbb{P}^2 = \frac{1}{(\underline{a}^T\underline{a})^2}\underline{a}(\underline{a}^T\underline{a})\underline{a}^T = \frac{1}{\underline{a}^T\underline{a}}\underline{a}\underline{a}^T = \mathbb{P}$

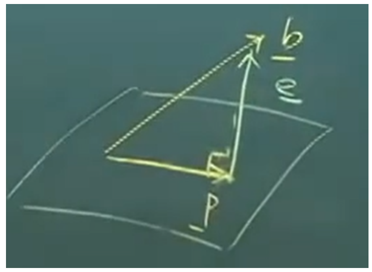
**——投影到某个子空间上**

首先，在$\mathbb{P}^m$上有$n$个线性无关的向量$a_1, a_2, \ldots, a_n$，这$n$个线性无关的向量张成了$\mathbb{R}^m$上$n$维的子空间$A$。

我们想要找到向量$\underline{b}$ 在子空间$A$上的投影$\underline{p} = \hat{x}_1\underline{a_1} + \hat{x}_2\underline{a_2} + \ldots + \hat{x}_n\underline{a_n}$(你要投影到这个子空间上，那投影向量一定是用这个子空间的基来线性表示的，所以是写成这个形式，也就是在$A$的列空间中)

我们可以把$A$写成$A=\begin{bmatrix} \underline{a_1} & \underline{a_2} & \cdots & \underline{a_n} \end{bmatrix}$

而向量$\underline{p}$可以改写成$\underline{p}=A\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix}=A\underline{\hat{x}}$




误差向量$\underline{e}=\underline{b}-A\underline{\hat{x}}$，有$\underline{e}\perp\underline{p}$，$\underline{p}$由$A$的列空间的一组基构成，那么$\underline{e}$就正交于$A$的列空间

> 因此，我们按照投影到直线的角度来思考这个问题。

$$
\begin{aligned}
\underline{p}^T\underline{e}&=\underline{0}\\
\underline{p}^T(\underline{b}-\underline{p})&=0\\
\underline{\hat{x}}^TA^T(\underline{b}-A\underline{\hat{x}})&=0\\
A^TA\underline{\hat{x}}&=A^T\underline{b}\\
(A^TA\text{一定是满秩的})\\
\underline{\hat{x}}&=(A^TA)^{-1}A^Tb
\end{aligned}
$$

其实这边还要证明一个非常简单的事实，$A^TA$是对称矩阵，且在$A$的列线性无关的情况下，$A^TA$可逆。

**证明**：

$$
\begin{gathered}
(1)\text{、}(A^TA)^T=A^TA\\
(2)\text{、只需要证明}rank(A)=rank(A^TA)=rank(AA^T)\\
\text{然后由于}A\text{和}A^TA\text{都有}n\text{列，那只需要证明他俩的零空间的维度一样就行}\\
\text{或者说，我证明他俩的零空间是一样的就行}\\
\text{所以有命题：}N(A)=N(A^TA)\\
\begin{aligned}
\Rightarrow:\qquad\qquad A\underline{x}&=0\\
A^TA\underline{x}&=0\\
\Leftarrow:\qquad\qquad A^TA\underline{x}&=0\\
\underline{x}^TA^TA\underline{x}&=0\\
(A\underline{x})^TA\underline{x}&=0\\
\text{令}\underline{y}&=A\underline{x}\\
\underline{y}^T\underline{y}&=0\\
\Rightarrow\qquad\quad \underline{y}&=0\\
\Rightarrow\qquad\qquad A\underline{x}&=0
\end{aligned}\\
\text{故说明}A^TA\text{和}A\text{的零空间互为对方的子集，这就说明了两者是一样的}\\
\text{因此}N(A)=N(A^TA)\text{，}rank(A)=rank(A^TA)\\
\text{类似的}, N(A)=N(AA^T)\text{，}rank(A)=rank(AA^T)\\
\text{然后，由于}A\text{有}n\text{列线性无关的列向量，所以说，}rank(A)=rank(A^TA)=n,\text{故}A^TA\text{可逆}\square
\end{gathered}
$$

所以说，$\underline{\hat{x}}=(A^TA)^{-1}A^T\underline{b}$

还有，$\underline{p}=A\underline{\hat{x}}=A(A^TA)^{-1}A^T\underline{b}$

故投影矩阵$\mathbb{P}=A(A^TA)^{-1}A^T$

另外，还有$\mathbb{P}^2=A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T=A(A^TA)^{-1}A^T=\mathbb{P}$

**例子**：

$$A=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix},\underline{b}=\begin{bmatrix}6\\0\\0\end{bmatrix}$$

$$A^TA=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}=\begin{bmatrix}3&3\\3&5\end{bmatrix},A^Tb=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}$$

$$\underline{\hat{x}}=(A^TA)^{-1}A^T\underline{b}=\begin{bmatrix}5\\-3\end{bmatrix}$$

$$\underline{p}=A\underline{\hat{x}}=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}\begin{bmatrix}5\\-3\end{bmatrix}=\begin{bmatrix}5\\2\\-1\end{bmatrix}$$

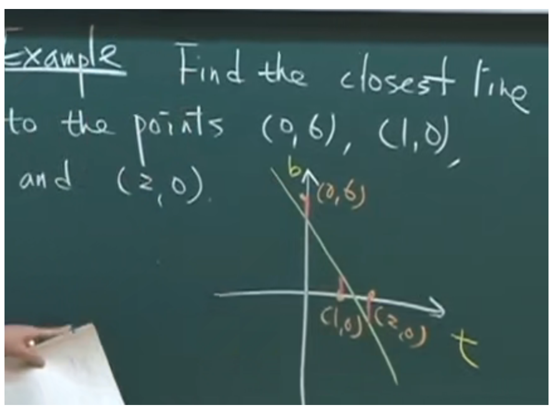
$$\underline{e}=\underline{b}-\underline{p}=\begin{bmatrix}6\\0\\0\end{bmatrix}-\begin{bmatrix}5\\2\\-1\end{bmatrix}=\begin{bmatrix}1\\-2\\1\end{bmatrix}$$

$$A^T\underline{e}=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}1\\-2\\1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

**——投影的应用—最小二乘法**

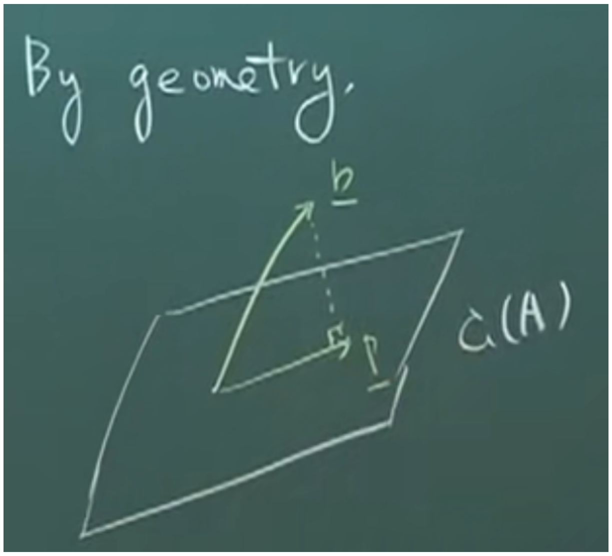
$\underline{e}=\underline{b}-\underline{p}$，这个$\underline{e}$通常不可能为$\underline{0}$。但是，我们可以让$||\underline{e}||^2$最小化；也就是最小二乘解。

通常来说，给定的数据集包含两部分，一部分是$X$，它是一个$m\times n$的矩阵，$m$代表样本容量，$n$代表特征个数，而$m$往往会远大于$n$；然后，还有一部分是因变量$y$，是一个$n\times 1$的向量；还有系数向量$\beta$；需要求解的是$X\beta=y$。然而，由于$m>>n$，它基本上是不存在解的；因为经过$Gauss\ Elimination$，下面的$m-r$行基本上不可能全为0。而我们的目标是，求解出使得$||\underline{e}||^2=||\underline{y}-X\beta||^2$尽可能小的一组$\beta$。从图像上来看，它是这样的。




$X\beta$就是图中的那根直线，而 $X$ 和 $\underline{y}$ 可以简单地理解为图上的那些离散点。而$||\underline{e}||^2=||\underline{y}-X\beta||^2$就是图上红色的那些部分平方再相加。

根据线性代数基本定理，$\mathbb{R}^m$中的向量$x$可以唯一分解为$\underline{x}=\underline{x}_c+\underline{x}_{ln}$。所以，我们可以把$y$先投影到$X$的列空间中，得到$\underline{y}_c$。此时，$\underline{y}_c\perp X\beta$，所以说这个时候，他俩之间的误差一定是最小的。(下图可见)然后再求解系数$\beta$，就得到了最小二乘解。




当然了，这是比较直观/图像化的解释，我们还需要严格地证明这是最小的。

**命题**：最小二乘解$\hat{\beta}$一定能够使得$||X\beta-\underline{y}||^2$最小。

根据线性代数基本定理，$\mathbb{R}^m$中的向量一定可以唯一分解成$\underline{x}=\underline{x}_c+x_{ln}$。

因此，在给定$m\times n$的数据集$X$和对应的向量$\underline{y}$的情况下。我们同样可以将$\underline{y}$分解成为$\underline{y}=\underline{y}_c+\underline{y}_{ln}$。$\underline{y}_c$是$\underline{y}$投影到$C(X)$上的向量。而$\underline{y}_{ln}=\underline{y}-\underline{y}_c$。$\underline{y}_c$和$\underline{y}_{ln}$分别位于$C(X)$和$N(X^T)$中，因此，此二向量一定是正交的。故$\underline{e}=\underline{y}_{ln}$

$$
\begin{aligned}
&||X\beta-\underline{y}||^2=(X\beta-\underline{y}_c-\underline{y}_{ln})^T(X\beta-\underline{y}_c-\underline{y}_{ln})\\
&=(X\beta-\underline{y}_c)^T(X\beta-\underline{y}_c)-(X\beta-\underline{y}_c)^T\underline{y}_{ln}-\underline{y}_{ln}(X\beta-\underline{y}_c)^T+\underline{y}_{ln}^T\underline{y}_{ln}\\
&=||X\beta-\underline{y}_c||^2+||\underline{y}_{ln}||^2(X\beta-\underline{y}_c\text{在}C(A)\text{中，}\underline{y}_{ln}\text{在}N(A^T)\text{中})
\end{aligned}
$$

如果要上式最小，要有$||X\beta-\underline{y}_c||^2$最小。即$\beta=\hat{\beta}$。$\min\limits_{\beta}||X\beta-\underline{y}||^2=||\underline{y}_{ln}||^2$。□

而最小二乘解的计算方法是：

$$
\begin{aligned}
A^T(\underline{y}-A\hat{\beta})&=0\\
A^TA\hat{\beta}&=A^T\underline{y}\\
\hat{\beta}&=(A^TA)^{-1}A^T\underline{y}
\end{aligned}
$$

### 4、正交基和施密特正交化

**定义**：一组向量$q_1,q_2,\ldots,q_n$被称为是正交的，当且仅当$q_i^Tq_t=0(i\neq t)$。

**性质**：如果非零向量$q_1,q_2\ldots,q_n$是一组正交向量，难么这组向量是线性无关的。

**证明**：

$$
\text{令}Q=\begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}
$$

$$
\text{要求解}Qx=0\text{，可以两边左乘}Q^T
$$

$$
\text{那么}Q^TQx=\begin{bmatrix} q_1^Tq_1 & 0 & 0 & \cdots & 0\\ 0 & q_2^Tq_2 & 0 & \cdots & 0\\ & \vdots & & & \\ 0 & 0 & 0 & \cdots & q_n^Tq_n \end{bmatrix}x=0
$$

$$
\text{由于}q_i\text{是非零向量，则}||q_i||^2\neq 0;\text{因此，只有}x=0\text{时才满足，故线性无关}\square
$$

**定义**：如果一组向量$q_1,q_2,\ldots,q_n$是一组正交基，满足以下性质：

$$
q_i^Tq_j=\begin{cases}1,\ i=j\\ 0,\ i\neq j\end{cases}
$$

**性质**：$n$个$m$维的正交向量构成的矩阵$Q$，$Q^TQ=I(n\times n)$

如果Q是一个$n \times n$的矩阵，由于那么$Q^T = Q^{-1}$。而这个$Q$就被称为**正交矩阵。**

**例子**：（旋转矩阵）

$$
\begin{gathered}
Q = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \\
Q\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} cos\theta \\ sin\theta \end{bmatrix} \\
Q\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -sin\theta \\ cos\theta \end{bmatrix}
\end{gathered}
$$

**例子**：(轴对称反转剧矩阵)

$$Q = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}$$

**例子**：(y=x反转剧矩阵)

$$Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$Q\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

**性质**：如果$Q$有orthonormal列$(Q^TQ = I)$，那么：

(Ⅰ)、$||Qx|| = ||x||$

(Ⅱ)、$(Qx)^T(Qy) = x^Ty$

**证明**：

$$
\begin{gathered}
(\text{Ⅰ})\text{、要证}||Qx|| = ||x||\text{，可以转化为}(Qx)^TQx = x^TQ^TQx = x^Tx \\
\text{故}||Qx||^2 = ||x||^2,\text{所以有}||Qx|| = ||x||\text{。}\square \\
(\text{Ⅱ})\text{、}(Qx)^T(Qy) = x^TQ^TQy = x^Ty.\square
\end{gathered}
$$

**——用正交基来投影**

有了正交基，投影就会非常方便。假设我们有$n$个$\mathbb{R}^m$中的正交基$q_1, q_2, \ldots, q_n$构成的矩阵$Q$。如果我们要把向量$b$投影到$Q$的列空间上去，那么，我们只需要：

$$
\begin{aligned}
Q^TQ\hat{x} &= Q^Tb \\
I\hat{x} &= Q^Tb \\
\hat{x} &= Q^Tb
\end{aligned}
$$

另外：

$$
\begin{aligned}
q = Q\hat{x} &= QQ^Tb \\
QQ^Tb &= \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} b \\
&= \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^Tb \\ q_2^Tb \\ \vdots \\ q_n^Tb \end{bmatrix} \\
&= q_1^Tbq_1 + q_2^Tbq_2 + \cdots + q_n^Tbq_n \\
&= (q_1q_1^T)b + (q_2q_2^T)b + \cdots + (q_nq_n^T)b
\end{aligned}
$$

简单来说，就是使用$q_iq_i^t$这个投影矩阵，把向量$b$分别投影到$Q$的每个列向量上，然后再给他线性组合起来。然后$q_i$是一个单位向量，所以说，$(q_i^tb)q$就是向量$b$投影到$q_i$方向上的长度再乘以方向向量。

然后，投影矩阵就是$\mathbb{P} = QQ^T$

当$Q$是一个方阵的时候，$C(Q)$就是$R^n$，且$Q^T=Q^{-1}$，其解为$\hat{x}=Q^{-1}b$。而此时的投影矩阵就是$\mathbb{P}=QQ^T=QQ^{-1}=I$，投影就是向量本身。

那么，向量$b$还可以分解为$b=q_1(q_1^Tb)+q_2(q_2^Tb)+\cdots+q_n(q_n^Tb)$

**——怎么找正交基？—Gram-Schmidt Orthogonalization**

**条件**：给定$n$个线性无关的向量$a_1,a_2.\cdots,a_n$，找到$n$个能够张成一样的空间的正交向量$q_1,q_2,\cdots,q_n$

**方法**：

第一次先取出$a_1$，把它单独视作一个子空间，然后找出这个子空间的正交基$\{q_1\}$。(这一步其实就是给向量单位化一下)

然后取出$q_1,a_2$，然后找到这个子空间的一组基。使用的原理就是上面的那个式子$a_2=q_1q_1^ta_2+q_2(q_2^ta_2)$，所以说，$q_2$的方向一定和$a_2-q_1q_1^ta_2$一样，那么，只需要把前面那个向量解出来，并给它单位化一下，得到这个子空间的正交基$\{q_1,q_2\}$。

以此类推，就能够不那么快速地得到一组正交基$\{q_1,q_2,\cdots,q_n\}$

{一点小补充，投影向量一定会垂直去减掉的那个向量}

给出一个通项，$A_i=a_i-\sum_{j=1}^{i-1}(q_j^Tb)q_j,q_i=\frac{A_i}{||A_i||}$

{这好像就是一个算法。。。}

然后，我们放宽条件，不需要假设是$n$个线性无关的向量。那假如前两个向量都是线性无关的，但是，第三个向量是前两个向量的线性组合，那算出来的$A_i$会是零向量，那直接丢掉，算接下来的就行。假如不是线性无关的几个向量，那最后的正交基一定会小于$n$个向量。

## 5、QR分解

对于一个$m\times n$的矩阵$A$，我们可以假设它有$r(1\le r\le min\{m,n\})$个线性无关的正交向量作为基，然后，我们可以获得一个$m\times r$的正交矩阵Q。由于$C(Q)$和$C(A)$是同一个东西，因此，用$Q$的列向量一定能线性表示$A$中的每一个列向量，这就可以用一个$r\times n$系数矩阵$R$来记录系数矩阵。而这，就是QR分解！具体的分解如下所示：

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}=A=QR=\begin{bmatrix} q_1 & q_2 & \cdots & q_r \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ & \vdots & & \\ r_{r1} & r_{r2} & \cdots & r_{rn} \end{bmatrix}$$

$$\text{其中，}a_i=Qr_i=\begin{bmatrix} q_1 & q_2 & \cdots & q_r \end{bmatrix}\begin{bmatrix} r_{1i} \\ r_{2i} \\ \cdots \\ r_{ri} \end{bmatrix}$$

但实际上，$QR$分解不是这样的。

根据$Gram-Schmidt$分解，我们会得到：

$$\begin{cases} a_1=q_1(q_1^T)a_1 \\ a_2=q_1q_1^Ta_2+q_2q_2^Ta_2 \\ \vdots \\ a_n=q_1q_1^Ta_n+q_2q_2^Ta_n+\cdots+q_nq_n^Ta_n \end{cases}$$

转换为矩阵形式就是：

$$A=QR$$

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}=\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}\begin{bmatrix} q_1^Ta_1 & q_1^Ta_1 & \cdots & q_1^Ta_1 \\ 0 & q_2^Ta_2 & \cdots & q_2^Ta_2 \\ 0 & 0 & \cdots & q_3^Ta_3 \\ & \vdots & \ddots & \\ 0 & 0 & & q_n^Ta_n \end{bmatrix}$$

$$Q^TA=Q^TQR=IR=R$$

可以考察矩阵$R$对角线上的元素：

$$
\begin{aligned}
&A_1=||A_1||q_1=a_1,q_1^Ta_1=||A_1||\\
&A_2=||A_2||q_2=a_2-q_1(q_1^Ta_2),q_2^Ta_2=||A_2||\\
&A_3=||A_3||q_3=a_3-q_1(q_1^Ta_3)-q_2(q_2^Ta_3),q_3^Ta_3=||A_3||\\
&\quad\vdots\\
&A_n=||A_n||q_n=a_n-\sum_{j=1}^{n-1}q_j(q_j^Ta_n),q_n^Ta_n=||A_n||
\end{aligned}
$$

因此，$R$是一个有着正对角线元素的上三角矩阵，故$R$可逆。

故，最小二乘法可以改写成：

$$
\begin{aligned}
A\hat{x}&=b\\
A^TA\hat{x}&=A^Tb\\
(QR)^T(QR)\hat{x}&=(QR)^Tb\\
R^TR\hat{x}&=R^TQ^Tb\\
R\hat{x}&=Q^Tb(R\text{是上三角阵，所以可以用回代直接解决})
\end{aligned}
$$

## 6、内积空间

**定义**：向量空间$V$上的**内积**是一个函数，对于任意两个向量（有序对）$v,w$来说都会被转化成一个标量$<v,w>$，对于任意向量$u,v,w$和标量$c$而言，会满足以下的性质：

(Ⅰ)、$<u+v,w>=<u,w>+<v,w>$

(Ⅱ)、$<cv,w>=c<v,w>$

(Ⅲ)、$<v,w>=<w,v>$

(Ⅳ)、$<v,v>\geq 0, if\ <v,v>=0 \iff v=0$

**定义**：如果一个向量空间具有内积，那么这个向量空间就被称作**内积空间**

**大栗子**：

首先，对于$V=C[a,b]$上的所有连续实函数是一个向量空间，它满足加法和数乘。

然后，我们可以定义再这个向量空间上的内积为：$<f,g>=$。这个东西满足内积的4条性质。

(Ⅰ)、$<f+h,g>=\int_a^b[f(t)+h(t)]g(t)dt=\int_a^b[f(t)g(t)+h(t)g(t)]dt=<f,g>+<h,g>$

(Ⅱ)、$<cf,g>=\int_a^b cf(t)g(t)dt=c\int_a^b f(t)g(t)dt=c<f,g>$

(Ⅲ)、$<f,g>=\int_a^b f(t)g(t)dt=\int_a^b g(t)f(t)dt=<g,f>$

(Ⅳ)、$<f,f>=\int_a^b f(t)f(t)dt\geq 0$，$<f,f>=0 \iff f=0$

下面就要引入鼎鼎大名的傅里叶级数了。

首先，考虑$V=C[0,2\pi],<f,g>=\int_0^{2\pi}f(t)g(t)dt$，以及以下函数：$\{1,cost,sint,cos2t,sin2t,\cdots,cos(Nt),sin(Nt)\}$

那么，有：$<1,cosnt>=\int_0^{2\pi}cosntdt=\frac{1}{n}sinnt|_0^{2\pi}=0(\text{一个周期刚好等于}0)$

$<1,sinnt>=\int_0^{2\pi}sinntdt=-\frac{1}{n}cosnt|_0^{2\pi}=0$

$<cosnt,sinnt>=\int_0^{2\pi}cosnt\cdot sinntdt=\frac{1}{n}\int_0^{2\pi}sinntdsinnt=\frac{1}{2n}sin^2nt|_0^{2\pi}=0$

$<cosnt,sinmt>=\int_0^{2\pi}cosnt\cdot sinmtdt=\frac{1}{2}\int_0^{2\pi}[sin(n+m)t-sin(n-m)t]dt=0$

$<cosnt,cosmt>=\int_0^{2\pi}cosnt\cdot cosmtdt=\frac{1}{2}\int_0^{2\pi}cos(n-m)t+cos(n+m)tdt=0$

$<sinnt,sinmt>=\int_0^{2\pi}sinnt\cdot sinmtdt=\frac{1}{2}\int_0^{2\pi}cos(n-m)t-cos(n+m)tdt=0$

另外，还有：$<1,1>=\int_0^{2\pi}1^2dt=2\pi$

$<cosnt,cosnt>=\int_0^{2\pi}cos^2ntdt=\pi$

$<sinnt,sinnt>=\pi$

所以说，$S_N=\{\frac{1}{\sqrt{2\pi}},\frac{1}{\sqrt{\pi}}cost,\frac{1}{\sqrt{\pi}}sint,\frac{1}{\sqrt{\pi}}cos2t,\frac{1}{\sqrt{\pi}}sin2t,\cdots,\frac{1}{\sqrt{\pi}}cosNt,\frac{1}{\sqrt{\pi}}sinNt\}$是一组正交的函数

因此，给定$C[0,2\pi]$上的实函数，可以给它投影到这组正交基上面去，求系数，用这组正交基来表示实函数。

我们可以将其表示为：

$f_N(t)=\frac{A_0}{\sqrt{2\pi}}+\frac{A_1}{\sqrt{\pi}}cost+\frac{B_1}{\sqrt{\pi}}sint+\cdots+\frac{A_N}{\sqrt{\pi}}cosNt+\frac{B_n}{\sqrt{\pi}}sinNt=\frac{A_0}{\sqrt{2\pi}}+\frac{1}{\sqrt{\pi}}\sum_{n=1}^{N}(A_ncos(nt)+B_nsin(nt))$

至于怎么计算$A_i$和$B_i$，那我们就可以做投影，把函数投影到$cosnt$和$sinnt$中去，来计算系数。

具体方法为：

$<f(t),\frac{1}{\sqrt{2\pi}}>=\frac{1}{\sqrt{2\pi}}\int_0^{2\pi}f(t)dt$

$<f(t),\frac{1}{\pi}cosnt>=\frac{1}{\pi}\int_0^{2\pi}f(t)cos(nt)dt=A_n$

$<f(t),\frac{1}{\pi}sinnt>=\frac{1}{\pi}\int_0^{2\pi}f(t)sin(nt)dt$

## (四)、行列式

### 1、行列式的10条性质

**定义**：$n\times n$实矩阵的**行列式**记作$detA$或$|A|$是一个满足以下三条性质的标量：

(Ⅰ)、$detI=1$

(Ⅱ)、在交换矩阵任意两行（列）时，行列式符号会发生变化

(Ⅲ)、行列式是每个行（列）的线性函数。

**栗子**：

$$det\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}=-det\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}=-1$$

如果说$P$是一个排列矩阵，那么$detP=(-1)^n$，$n$是将排列矩阵恢复成单位矩阵$I$所需要的行交换次数。

**栗子**(线性性质)：

$$\begin{vmatrix}ta&tb\\c&d\end{vmatrix}=t\begin{vmatrix}a&b\\c&d\end{vmatrix}$$

$$\begin{vmatrix}s(a+a')&t(b+b')\\c&d\end{vmatrix}=s\begin{vmatrix}a&b\\c&d\end{vmatrix}+t\begin{vmatrix}a'&b'\\c&d\end{vmatrix}$$

**栗子**：

$$\begin{vmatrix}2&0\\0&2\end{vmatrix}=2^2\begin{vmatrix}1&0\\0&1\end{vmatrix}=2^2I=4$$

故$det(tI)=t^n$

用以上的三条性质，可以推出性质(Ⅳ)-(Ⅹ)：

（Ⅳ）、如果两行相等，那么$det(A)=0$

**证明**：根据性质（Ⅱ），如果交换两行，符号应该加负号。那如果交换两个相同的行，那么符号也应该跟着变反。但是，如果交换了两个一样的行，那得到的结果应该跟没交换前是一样的，而唯一满足这个性质的数是0。所以说，如果有两行一样的矩阵，其行列式应当为0。□

（Ⅴ）、从行列式的第$i$行中减去第$j(i\neq j)$的倍数（可正可负）不会影响行列式的值。

**证明**：可以根据性质（Ⅲ）和性质（Ⅳ）来推出该性质。

$$\begin{vmatrix}a_1+ka_i\\a_2\\\vdots\\a_n\end{vmatrix}=\begin{vmatrix}a_1\\a_2\\\vdots\\a_n\end{vmatrix}+\begin{vmatrix}ka_i\\a_2\\\vdots\\a_n\end{vmatrix}=detA+0=detA.\square$$

**性质**：假设$A$通过高斯消去法得到了矩阵$U$（未经过行交换），那么$detA=detU$。

（Ⅵ）、如果$A$存在0行，那么$detA=0$。

**证明**：从性质（Ⅴ）和（Ⅳ）可以推出本性质。

$$\begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ 0 \\ \vdots \\ a_n \end{vmatrix} = \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{vmatrix} = 0.\square$$

**性质**：

（Ⅶ）、如果$A$是一个三角矩阵（上三角或下三角），那么，$detA = a_{11}a_{22}\cdots a_{nn} = products\ of\ diagonal\ entries$

**证明**：

**情形一**：对角线全是非零元素

那么，根据消去法，可以用最后一行先把最后一列的上$n-1$个元素消成0，然后再利用倒数第二行消去倒数第二列，$\cdots$，以此类推，直到消去所有非对角线元素，变成$detD$。根据性质（VI），研究$detD$等价于研究$detA$。然后，再利用线性性质（Ⅲ），将每一行的数提出，变成$detA = a_{11}a_{22}\cdots a_{nn}detI = a_{11}a_{22}\cdots a_{nn}$（性质（Ⅰ））

**情形二**：至少有一个对角线元素是0

那么，矩阵$A$是奇异的，也就是至少存在一个零行。根据性质（Ⅳ），$detA = 0$。

综上，可以得到结论：$detA = a_{11}a_{22}a_{33}\cdots a_{nn}\square$

（Ⅷ）、如果矩阵$A$是奇异矩阵，那么$detA = 0$。反之，如果矩阵$A$是可逆矩阵，那$detA \neq 0$。

**证明**：

$detA = detU$，那么，只需要考察 $U$ 的性质就行。如果$U$存在$n$个$pivot$，那么表示这个矩阵是满秩的（不存在0行），那么$detA \neq 0$；反之，如果$pivot$数目少于$n$个，那表示一定存在0行，那$detA = 0$。$\square$

**性质**：假如$p_1, p_2, \cdots, p_n$是$pivot\ number$，那么，$detA = \pm p_1 p_2 \cdots p_n$

（Ⅸ）、$|AB| = |A||B|$

**证明**：

（1）、考虑$|B| = 0$的情形。那么$B$是奇异矩阵。则$AB$也是奇异矩阵。【因为$AB$的行向量就是$B$的行空间构成的，那$B$不满秩，那$AB$铁定不满秩///或，$Bx = 0$有非零解，那么左乘$A$，可以得到$ABx = 0$也有非零解，因此，$AB$也是奇异的】故$|AB| = 0, 且|A||B| = 0$。

（2）、若$|B| \neq 0$，记$d(A) = \frac{|AB|}{|B|}$。只需要证明$d(A)$满足性质（Ⅰ）、（Ⅱ）、（Ⅲ）即可。

性质（Ⅰ）：若$A = I$，则$d(A) = \frac{|IB|}{|B|} = 1$

性质（Ⅱ）：交换$A$的任意两行，那$AB$也会有对应的两行被互换，满足条件。

性质（Ⅲ）：trivial

$\square$

（Ⅹ）、$|A^T| = |A|$

**证明**：

（1）、当$A$是奇异时，$A^T$也是奇异的，那么$detA = 0 = detA^T$

（2）、$A$如果不是奇异矩阵，就会有$PA = LU \Rightarrow A^T P^T = U^T L^T$

根据性质（Ⅸ），就会有$|P||A| = |L||U|与|A^T||P^T| = |U^T||L^T|$

所以，只需要证明$|P| = |P^T|, |U| = |U^T|, |L| = |L^T|$，那就一定有$|A| = |A^T|$

首先，因为$L记录了消去的过程，对角线都是1$，因此，根据性质（Ⅶ），$|L| = |L^T| = 1$

其次，$|U| = |U^T|$

最后，由于$PP^T = I(P^{-1} = P^T) \Rightarrow |P||P^T| = 1 \Rightarrow |P|和|P^T|为\pm 1 \Rightarrow |P| = |P^T|$

故，$|A| = |A^T|。\square$

**性质**：对于行来说的每一条性质，对于列而言也都适用。

## 2、排列与Cofactors

**——主元公式**

$PA = LU, detPdetA = detLdetU \Rightarrow detA = \pm detU = \pm p_1p_2\cdots p_n$

如果$pivot$数目少于$n$个，那么$detA = 0$。

由于矩阵$A$前$k$个$pivot$只和左上角的$k\times k$矩阵有关系，因此，在没有行交换的情况下，有$detA_k = p_1p_2\cdots p_k, \forall k\in\{1,2,3,\cdots,n\}$，而第$k$个$pivot$的值为$p_k = \frac{p_1p_2\cdots p_k}{p_1p_2\cdots p_{k-1}} = \frac{A_k}{A_{k-1}}$。如果说，每个矩阵$A_k$的行列式均不为0，那就不需要做行交换。

**——Big Formula**

**举个例子**：

$$\begin{aligned}\begin{vmatrix}a & b\\ c & d\end{vmatrix} &= \begin{vmatrix}a & 0\\ c & d\end{vmatrix} + \begin{vmatrix}0 & b\\ c & d\end{vmatrix}\\ &= \begin{vmatrix}a & 0\\ c & d\end{vmatrix} + \begin{vmatrix}a & 0\\ 0 & d\end{vmatrix} + \begin{vmatrix}0 & b\\ c & 0\end{vmatrix} + \begin{vmatrix}0 & b\\ 0 & d\end{vmatrix}\\ &= ad - bc\end{aligned}$$

所以说，按这种方法拆分，按第一行拆开可以拆成$n$个，再按第二行拆可以拆出$n^2$个，以此类推，一共可以拆出$n^n$个行列式。但由于其中大部分都是0，最后只会剩下分布在行列均没有重复元素的那些矩阵，共$n!$个。

$$\begin{aligned}\begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix} &= \begin{vmatrix}a_{11} & 0 & 0\\ 0 & a_{22} & 0\\ 0 & 0 & a_{33}\end{vmatrix} + \begin{vmatrix}0 & a_{12} & 0\\ 0 & 0 & a_{23}\\ a_{31} & 0 & 0\end{vmatrix} + \begin{vmatrix}0 & 0 & a_{13}\\ a_{21} & 0 & 0\\ 0 & a_{32} & 0\end{vmatrix}\\ &+ \begin{vmatrix}a_{11} & 0 & 0\\ 0 & 0 & a_{23}\\ 0 & a_{32} & 0\end{vmatrix} + \begin{vmatrix}0 & 0 & a_{13}\\ 0 & a_{22} & 0\\ a_{31} & 0 & 0\end{vmatrix} + \begin{vmatrix}0 & a_{12} & 0\\ a_{21} & 0 & 0\\ 0 & 0 & a_{33}\end{vmatrix}\\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}\\ &- a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}\end{aligned}$$

对于一个$n\times n$的矩阵$A$来说，其行列式为：

$$detA = \sum_{\sigma=(\alpha,\beta,\cdots,\omega)}(detP_\sigma)a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$$
$$[\sigma\ is\ the\ permutation\ of\ (1,2,3\cdots,n)]$$
$$[P_\sigma\ is\ the\ determinant\ of\ \sigma]$$
$$There\ are\ n!\ possible\ permuation\ of\ (1,2,3,\cdots,n),\ thus\ we\ can\ get\ n!\ summation.$$

**——余子式**

$$\begin{aligned}\begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix} &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})\\ &= a_{11}(\begin{vmatrix}a_{22} & 0\\ 0 & a_{33}\end{vmatrix} - \begin{vmatrix}a_{23} & 0\\ 0 & a_{32}\end{vmatrix})\\ &+ a_{12}(\begin{vmatrix}a_{23} & 0\\ 0 & a_{31}\end{vmatrix} - \begin{vmatrix}a_{21} & 0\\ 0 & a_{33}\end{vmatrix})\\ &+ a_{13}(\begin{vmatrix}a_{21} & 0\\ 0 & a_{32}\end{vmatrix} - \begin{vmatrix}a_{22} & 0\\ 0 & a_{31}\end{vmatrix})\\ &= \begin{vmatrix}a_{11} & 0 & 0\\ 0 & a_{22} & a_{23}\\ 0 & a_{32} & a_{33}\end{vmatrix} + \begin{vmatrix}0 & a_{12} & 0\\ a_{21} & 0 & a_{23}\\ a_{31} & 0 & a_{33}\end{vmatrix} + \begin{vmatrix}0 & 0 & a_{13}\\ a_{21} & a_{22} & 0\\ a_{31} & a_{32} & 0\end{vmatrix}\\ &= a_{11}detM_{11} - a_{12}detM_{12} + a_{13}detM_{13}\\ &\triangleq a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.\ (C_{ij} = (-1)^{i+j}detM_{ij})\end{aligned}$$

对于$n\times n$的矩阵$A$，其行列式能够写成余子式的组合：

$$\begin{aligned}detA &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}\\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}\end{aligned}$$

其中，$C_{ij} = (-1)^{i+j}M_{ij}, M_{ij}$是去掉了$i$行$j$列的$(n-1)\times(n-1)$的$A$的子矩阵

**——克莱默法则、逆和体积**

**（1）克莱默法则**

要解$Ax=b$，有：

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b\end{bmatrix}$$

为了使用行列式的性质，可以对向量进行扩展：

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}\begin{bmatrix}x_1 & 0 & 0\\x_2 & 1 & 0\\x_3 & 0 & 1\end{bmatrix}=\begin{bmatrix}b & a_2 & a_3\end{bmatrix}$$

两边取行列式：

$$|A|x_1=|B|$$
$$x1=\frac{|B_1|}{|A|}$$

同理：

$$\begin{cases}x_1=\dfrac{|B_1|}{|A|}\\ x_2=\dfrac{|B_2|}{|A|}，当|A|\neq 0时\\ x_3=\dfrac{|B_3|}{|A|}\end{cases}$$

**（2）逆**

用上面的克莱默法则来求矩阵的逆，等价于求$n$组线性方程组。

有

$$AA^{-1}=I$$
$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\ & \vdots & \ddots & \\a_{n1} & a_{n2} & & a_{nn}\end{bmatrix}\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1n}\\x_{21} & x_{22} & \cdots & x_{2n}\\ & \vdots & \ddots & \\x_{n1} & x_{n2} & & x_{nn}\end{bmatrix}=\begin{bmatrix}1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1\end{bmatrix}$$
$$x_{11}=\frac{det\begin{bmatrix}1 & a_{12} & \cdots & a_{1n}\\0 & a_{22} & \cdots & a_{2n}\\ & \vdots & \ddots & \\0 & a_{n2} & & a_{nn}\end{bmatrix}}{detA}=\frac{(-1)^{1+1}detM_{11}}{detA}$$
$$x_{21}=\frac{det\begin{bmatrix}a_{11} & 1 & \cdots & a_{1n}\\a_{21} & 0 & \cdots & a_{2n}\\ & \vdots & \ddots & \\a_{n1} & 0 & & a_{nn}\end{bmatrix}}{detA}=\frac{(-1)^{1+2}detM_{12}}{detA}$$

以此类推。。。$x_{ij}=\frac{(-1)^{i+j}M_{ij}}{detA}$

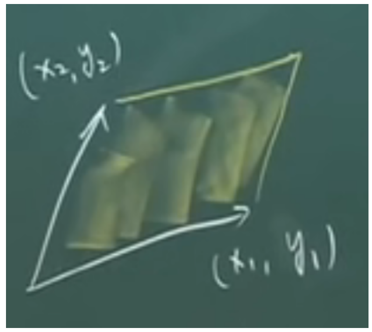
根据这个玩意，可以写出矩阵的伴随矩阵：

$$adj(A)=\begin{bmatrix}M_{11} & -M_{21} & \cdots & (-1)^{1+n}M_{n1}\\-M_{21} & M_{22} & \cdots & (-1)^{2+n}M_{n2}\\ & \vdots & \ddots & \\(-1)^{n+1}M_{1n} & (-1)^{n+2}M_{2n} & & M_{nn}\end{bmatrix}$$

另外：

$$A^{-1}_{ij}=\frac{(-1)^{i+j}M_{ji}}{detA}$$

**（3）体积**




所谓体积，就是多个向量张成的平行四边形/平行六面体/还有四维朝上我不知道怎么叫体的体积。

这个体积是带符号的，如果$(x_1, y_1) \rightarrow (x_2, y_2)$是顺时针的话，那就是正体积；反之就是负体积。

**性质**：面积= $det \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = detA$

**证明**：只需要证明这个公式满足行列式的三个性质即可。

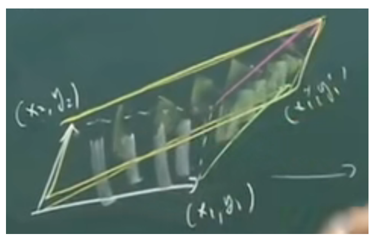
(Ⅰ)、如果$A = I$，那么$detA = 1$，由于$(0, 1)$和$(1, 0)$张成的平行四边形是一个正方形，因此，面积是1，故$area = detA$。

(Ⅱ)、当行列式的两行互换时，面积就会变号，故满足行列式的第二条性质。

(Ⅲ)、a.如果给向量的某个值乘以$t$倍，而它的面积也会相应地变成原来的$t$倍；满足线性性质。

b.如果给$(x_1, y_1)$加上一个向量$(x_1^{'}. y_1^{'})$，要证明线性性质，就要证明$(x_1, y_1)$到$(x_2, y_2)$和$(x_1^{'}, y_1^{'})$到$(x_2, y_2)$的面积和，与$(x_1 + x_1^{'}, y_1 + y_1^{'})$的面积时一样的。

通过图可知：



中间有两块重叠部分，又张成下半个三角形和上面那个三角形的俩向量是一样的，所以说两组向量的面积之和与先将向量相加后得到的面积之和是一样的。又

$$det \begin{bmatrix} x_1 + x_1^{'} & y1 \\ x_2 + x_2^{'} & y_2 \end{bmatrix} = det \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} + det \begin{bmatrix} x_1^{'} & y_1 \\ x_2^{'} & y_2 \end{bmatrix}$$

故面积也满足线性性质。

由于两向量的面积满足行列式的三个性质，因此，面积可以用行列式来计算。□

**性质**：有三个向量$a_1, a_2. a_3$构成的平行四面体，其体积=$det \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$

接下来从另一个角度讲讲为什么行列式就是“体积”。

这是因为正交化的过程就是减去另外几个行（列）向量的倍数的过程；而行列式又正好满足减去另外一行的倍数值不发生变化这一性质，这就导致了，行列式的值就是经过Gram-Schmidt正交化后的那几个向量的模的乘积再加个符号。

# 四、特征值与特征向量

## 1、引入特征值

**例子**：在一个小镇上，每年有30%的已婚女性离婚，又有20%的未婚女性结婚。然后初值是8000已婚女性，2000单身女性，同时，女性总数不会发生变化。

令$w_0 = \begin{bmatrix} 8000 \\ 2000 \end{bmatrix}$

$w_1 = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} w_0 = \begin{bmatrix} 6000 \\ 4000 \end{bmatrix}$

$w_2 = Aw_1 = \begin{bmatrix} 5000 \\ 5000 \end{bmatrix} = A^2 w_0$

而$n$年后的数据，可以用$w_n = A^n w_0$来计算。

而$w_{10} = \begin{bmatrix} 4004 \\ 5996 \end{bmatrix}, \cdots, w_{20} = \begin{bmatrix} 4000 \\ 6000 \end{bmatrix}, \cdots, w_{100} \approx \begin{bmatrix} 4000 \\ 6000 \end{bmatrix}$

也就是趋近于稳态了。

那假设$x_1 = [2 \;\; 3]^T, Ax_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$，还有$x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, Ax_2 = \dfrac{1}{2} x_2$

$w_0 = \begin{bmatrix} 8000 \\ 2000 \end{bmatrix} = 2000x_1 - 4000x_2$

$w_1 = Aw_0 = A(2000x_1 - 4000x_2) = 2000Ax_1 - 4000Ax_2 = 2000x_1 - 4000(\frac{1}{2}x_2)$

$w_2 = Aw_1 = A(2000x_1 - 4000(\frac{1}{2}x_2)) = 2000x_1 - 4000(\frac{1}{2^2}x_2)$

故，$w_n = A^n w_0 = 2000x_1 - 4000(\frac{1}{2^n}x_2)$

**定义**：对于一个$n \times n$的矩阵$A$，如果$Ax = \lambda x$，且$x$不是零向量，那么称λ是特征值，$x$是对应的特征向量。

那如何求解特征值和特征向量呢？有如下推断：

$$\begin{aligned} &\text{首先，如果一个矩阵有特征向量，那根据定义，一定有：} Ax = \lambda x \\ &\iff (A - \lambda I)x = 0 \\ &\iff N(A - \lambda I) \neq \{0\} \\ &\iff A\text{是不可逆/奇异矩阵} \\ &\iff detA = 0 \end{aligned}$$

**栗子**：

$$\begin{aligned} A &= \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \\ det(A - \lambda I) &= \begin{vmatrix} 0.7 - \lambda & 0.2 \\ 0.3 & 0.8 - \lambda \end{vmatrix} \\ &= (\lambda - 0.7)(\lambda - 0.8) - 0.06 \\ &= \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} \\ &= (\lambda - 1)(\lambda - \frac{1}{2}) = 0 \\ \lambda &= 1\text{或}\frac{1}{2} \end{aligned}$$

$$\text{当}\lambda = 1\text{时，} A - I = \begin{bmatrix} -0.3 & 0.2 \\ 0.3 & -0.2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.3 & -0.2 \\ 0 & 0 \end{bmatrix}$$

$$\text{可以知道，此处的特征向量} x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\text{当}\lambda = \frac{1}{2}\text{时，} A - \frac{1}{2}I = \begin{bmatrix} 0.2 & 0.2 \\ 0.3 & 0.3 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\text{此时，} \lambda = \frac{1}{2}\text{的特征向量为} x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

**栗子**：

$$A=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{bmatrix}$$
$$A^T=A,A^2=A(\text{这是一个投影矩阵})$$
$$det(A-\lambda I)=0$$
$$\begin{vmatrix}\frac{1}{2}-\lambda&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}-\lambda\end{vmatrix}=0$$
$$(\lambda-\frac{1}{2})^2-\frac{1}{4}=0$$
$$\lambda(\lambda-1)=0$$
$$\lambda=0\text{或}1$$
$$\text{对于}\lambda=1,A-\lambda I=\begin{bmatrix}-\frac{1}{2}&\frac{1}{2}\\0&0\end{bmatrix}$$
$$\text{对应的特征向量为}x=\begin{bmatrix}1\\1\end{bmatrix}$$
$$\text{对于}\lambda=0,A=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{bmatrix}\Rightarrow\begin{bmatrix}1&1\\0&0\end{bmatrix}$$
$$\text{对应的特征向量为}x=\begin{bmatrix}-1\\1\end{bmatrix}$$

行列式通常被称为特征多项式。

**栗子**:

$$Q=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$$
$$det(Q-\lambda I)=\begin{vmatrix}-\lambda&1\\-1&-\lambda\end{vmatrix}=\lambda^2+1=0$$
$$\lambda=\pm i(\text{没有实特征值和特征向量，因为这是一个旋转矩阵，直观来看，旋转以后不可能保持原方向})$$

上面这个栗子说明了，如果一个矩阵可能会存在没有特征值和特征向量的情况。且，因为有共轭复数的存在，因此复根肯定是成对出现的。

如果$Ax=\lambda x,$那么，$A^2x=\lambda Ax=\lambda^2x$，不难发现，如果是λ是$A$的特征值，那么，对于$A^2$来说，其特征值就是$\lambda^2$。更进一步，对于$A^n$而言，特征值是$\lambda^n$。同时，它们对应的特征向量是一样的。

**栗子**:

$$A=\begin{bmatrix}0.7&0.2\\0.3&0.8\end{bmatrix}$$
$$\lambda_1=1,\lambda_2=\frac{1}{2}$$
$$\begin{cases}\lambda_1+\lambda_2=\frac{3}{2}\\\lambda_1\lambda_2=\frac{1}{2}\end{cases}$$
$$\text{然后，}A\text{的主对角线元素相加是}1.5(trace)\text{，行列式是}0.5$$

**性质**: （Ⅰ）$\lambda_1\lambda_2\cdots\lambda_n=detA$

（Ⅱ）$\sum_{i=1}^n\lambda_i=trace(A)$

**证明**:

（Ⅰ）

$$det(A-\lambda I)=\begin{vmatrix}a_{11}-\lambda&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}-\lambda&&a_{2n}\\\vdots&&\ddots&\\a_{n1}&&&a_{nn}\end{vmatrix}$$
$$=(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda)$$

当$\lambda=0$时，就是行列式的值，因此，$detA=\prod_{i=1}^n\lambda_i.\square$

（Ⅱ）

$$det(A-\lambda I)=\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & & & a_{nn} \end{vmatrix}$$
$$=(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda)$$

如果要选到特征值相加的项的话，一定是取在对角线这一条上面的，且多项式的次数为$n-1$次。

而在原本的行列式中，包含$\lambda^{n-1}$的项也一定在对角线元素相乘的这一项中取到，也就是$(a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda)$。根据多项式乘法的性质，可以得到$\sum_{i=1}^{n} a_{ii}=trace(A)=\sum_{i=1}^{n}\lambda_i.\square$

## 2、矩阵的对角线化

接下来，从矩阵乘法的角度出发，我们希望能够简单地获得$A^n$的值，那么根据特征值的含义，$AS=S\Lambda$，$S$的列向量是特征向量，$\Lambda$是特征值矩阵，对角线上是特征向量对应的特征值。假如说$S$是可逆的，那么会有$A=S\Lambda S^{-1}$，那$A^n$就可以很容易地表示成$S\Lambda^n S^{-1}$。

因此，假设$n\times n$的矩阵$A$有$n$个**线性无关**的特征向量$x_1,x_2,\cdots,x_n$。

$$\begin{cases} Ax_1=\lambda_1x_1 \\ Ax_2=\lambda_2x_2 \\ \quad\vdots \\ Ax_n=\lambda_nx_n \end{cases}$$

$$令S=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

$$可以得到AS=S\Lambda\Rightarrow S^{-1}AS=\Lambda(或A=S\Lambda S^{-1})$$

那么$A^2=AA=(S\Lambda S^{-1})(S\Lambda S^{-1})=S\Lambda^2S^{-1}$，以及$A^k=S\Lambda^kS^{-1}$

**栗子**:

$$A=\begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}$$

$$\lambda_1=1,x_1=\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\lambda_2=\frac{1}{2},x_2=\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$A=\begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}$$

$$A^k=\begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & (\frac{1}{2})^k \end{bmatrix}\begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}$$
$$=\begin{bmatrix} \frac{2}{5}+\frac{3}{5}(\frac{1}{2})^k & \frac{2}{5}-\frac{2}{5}(\frac{1}{2})^k \\ \frac{3}{5}-\frac{3}{5}(\frac{1}{2})^k & \frac{3}{5}+\frac{2}{5}(\frac{1}{2})^k \end{bmatrix}$$

$$当k\rightarrow+\infty时，A^k=\begin{bmatrix} \frac{2}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{3}{5} \end{bmatrix}$$

$$故，如果存在一个向量x=\begin{bmatrix} a \\ b \end{bmatrix}，A^kx=\begin{bmatrix} \frac{2}{5}(a+b) \\ \frac{3}{5}(a+b) \end{bmatrix}$$

**性质**：如果特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$均不相等，那么，对应的特征向量$x_1,x_2,\cdots,x_n$是线性无关的。

**证明**:

$$\text{假设}c_1x_1+c_2x_2+\cdots+c_nx_n=0(\text{Ⅰ})$$
$$\Rightarrow A(c_1x_1+c_2x_2+\cdots+c_nx_n)=0$$
$$\Rightarrow c_1\lambda_1x_1+c_2\lambda_2x_2+\cdots+c_n\lambda_nx_n=0(\text{Ⅱ})$$
$$\text{给(Ⅰ)乘一个}\lambda_n\text{，再减去(Ⅱ)，会有：}$$
$$c_1(\lambda_1-\lambda_n)x_2+c_3(\lambda_1-\lambda_n)x_3+\cdots+c_{n-1}(\lambda_1-\lambda_{n-1})x_{n-1}=0(\text{Ⅲ})$$
$$\text{然后左乘一个}A\text{，就得到了}c_1\lambda_1(\lambda_1-\lambda_n)+c_2\lambda_2(\lambda_2-\lambda_n)+\cdots+c_{n-1}\lambda_{n-1}(\lambda_{n-1}-\lambda_n)x_{n-1}=0(\text{Ⅳ})$$
$$\text{再(Ⅲ)成一个}\lambda_{n-1}\text{，再和(Ⅳ)减一减，就只剩下}n-2\text{项了}$$
$$\text{这样反复消去，最后只会留下一项：}c_1(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)\cdots(\lambda_1-\lambda_n)x_1=0$$
$$\text{又因为}(\lambda_1-\lambda_i)(i\neq 1)\text{都不为0，因此，}c_1\text{为0}$$
$$\text{再重复}n-1\text{遍上述操作，}c_2\text{，}c_3,\cdots,c_n\text{也都为0}$$
$$\text{因此，这些向量线性无关.}\square$$

问：什么时候$A^k$会趋近于0矩阵？

答：$|\lambda|<1$的时候

前面提到的那个性质是矩阵可对角线化的一个充分条件，现在来看看在什么情况下矩阵可以对角化。

**性质**：当$n\times n$的矩阵$A$有$n$个线性无关的特征向量时，那矩阵$A$就是可以对角线化的。

---

**证明**：

$$\begin{aligned}&\Rightarrow \text{已经证明过}\\&\Leftarrow \text{因为矩阵}A\text{能够对角线化，因此有}S^{-1}AS=\Lambda\Rightarrow AS=S\Lambda\\&\text{令}S=[x_1\quad x_2\quad\cdots\quad x_n]\\&A\,[x_1\quad x_2\quad\cdots\quad x_n]=[x_1\quad x_2\quad\cdots\quad x_n]\begin{bmatrix}\lambda_1&&&\\&\lambda 2&&\\&&\ddots&\\&&&\lambda_n\end{bmatrix}\\&\text{可以得到}Ax_i=\lambda_ix_i(i=1,2,\cdots,n)\\&\text{根据定义，不但发现}x_i\text{就是特征向量；又假设}x_i\text{线性无关，所以}S\text{有}n\text{个无关的特征向量}\end{aligned}$$

对于一个特征值λ来说，他有两个描述性质的数：

（Ⅰ）代数重数（AM）：特征多项式解完以后，这个特征值有几重根

（Ⅱ）几何重数（GM）：特征值对应的特征向量有几个，也可以理解成$A-\lambda I$零空间的基的个数（$N(A-\lambda I)$也被称为λ对应的特征空间)

**性质**：对于每个不同的特征根，GM≤AM

**性质**：如果一个$n\times n$的矩阵是能够对角线化的，满足两条性质：

$$\begin{aligned}&(1)\text{、}\sum_{i=1}^{n}AM_i=n\\&(2)\text{、}\sum_{i=1}^{n}GM_i=n\text{，因为}GM_i\leq AM_i,\text{因此，}GM_i=AM_i\end{aligned}$$

**栗子**：

$$A=\begin{bmatrix}2&0&0\\-1&4&0\\-3&6&2\end{bmatrix},B=\begin{bmatrix}2&0&0\\0&4&0\\1&0&2\end{bmatrix}$$
$$\text{由于}A\text{和}B\text{都是下三角矩阵，所以说，其特征方程就是对角线相乘}$$
$$\text{因此，}det(A-\lambda I)=det(B-\lambda I)=(\lambda-2)^2(\lambda-4)$$
$$\text{对于矩阵}A\text{而言，当}\lambda=2\text{时，}AM=2,GM=2;\lambda=4\text{时，}AM=1,GM=1,\text{故可对角线化}$$
$$\text{对于矩阵}B\text{而言，当}\lambda=2\text{时，}AM=2\text{，}GM=2;\lambda=4\text{时，}AM=1,GM=1\text{，故可对角线化}$$

**——对角化应用—解差分方程**

斐波那契数列：$F_{n+2} = F_{n+1} + Fn, n = 0\{F_1 = 0, F_1 = 1\}$

令$u_k = \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$

可以将斐波那契数列表示为矩阵形式：

$$u_{k+1} = \begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = Au_k$$

这时就变成了一个线性代数问题。

$$u_{k+1} = A^k u_k$$

所以，只要解$A^k$即可，也就是用到对角化。

$$\begin{aligned}
det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} \\
&= \lambda^2 - \lambda - 1 \\
\lambda &= \frac{1 \pm \sqrt{5}}{2} \\
A - \lambda_1 I &= \begin{bmatrix} \frac{1-\sqrt{5}}{2} & 1 \\ 1 & -\frac{1+\sqrt{5}}{2} \end{bmatrix} \\
x_1 &= \begin{bmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{bmatrix} \\
A - \lambda_2 I &= \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 1 \\ 1 & -\frac{1-\sqrt{5}}{2} \end{bmatrix} \\
x_2 &= \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix} \\
\text{经过对角线化，} A &= S\Lambda S^{-1} \\
&= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \\
\text{如果我们要计算} u_k \text{，就需要先把} u_0 \text{写成} x_1 \text{和} x_2 \text{的线性组合} \\
u_0 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\
\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
\text{故} u_0 &= \frac{1}{\lambda_1 - \lambda_2}(x_1 - x_2) \\
u_k &= A^k u_0 = A^k \frac{1}{\lambda_1 - \lambda_2}(x_1 - x_2) \\
&= \frac{1}{\lambda_1 - \lambda_2}(\lambda_1^k x_1 - \lambda_2^k x_2) \\
F_k &= \frac{1}{\lambda_1 - \lambda_2}(\lambda_1^k - \lambda_2^k) \\
&= \frac{1}{\sqrt{5}}[(\frac{1+\sqrt{5}}{2})^k - (\frac{1-\sqrt{5}}{2})^k] \\
\\
\text{然后，由于} (\frac{1-\sqrt{5}}{2})^k \text{在} k \to +\infty \text{时趋近于} 0 \\
\text{因此，可以用} \frac{1}{\sqrt{5}}(\frac{1+\sqrt{5}}{2})^k \text{来近似} F_k \text{，它是最接近} F_k \text{是最接近近似值的整数} \\
\text{然后，} \frac{F_k}{F_{k-1}} \text{非常接近} \frac{1+\sqrt{5}}{2}(k \to \infty)
\end{aligned}$$

更一般地，差分方程可以写成：

$$G_{k+n} = a_1 G_{k+n-1} + a_2 G_{k+n-2} + \cdots + a_n G_k [n\text{阶线性齐次常系数差分方程}]$$

写成矩阵形式：

$$\begin{bmatrix} G_{k+n} \\ G_{k+n-1} \\ \vdots \\ G_{k+1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ 1 & 0 & & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \\ 0 & & & 1 & 0 \end{bmatrix} \begin{bmatrix} G_{k+n-1} \\ G_{k+n-2} \\ \vdots \\ G_k \end{bmatrix} = Au_k$$

为了使得差分方程有定义，还应当有$n$个初值$G_1, G_2, \cdots, G_{n-1}$

如果要找到差分方程的解，就需要求A的对角线化形式

经过对角化，矩阵$A \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

我们再把$u_0$写成特征向量的线性组合：$u_0 = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

$u_k = A^ku_0 = A^k(c_1x_1 + c_2x_2 + \cdots + c_nx_n) = c_1\lambda_1^kx_1 + c_2\lambda_2^kx_2 + \cdots + c_n\lambda_n^kx_n$

$G_k$就是$u_k$的第$n$个元素

**——对角化应用—解微分方程**

$$\begin{bmatrix} \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix} = \frac{du}{dt} = Au = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}, \text{初值}: \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$
$$\frac{dy}{dt} = -2y + z$$
$$\frac{dz}{dt} = y - 2z$$

对于高等数学中学过的n阶常系数齐次线性微分方程，可以知道，需要求解特征方程得到特征根$\lambda_1, \lambda_2, \cdots \lambda_n$，如果n个特征根相异，可以得到方程的通解：$y(x) = c_1e^{\lambda_1 x} + c_2e^{\lambda_2 x} + \cdots + c_ne^{\lambda_n x}$

对于矩阵形式的常系数齐次线性微分方程也是这样，先求解矩阵的特征根，然后找到对应的特征向量，可以将解写作：$u = c_1e^{\lambda_1 t}x_1 + c_2e^{\lambda_2 t}x_2$

求解一开始提到的那个例子：

$$det(A - \lambda I) = \begin{vmatrix} -(2+\lambda) & 1 \\ 1 & -(2+\lambda) \end{vmatrix}$$
$$= \lambda^2 + 4\lambda + 3 = 0$$
$$\text{不难得到}, \lambda = -3\text{或} - 1$$
$$\text{对于}\lambda_1 = -1, x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$\text{对于}\lambda_2 = -3, x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
$$\text{故微分方程的解为}: u = c_1e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2e^{-3t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
$$\text{又初值为:} \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \text{解得} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$\text{故}u = e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{-3t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-3t} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= Se^{\Lambda t}c$$

在微积分中，$e^x$的泰勒展开式为：$1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$

同样的，使用矩阵同样也能定义同样的东西：

$$\begin{aligned}e^{At} &\triangleq I+At+\frac{1}{2!}(At)^2+\frac{1}{3!}(At)^3+\cdots \\ \frac{d}{dt}e^{At} &= A+A^2t+\frac{1}{2!}A^3t^2+\cdots \\ &= A[I+At+\frac{1}{2!}(At)^2+\cdots] \\ &= Ae^{At}\end{aligned}$$

因此，$u=e^{At}c$，是$\frac{du}{dt}=Au$的解。

所以说，现在要来看一下$e^{At}$是怎么对角化的

$$\begin{aligned}e^{At} &= I+At+\frac{1}{2!}(At)^2+\frac{1}{3!}(At)^3+\cdots \\ &= I+S\Lambda S^{-1}+\frac{1}{2!}S\Lambda^2S^{-1}+\frac{1}{3!}S\Lambda^3S^{-1}+\cdots \\ &= S[I+\Lambda+\frac{1}{2!}\Lambda^2+\frac{1}{3!}\Lambda^3+\cdots]S^{-1} \\ &= Se^{\Lambda t}S^{-1}\end{aligned}$$

因此，$u=e^{At}u_0=Se^{\Lambda t}S^{-1}u_0=Se^{\Lambda t}c$

其中，

$$\begin{aligned}e^{\Lambda t} &= I+\Lambda t+\frac{1}{2}\Lambda^2t+\cdots \\ &= \begin{bmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix}+\begin{bmatrix}\lambda_1t & & & \\ & \lambda_2t & & \\ & & \ddots & \\ & & & \lambda_nt\end{bmatrix}+\begin{bmatrix}\frac{1}{2}\lambda_1^2t & & & \\ & \frac{1}{2}\lambda_2^2t & & \\ & & \ddots & \\ & & & \frac{1}{2}\lambda_n^2t\end{bmatrix}+\cdots \\ &= \begin{bmatrix}e^{\lambda_1t} & & & \\ & e^{\lambda_2t} & & \\ & & \ddots & \\ & & & e^{\lambda_nt}\end{bmatrix}\end{aligned}$$

而$Se^{\Lambda t}c=c_1e^{\lambda_1t}x_1+c_2e^{\lambda_2t}x_2+\cdots+c_ne^{\lambda_nt}x_n$

**栗子**：$y^{'''}-3y^{''}+2y^{'}=0$

$$令w=y^{''},v=y^{'},u=\begin{bmatrix}w\\v\\y\end{bmatrix}$$

$$\begin{cases}w^{'}=y^{'''}=3y^{''}-2y^{'}=3w-2v \\ v^{'}=y^{''}=w \\ y^{'}=y^{'}=v\end{cases}$$

$$故，可以写成：\frac{du}{dx}=\begin{bmatrix}w^{'}\\v^{'}\\y^{'}\end{bmatrix}=\begin{bmatrix}3 & -2 & 0\\1 & 0 & 0\\0 & 1 & 0\end{bmatrix}\begin{bmatrix}w\\v\\y\end{bmatrix}$$

$$然后算下去就行了。$$

**——微分方程解的稳定性（？）**

如果$\lambda=a+bi$，$e^{\lambda t}=e^{at}e^{bti}=e^a(cosbt+isinbt)$

$|e^{\lambda t}|=e^{at}$

因此，当$Re(\lambda_i)<0$时，$\frac{du}{dt}=Au$是**稳定**（stable）的且满足$e^{At}\to 0(t\to\infty)$

当$Re(\lambda_i)\le 0$时，称微分方程是neutral的。

当$\exists i,Re(\lambda_i)>0$时，称微分方程时不稳定(unstable)的。

### 3、对称矩阵的性质

现在，我们有一个对称矩阵$A$，即$A^T=A$，具有以下性质：

（Ⅰ）对称矩阵只具有实特征值。

（Ⅱ）特征向量一定存在，并且可以是单位正交（orthonormal）的，即$A=Q\Lambda Q^{-1}=Q\Lambda Q^T, Q$是orthogonal matrix。而这个性质又被称为谱定理。

**栗子**：

$$
\begin{aligned}
A=\begin{bmatrix}1&2\\2&4\end{bmatrix}\\
\lambda_1=0, x_1=\frac{1}{\sqrt{5}}\begin{bmatrix}-2\\1\end{bmatrix}\\
\lambda_1=5, x_1=\frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}\\
\text{有}x_1^Tx_2=0\\
\text{因此，有}A&=Q\Lambda Q^T\\
&=\frac{1}{\sqrt{5}}\begin{bmatrix}2&1\\-1&2\end{bmatrix}\begin{bmatrix}0&0\\0&5\end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix}2&-1\\1&2\end{bmatrix}
\end{aligned}
$$

**性质**：实对称矩阵的特征值都是实数。

**证明**：

考虑特征值的原始定义：$Ax=\lambda x$

假如我要证明λ是实数，那我就要证明它的虚部是0，实际上就是要证明它和它的共轭是相等的，即$\bar{A}\bar{x}=\bar{\lambda}\bar{x}$。

由于$A$是实矩阵，故$A\bar{x}=\bar{\lambda}\bar{x}$。两边取转置$\bar{x}^TA=\bar{\lambda}\bar{x}^T$

然后，就有$\bar{\lambda}\bar{x}^Tx=\bar{x}^TAx=\lambda\bar{x}^Tx$

由于$\bar{x}^Tx$是大于0的实数，故可以得出$\bar{\lambda}=\lambda$，故特征值都是实数.□

**性质**：实对称矩阵不同特征值对应的特征向量一定是正交的。

**证明**：

假设$Ax_1=\lambda_1x_1$和$Ax_2=\lambda_2x_2$

对左边那个式子取转置得$x_1^TA=\lambda_1x_1^T$

两边再右乘一个$x_2$，得到$x_1^TAx_2=\lambda_1x_1^Tx_2$

$\Rightarrow \lambda_2x_1^Tx_2=\lambda_1x_1^Tx_2$

由于$\lambda_2\neq\lambda_1$，故只能得到$x_1^Tx_2=0$，即$x_1\perp x_2$

因此，不同特征值对应的特征向量一定是正交的。□

当然，只知道不同特征值对应的特征向量正交并不足以证明谱定理。因为，谱定理暗含了一个对称矩阵，一定有足够多的特征值(n个)。

对于一个$3\times 3$的对称矩阵而言，有：

$$
A=Q\Lambda Q^T=\begin{bmatrix}\\x_1&x_2&x_3\\\\\end{bmatrix}\begin{bmatrix}\lambda_1&&\\&\lambda_2&\\&&\lambda_3\end{bmatrix}\begin{bmatrix}&x_1^T&\\&x_2^T&\\&x_3^T&\end{bmatrix}=\lambda_1x_1x_1^T+\lambda_2x_2x_2^T+\lambda_3x_3x_3^T
$$

如果$\lambda_1=\lambda_2$，会有$A=\lambda_1(x_1x_1^T+x_2x_2^T)+\lambda_3x_3x_3^T$

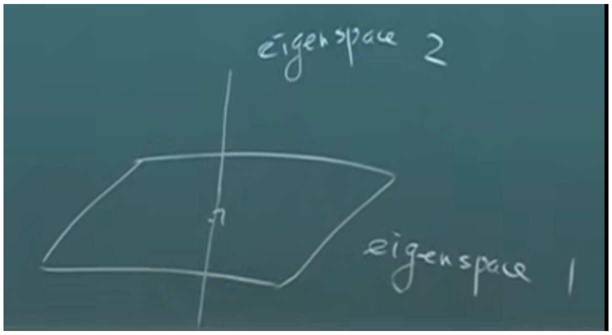
此处联想前面学的投影。对于$\mathbb{R}^n$上的子空间$P$的一组单位正交基$q_1,q_2,\ldots,q_m(m\leq n)$。存在$R^n$上的向量$b$，要将$b$投影到$P$上，可以得到投影向量$p=q_1(q_1^Tb)+q_2(q_2^Tb)+\cdots+q_m(q_m^Tb)$。将$b$提出来，可以得到$p=(q_1q_1^T+q_2q_2^T+\cdots+q_mq_m^T)b=\mathbb{P}b$

故，投影矩阵为$\mathbb{P}=q_1q_1^T+q_2q_2^T+\cdots+q_mq_m^T$

因此，再回到上面那个$3\times3$矩阵的例子，当$\lambda_1=\lambda_2$时，可以知道的是，矩阵A可以分为两个部分，一个是$\lambda_1$对应的两个特征向量张成的特征子空间（即$N(A-\lambda_1 I)$），还有$\lambda_3$对应的一个特征向量张成的特征子空间（即$N(A-\lambda_3 I)$）。

可以写成$A=\lambda_1P_1+\lambda_3P_2$

从图上来看，就是这样的：




广义来看，如果一个实矩阵$A$有$k$个不同的特征值$\lambda_1,\lambda_2,\cdots,\lambda_k$，那它就可以写成：$A=\lambda_1P_1+\lambda_2P_2+\cdots+\lambda_kP_k$，其中，$P_i$是投影到$\lambda_i$的特征空间上的投影矩阵；而这种分解方法被称为谱分解。

**栗子**：

$$A=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}$$

$$\lambda_1=1,\text{对应的特征向量是}x_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix},\text{投影矩阵是}P_1=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$\lambda_2=3,\text{对应的特征向量是}x_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix},\text{投影矩阵是}P_2=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$A=P_1+3P_2$$

**性质**：对于实矩阵而言，复特征根和特征向量都是以共轭对的形式出现的。

**证明**：假设$Ax=\lambda x$，给矩阵取下共轭，有$\bar{A}\bar{x}=\bar{\lambda}\bar{x},\text{推出}A\bar{x}=\bar{\lambda}\bar{x}.\square$

**性质**：如果$A$是实对称矩阵，则非零特征值的个数=rank=非零主元的个数

**证明**：$A$是实对称矩阵且可对角化的，那么，0特征值的代数重数=0特征值的几何重数=N(A)的维数=n-rank(A)。□

**舒尔定理**：对于任意的正方形矩阵，都可以分解成$A=QTQ^T$，$T$是一个上三角矩阵，$\bar{Q}^T=Q^{-1}$；如果$A$有实特征值，则$Q$和$T$可以是实矩阵，即$Q^T=Q^{-1}$（如果$\bar{Q}^T=Q^{-1}$，则称$Q$为酉矩阵）

**证明**：（数学归纳法）

首先，当$n=1$时，成立$a=1\cdot a\cdot 1^{-1}$。

假设当$k=n$时，$n\times n$矩阵$A$满足舒尔定理$A=QTQ^T=QT\bar{Q}^{-1}$。

当$k=n+1$时，对于矩阵$A$来说，存在特征值$\lambda_1$和对应的特征向量$q_1$。（有特征多项式，如果没有实根，那就一定有复根，且是以共轭特征根的形式出现）

然后，使用$q_1$对矩阵$A$进行施密特正交化，可以得到一组单位正交向量$q_2,q_3,\cdots,q_{n+1}$，而这组向量是在$\mathcal{C}^{n+1}$的一组单位正交基。

可以令$Q_1=[q_1 \quad q_2 \quad \cdots \quad q_{n+1}]$，且$Q_1$为酉矩阵。

因此，有：

$$Q_1^H A Q_1 = \begin{bmatrix} \bar{q}_1^T \\ \bar{q}_2^T \\ \vdots \\ \bar{q}_{n+1}^T \end{bmatrix} [Aq_1 \quad Aq_2 \quad \cdots \quad Aq_{n+1}]$$

$$= \begin{bmatrix} \bar{q}_1^T \\ \bar{q}_2^T \\ \vdots \\ \bar{q}_{n+1}^T \end{bmatrix} [\lambda_1 q_1 \quad Aq_2 \quad \cdots \quad Aq_{n+1}] (q_1\text{是特征向量})$$

$$= \begin{bmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix} = T$$

由于$A_2$是一个$n \times n$的矩阵，因此，可以对其进行分解，得到$A_2 = Q_2 T_2 Q_2^T$，其中$Q_2$是酉矩阵，$T_2$是上三角矩阵。

令$Q = Q_1 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2 & \\ 0 & & & \end{bmatrix}$

有$Q\bar{Q}^T = Q_1 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2 & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \bar{Q}_2^T & \\ 0 & & & \end{bmatrix} \bar{Q}_1^T = I$

故$Q$也是酉矩阵。

因此，对$Q_1^H A Q_1$左乘$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2^H & \\ 0 & & & \end{bmatrix}$，右乘$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2 & \\ 0 & & & \end{bmatrix}$

能够得到：

$$Q^H A Q = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2^H & \\ 0 & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2 & \\ 0 & & & \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_2^H & \\ 0 & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & & & \\ \vdots & & A_2 Q_2 & \\ 0 & & & \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & & & \\ \vdots & & Q_2^H A_2 Q_2 & \\ 0 & & & \end{bmatrix} = \begin{bmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & & & \\ \vdots & & T_2 & \\ 0 & & & \end{bmatrix}$$

由于$T_2$是上三角矩阵，因此，$T$也是上三角矩阵，故得证，在$k = n+1$时，能够将矩阵$A$分解为$A = QT\bar{Q}^T$。

根据数学归纳法，舒尔分解对于任意$k$均成立。□

然后，可以用**舒尔定理**证明谱分解定理。

**证明**：由于实对称矩阵只有实特征值，因此，根据Schur Theorem，$A = QTQ^T$，而$Q$是正交矩阵。

两边乘矩阵的逆，可以得到$Q^T AQ = T$，由于$A$是实对称矩阵，因此，取转置以后应该和原来一样，$Q^T AQ = T^T$，因此，$T^T = T$，又$T$是上三角矩阵，故$T$是对角矩阵；因此，可以将实对称矩阵$A$分解成$A = Q\Lambda Q^T$.□

## 4、正定矩阵

**栗子**：

$$A=\begin{bmatrix}1 & -2\\ -2 & 6\end{bmatrix}$$
$$\begin{aligned}x^TAx&=\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}1 & -2\\ -2 & 6\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}\\ &=x_1^2-4x_1x_2+6x_2^2\end{aligned}$$

**另一个栗子**：

$$B=\begin{bmatrix}1 & -1\\ -3 & 6\end{bmatrix}$$
$$\begin{aligned}x^TBx&=x_1^2-4x_1x_2+6x_2^2\\ &=x^TAx\end{aligned}$$

$A$和$B$的二次型是一样的，而$A$作为实对称矩阵，性质更好，一定能找到$n$个正交的特征向量，因此，从二次型的角度考虑，一般会从使用实对称矩阵进行分析。

对于一个不对称的矩阵$B$，如果要分析二次型，那么可以取$A=\frac{1}{2}(B+B^T)$。

**定义**：当实对称矩阵$A$的二次型$x^TAx>0$时（不包含零向量），则称矩阵$A$是正定的；反之，如果二次型恒小于0，则称矩阵$A$是负定的。

**性质**：对于一个实对称矩阵$A$而言，以下几条性质是等价的：

（Ⅰ）对于任意非零向量$x$，$x^TAx>0.$

（Ⅱ）特征值均为正

（Ⅲ）$n$个$upper\ left$ 行列式都是正的。

（Ⅳ）所有主元都是正的。

（Ⅴ）一定有分解$A=R^TR$，$R$有线性无关的列

**证明**：

（Ⅰ）$\Rightarrow$（Ⅱ）

假设$\lambda_i$是一个特征值，$x_i$是其对应的特征向量（已经单位化）。会有$Ax_i=\lambda_ix_i$。那么$x_i^TAx_i=\lambda_ix_i^Tx_i=\lambda_i>0.\square$

（Ⅱ）$\Rightarrow$（Ⅰ）

由于实对称矩阵的特征值均为正，而又有$A=\lambda_1P_1+\lambda_2P_2+\cdots+\lambda_nP_n$，即任何向量左乘$A$都是用投影到特征空间上向量的线性组合。因此，设特征值为$\lambda_i$，对应的单位化的特征向量为$x_i$。而任意向量$x$均可以表示为$x=c_1x_1+c_2x_2+\cdots+x_n$。

$x^TAx=(c_1x_1^T+c_2x_2^T+\cdots+c_nx_n^T)A(c_1x_1+c_2x_2+\cdots+c_nx_n)=c_1^2\lambda_1+c_2^2\lambda_2+\cdots+c_n^2\lambda_n>0$

因此，每一个特征向量对应的二次型都是正的，那么，经过线性组合以后，它们的二次型还会是正的。$\square$

（Ⅰ）$\Rightarrow$（Ⅲ）

对于$A$而言，因为$\lambda_i>0$，又$|A|=\lambda_1\lambda_2\cdots\lambda_n$，故$|A|>0.$

因此，只需要证明任意的左上子矩阵$A_k$都是正定矩阵，那么一定满足$|A_k|>0$。

只需要说明$x_k^TA_kx_k>0$，

因为$x^TAx>0$，要说明$x_k^TA_kx_k$大于0，只需要给向量$x$的下$n-k$的元素取0即可。

即$x^TAx=\begin{bmatrix}x_k^T & 0 & \cdots & 0\end{bmatrix}A\begin{bmatrix}x_k\\ 0\\ \vdots\\ 0\end{bmatrix}>0$

故$A_k$是正定的，因此$|A_k|>0.\square$

（Ⅲ）$\Rightarrow$（Ⅳ）

由于（Ⅲ），可以知道矩阵$A$每行不需要经过行交换就能存在主元

且，有$p_k = \frac{detA_k}{detA_{k-1}}$

由于$p_1 > 0$，因此，根据归纳法，后面的每一个$p_k > 0$，故一定存在$n$个正主元.□

（Ⅳ）⇒（Ⅰ）

不经过行交换，一定能够将矩阵$A$分解成为$A = LDU$，由于$A$是实对称矩阵，因此，$A = LDL^T$，其中$D$是对角线矩阵，对角线上是主元，且$L$是下三角矩阵。那么

$$x^TAx = x^TLDL^Tx = (L^Tx)^TD(L^Tx) = \begin{bmatrix} l_1^Tx & l_2^Tx & \cdots & l_n^Tx \end{bmatrix} \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{bmatrix} \begin{bmatrix} l_1^Tx \\ l_2^Tx \\ \vdots \\ l_n^Tx \end{bmatrix} = p_1(l_1^Tx)^2 + p_2(l_2^Tx)^2 + \cdots + p_n(l_n^Tx)^2 >$$

（Ⅴ）⇒（Ⅰ）

$A = R^TR$，则$x^TR^TRx = (Rx)^TRx = ||Rx||^2$

由于$R$有线性无关的列向量，故$Rx$不会是0（$x$不是零向量），故$x^TAx = ||Rx||^2 > 0.\square$

（Ⅰ）⇒（Ⅴ）

有几种分法：

（i）由于A是实对称矩阵，因此，可以分解为$A = Q\Lambda Q^T = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^T = (Q\sqrt{\Lambda})(Q\sqrt{\Lambda})^T$

故$A$可以分解为$A = RR^T$，其中$R = Q\sqrt{\Lambda}$

（ii）$A = LDL^T$，由于$A$是实对称矩阵，$D$是主元，因为$D$的主元全为正，因此可以拆成$A = L\sqrt{D}\sqrt{D}L^T = SS^T$

因为$L$是下三角矩阵，$\sqrt{D}$是对角线矩阵，因此，$S$也是一个下三角矩阵。而将矩阵分解成$SS^T$，这种分解方法又被称为Cholesky分解。

（iii）给定任意一个符合条件的$R$

现在给$R$乘上一个正交矩阵$Q$(不一定要正方形矩阵)

有$A = (QR)^TQR = R^TQ^TQR = R^TIR = R^TR = A.\square$

**栗子**：

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

$$\begin{aligned} x^TAx &= 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2 \\ &= 2(x_1^2 - x_1x_2 + \frac{1}{4}x_2^2) + \frac{3}{2}x_2^2 - 2x_2x_3 + 2x_3^2 \\ &= 2(x_1 - \frac{1}{2}x_2)^2 + \frac{3}{2}(x_2^2 - \frac{4}{3}x_2x_3 + \frac{4}{9}x_3^2) + \frac{4}{3}x_3^2 \\ &= 2(x_1 - \frac{1}{2}x_2)^2 + \frac{3}{2}(x_2 - \frac{2}{3}x_3)^2 + \frac{4}{3}x_3^2 > 0 \end{aligned}$$

特征值为：$2-\sqrt{2}, 2, 2+\sqrt{2}$

**定义**🏉:对于实对称矩阵$A$而言，如果有任意非零向量$x$，有$x^TAx \geq 0$，则称矩阵是半正定的。

**性质**：对于半正定矩阵而言，以下性质是等价的：

（Ⅰ）对于任意向量$x$，有$x^TAx \geq 0$

（Ⅱ）所有的特征值大于等于0

（Ⅲ）所有的左上行列式都是大于等于0的

（Ⅳ）所有主元大于等于0

（Ⅴ）$A = RR^T$，且列向量可能是相关的。

在概率论中，协方差矩阵就是半正定矩阵。

**栗子**：找到斜椭圆的长短轴。

$5x^2+8xy+5y^2=1$

令$A=\begin{bmatrix}5 & 4\\4 & 5\end{bmatrix}$，$x^TAx=5x^2+8xy+5y^2, x=\begin{bmatrix}x\\y\end{bmatrix}$

因为要找到它的长短轴，所以说需要先给它对角线化，使得只有对角线上才有元素。

由于谱定理，一定可以有分解$A=Q\Lambda Q^T=\dfrac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}9 & 0\\0 & 1\end{bmatrix}\dfrac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}$

$x^TAx=x^TQ\Lambda Q^Tx=(Q^Tx)^T\Lambda(Q^Tx)=\begin{bmatrix}\frac{1}{\sqrt{2}}x+\frac{1}{\sqrt{2}}y & \frac{1}{\sqrt{2}}x+\frac{1}{\sqrt{2}}y\end{bmatrix}\begin{bmatrix}9 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}x+\frac{1}{\sqrt{2}}y\\\frac{1}{\sqrt{2}}x-\frac{1}{\sqrt{2}}y\end{bmatrix}=9(\dfrac{x+y}{\sqrt{2}})^2+(\dfrac{x-y}{\sqrt{2}})^2$

令$X=\frac{x+y}{\sqrt{2}},Y=\frac{x-y}{\sqrt{2}}$

会得到$9X^2+Y^2=1\Rightarrow\frac{X^2}{\frac{1}{9}}+\frac{Y^2}{1}=1$

因此，椭圆是以$\dfrac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$和$\dfrac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$为主轴的椭圆。

因此，假设$A=Q\Lambda Q^T$是正定矩阵，那么$\lambda_i>0$，则$x^TAx=1$是一个主轴是特征向量方向，长短轴长度为$\frac{1}{\sqrt{\lambda_1}},\frac{1}{\sqrt{\lambda_2}}$的椭圆。

而$x^TQ\Lambda Q^Tx=[X \quad Y]\Lambda\begin{bmatrix}X\\Y\end{bmatrix}$

**——瑞利商**

对于一个$n\times n$的实对称矩阵，它的特征值为$\lambda_1\le\lambda_2\le\lambda_3\le\cdots\le\lambda_n$，对应的特征向量也是单位正交的。

对于任意非零向量$x$，考虑$R(x)\triangleq\frac{x^TAx}{x^Tx}$

令$x=c_1x_1+c_2x_2+\cdots+c_nx_n$

那么，$x^TAx=(c_1x_1+c_2x_2+\cdots+c_nx_n)^TA(c_1x_1+c_2x_2+\cdots+c_nx_n)=c_1^2\lambda_1+c_2^2\lambda_2+\cdots+c_n^2\lambda_n$

且，$x^Tx=c_1^2+c_2^2+\cdots+c_n^2$

因此，瑞利商可以表示为：

$$R(x)=\frac{c_1^2\lambda_1+c_2^2\lambda_2+\cdots+c_n^2\lambda_n}{c_1^2+c_2^2+\cdots+c_n^2}$$

有$\lambda_1(c_1^2+c_2^2+\cdots+c_n^2)\le c_1^2\lambda_1+c_2^2\lambda_2+\cdots+c_n^2\lambda_n\le\lambda_n(c_1^2+c_2^2+\cdots+c_n^2)$

故$\lambda_1\le R(x)\le\lambda_n$，同时，等号是能成立的。、

因此，瑞利商最大值是最大的特征值，最小值是最小的特征值。

## 5、相似矩阵

对于存在n个线性无关的特征向量的矩阵而言，可以进行相似对角化：$A=S\Lambda S^{-1}$

**定义**：令$M$是可逆矩阵，且$B=M^{-1}AM$，则称$B$与$A$相似。

称$A$到$B$的这个过程为**相似变换**。

**反身性**：$A$和它本身相似，因为$A=I^{-1}AI$

**对称性**：若$B=M^{-1}AM$，即$B$相似于$A$，那么，一定有$A=MBM^{-1}$，即$A$相似于$B$。

**传递性**：如果$B$相似于$A$（$B=M_1^{-1}AM_1$），且$C$相似于$B$（$C=M_2^{-1}BM_2$），则有$C$相似于$A$。（因为$C=M_2^{-1}M_1^{-1}AM_1M_2=(M_1M_2)^{-1}AM_1M_2$）

满足这三个性质的二元关系被称为**等价关系**，离散数学里定义过这个概念。如果有一个**等价关系**，那么，一个集合就可以根据这个等价关系分成好几类。

**性质**：如果两个矩阵相似，那么这两个矩阵拥有一样的特征值。

**证明**：有$B=M^{-1}AM$，且有$A=MBM^{-1}$

由于$Ax=\lambda x\Rightarrow MBM^{-1}x=\lambda x$

因为$M$是可逆的，所以有$B(M^{-1}x)=\lambda(M^{-1}x)$

故$B$的特征值是$\lambda$，对应的特征向量是$M^{-1}x.\square$

**另一种证明**：

$$\begin{aligned}det(B-\lambda I)&=det(M^{-1}AM-\lambda I)=det(M^{-1}(A-\lambda I)M)\\&=det(M^{-1})det(A-\lambda I)det(M)=det(A-\lambda I)\end{aligned}.\square$$

对于可以对角化的矩阵，如果他们的特征值一样，则相似。

如果两个矩阵相似，那它们哪些东西会一样呢？

1、特征值

2、迹和行列式[迹为特征值之和，行列式为特征值之积]

3、秩

**说明**：考虑$Ax=0$，有$MBM^{-1}x=0$，因为$M$可逆，两边左乘一个$M^{-1}$会得到$B(M^{-1}x)=0$，如果$x\in N(A)$，那么$Mx\in N(B)$。又$c_1x_1+c_2x_2+\cdots+c_nx_n=0\iff c_1M^{-1}x_1+c_2M^{-1}x_2+\cdots+c_nM^{-1}x_n=0$故，$dim(N(A))=dim(N(B))=n-r$，也就是说两个矩阵的秩一样。

4、线性无关的特征向量个数

5、约当标准型一致（充要条件）

**——约当形**

老师说，这是不满$n$个特征向量的矩阵最接近对角线矩阵的一种形态。

如果矩阵$A$有$s$个线性无关的特征向量，则它相似于矩阵$J$。$J$在对角线上有$s$的约当块($Jordan\ Block$)。

$$M^{-1}AM=\begin{bmatrix}J_1&&&\\&J_2&&\\&&\ddots&\\&&&J_n\end{bmatrix}=J$$

约当块$J_i$的对角线上是特征值$\lambda_i$，并且，在对角线上的一个元素全为$1$，即$J_i=\begin{bmatrix}\lambda_i&1&&&\\&\lambda_i&1&&\\&&\ddots&\ddots&\\&&&&1\\&&&&\lambda_i\end{bmatrix}$

**栗子**：矩阵$A$有特征值$10,8,8,0,0,0$，每个特征值的代数重数和几何重数的数目分别为：10：{1，1}，8：{2，1}，0：{3，1}.其约当型为：

$$J=\begin{bmatrix}10&&&&&\\&8&1&&&\\&0&8&&&\\&&&0&1&\\&&&0&0&\\&&&&&0\end{bmatrix}$$

**性质**：如果$A$可以对角线化，那么$J=\Lambda$，即对角线元素为特征值的对角矩阵。（可以对角线化表示有$n$个线性无关的特征向量，所以有$n$个$1\times1$的约当块，其实就是一个对角线矩阵）

上面的五个条件中，只有约当形一致是充要条件，其他的全是充分条件，也就是说，如果两个矩阵的特征值一样，比如说都是$\{2,0\}$，不能说明两个矩阵相似；反例可以从多个角度给出，1）、两个矩阵形状不一样，$2\times2$和$3\times3$的矩阵拥有一样的特征值，但两个矩阵一定不相似；2）、假如同样都是$3\times3$的矩阵，那其中一个矩阵可能可以相似对角化，但另一个不可以，那么，这俩矩阵就不相似。

**性质**：$A$和$B$相似 $\Longleftrightarrow$ $A$和$B$的约当形一致。

（约当形部分未完全证明，只是粗略介绍。一个矩阵有唯一的约当形，但出现$AM=4, GM=2$的情况就不能通过上面介绍的方法得出约当形的形式，因为$4=3+1=2+2$，这没法直接看出来，课后补充）

## 6、奇异值分解（Singular Value Decomposition）

**——介绍**

**奇异值分解**可以将$m\times n$的矩阵进行分解：

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

其中，$U$是$m\times m$的矩阵，$\Sigma$是$m\times n$的矩阵，$V$是$n\times n$矩阵；且

$$U^T=U^{-1}, V^T=V^{-1}, \Sigma=\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & 0 \\ & & \ddots & \\ & 0 & & \sigma_r \\ & & & 0\end{bmatrix}$$

经过变形，可以获得$AV=U\Sigma$

$$A\begin{bmatrix}v_1 & v_2 & \cdots & v_n\end{bmatrix}=\begin{bmatrix}u_1 & v_u & \cdots & u_m\end{bmatrix}\begin{bmatrix}\sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_r & \\ & 0 & & & 0 \\ & & & & \ddots\end{bmatrix}$$

其中，$Av_i=\sigma_i u_i$, 对于$i=\{1,2,3,\cdots,r\}$

对于$V$而言，前$r$个向量是行空间的基（还是$A^TA$的特征向量），后$n-r$个向量是零空间的基。

对于$U$而言，前$r$个向量是列空间的基（还是$AA^T$的特征向量），后$m-r$个向量是左零空间的基。

**——推导**

给定$m\times n$的矩阵$A$，$rank(A)=r$，$A^TA$是$n\times n$矩阵，且对称。故$A^TA$一定可以对角线化，且有$n$个线性无关的特征向量，即$A^TAv_i=\lambda_i v_i$，$i=\{1,2,3,\cdots,n\}$

$||Av_i||^2=v_i^TA^TAv_i=\lambda_i||v_i||^2=\lambda_i\geq 0$

由于$rank(A^TA)=rank(A)=r$

故$A^TA$零特征值的个数为$n-r$个，即零主元的个数；而非零特征值一共有$r$个。故，可以假设$\lambda_1,\lambda_2,\cdots,\lambda_r$是非零特征值，而$\lambda_{r+1},\cdots,\lambda_n$是零特征值。

令$\sigma_i=\sqrt{\lambda_i}>0, i=\{1,2,\cdots,r\}$，这$r$个值被称为**奇异值。**

因此，可以得到$u_i=\frac{Av_i}{\sigma_i}, ||u_i||^2=\frac{1}{\sigma_i^2}v_iA^TAv_i=\frac{\lambda_i}{\lambda_i}=1$

另外，$u_i^Tu_j=\frac{1}{\sigma_i\sigma_j}v_iA^TAv_j=\frac{1}{\sigma_i\sigma_j}\lambda_j v_i^Tv_j=0,$ 当$i\neq j$

因此，$u_1,u_2,\cdots,u_r$也是单位正交的；通过$Gram-Schmidt$正交化，可以找到$R^m$意义下的另外$m-r$个单位正交向量，使得$u_1,u_2,\cdots,u_r,r_{r+1},\cdots,u_m$是$\mathbb{R}^m$的一组单位正交基。

所以有：$Av_i=\sigma_i u_i, i=\{1,2,3,\cdots,r\}$

写成矩阵就是$AV_1=U_1\Sigma$

$$A\begin{bmatrix}v_1 & v_2 & \cdots & v_r\end{bmatrix}=[u_1 \quad u_2 \quad \cdots \quad u_r]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$$

然后，再把$n-r$个零空间的解（0特征值对应的特征向量）加到矩阵$V_1$后面，变成$V$；还有$m-r$个线性无关的特征向量加到$U_1$后面变成$U$，相应地，为了维持矩阵的形状，$\Sigma$就变成了$m\times n$矩阵，后面$n-r$列全添0，因为$Av_i=0v_i, i=\{r+1,\cdots,n\}$

这样就得到$AV=U\Sigma\Rightarrow A=U\Sigma V^T$

**注记1**：对于半正定矩阵（$n\times n$的对称矩阵），奇异值分解就是矩阵对角化，$A=Q\Lambda Q^T$

**注记2**：$AA^T=(U\Sigma V^T)(V\Sigma^T U^T)=U\Sigma\Sigma^T U^T=U\Lambda U^T$

其中，$\Sigma\Sigma^T=\begin{bmatrix}\sigma_1^2 & & & & & \\ & \sigma_2^2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r^2 & & \\ & & & & 0 & \\ & & & & & \ddots\end{bmatrix}=\Lambda$

故，$u_1,u_2,\cdots,u_m$是$AA^T$的（单位正交）的特征向量矩阵，对应的特征值是$\lambda_1,\lambda_2,\cdots,\lambda_r,0,\cdots,0(r+m-r=m)$

前面说过，$v_1,v_2,\cdots,v_n$是$A^TA$的（单位正交）特征向量矩阵，对应的特征值是$\lambda_1,\lambda_2,\cdots,\lambda_r,0,\cdots,0(r+n-r=n)$

**注记3**：$A^TAv_i=\lambda_i v_i$，$v_i=A^T\frac{Av_i}{\lambda_i}$，$i=\{1,2,3,\cdots,r\}$，故$v_i$在$A^T$的列空间中，$A$的行空间中 。由于$dim(C(A^T))=dim(C(A))=r$，又$v_1,v_2,\cdots,v_r$是$r$个线性无关的单位正交的特征向量，因此，$v_1\rightarrow v_r$是$C(A^T)$的一组基。由于，$V$中后$n-r$个向量是0特征值的特征向量，也就是$N(A)$，故，$v_{r+1},v_{r+2},\cdots,v_n$是$N(A)$的正交基。

另外，还有$Av_i=\sigma_i u_i,i=\{1,2,3,\cdots,r\}$，同样的，$u_i$再$A$的列空间中，故$u_1,u_2,\cdots,u_r$是$C(A)$的正交基。又$u_{r+1},\cdots,u_m$是$u_1\rightarrow u_r$的正交补空间中线性无关的$m-r$个向量，因此，$u_{r+1},\cdots,u_m$是$N(A^T)$的单位正交基。

**——应用**

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

$$A=\begin{bmatrix}u_1 & u_2 & \cdots & u_m\end{bmatrix}\begin{bmatrix}\sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0\end{bmatrix}\begin{bmatrix}v_1^T \\ v_2^T \\ \vdots \\ v_n^T\end{bmatrix}$$
$$=\sigma_1u_1v_1^T+\sigma_2u_2v_2^T+\cdots+\sigma_ru_rv_r^T\{\sigma_1>\sigma_2>\cdots>\sigma_r\}$$

可以用奇异值分解来存图，这样基本上就是要存$r(m+n+1)$个值。对于一张$1920*960$的图来说最多可能要有$960(1920+960+1)$这么多。但其中，可能会有很多$\sigma$的值很小，比如$1\times10^{-6}$这种，对与只有$255$个状态的图来说，其实贡献一点也没有的，所以可以将这种很小的$sigma$对应的值给丢掉，这样就可以进行有损压缩(lossy)。