北京航空航天大學
BEIHANG UNIVERSITY

矩阵理论

王磊

自动化科学与电气工程学院

函 数 矩 阵

定义4.8.1(函数矩阵) 以变量t的函数为元素的矩阵$A(t)=(a_{ij}(t))_{m\times n}$称为函数(值)矩阵;

若矩阵$A(t)$的每个元素$a_{ij}(t)$在$[a,b]$上是连续、可微或可积时,则称$A(t)$在$[a,b]$上连续、可微或可积,并定义

$$A'(t)=\frac{d}{dt}A(t)=(a'_{ij}(t))_{m\times n}$$

$$\int_a^b A(t)\,dt=(\int_a^b a_{ij}(t)dt)_{m\times n}$$

例4.8.1 求函数矩阵$A(t)=\begin{bmatrix} \sin t & \cos t & t \\ e^t & e^{2t} & e^{3t} \\ 0 & 1 & t^2 \end{bmatrix}$的导数.

命题4.8.1 设$A(t)$和$B(t)$是适当阶的可微矩阵，则

（1）$\frac{d}{dt}\big(A(t)+B(t)\big)=\frac{d}{dt}A(t)+\frac{d}{dt}B(t)$；

（2）$\lambda(t)$为可微标量函数时，

$$\frac{d}{dt}\big(\lambda(t)A(t)\big)=\frac{d\lambda(t)}{dt}A(t)+\lambda(t)\frac{d}{dt}A(t);$$

（3）$\frac{d}{dt}\big(A(t)B(t)\big)=\frac{dA(t)}{dt}B(t)+\frac{dB(t)}{dt}A(t)$；

（4）$u=f(t)$可微时，$\frac{d}{dt}\big(A(u)\big)=f'(t)\frac{d}{dt}A(t)$；

命题4.8.1（续）

（5）当$A(t)$是可逆矩阵时，有

$$\frac{d}{dt}\left(A^{-1}(t)\right)=-A^{-1}(t)\left(\frac{d}{dt}A(t)\right)A^{-1}(t).$$

命题4.8.2 设$A \in \mathbb{C}^{n\times n}$,则有

（1） $\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A$;

（2） $\frac{d}{dt}\sin(At) = A\cos(At) = (\cos At)A$;

（3） $\frac{d}{dt}\cos(At) = -A\sin(At) = -(\sin At)A$.

推论4.8.1设$A \in \mathbb{C}^{n\times n}$，则有

（1） $\int_{t_0}^{t} Ae^{As}\,ds = e^{At} - e^{At_0}$；

（2） $\int_{t_0}^{t} A\sin(As)\,ds = \cos(At_0) - \cos(At)$；

（3） $\int_{t_0}^{t} A\cos(As)\,ds = \sin(At) - \sin(At_0)$.

命题4.8.3 设$A(t)$和$B(t)$是$[a,b]$上适当阶的可积矩阵，$\lambda\in\mathbb{C}$，则有

（1） $\int_a^b \big(A(t)+B(t)\big)\,dt=\int_a^b A(t)\,dt+\int_a^b B(t)\,dt$

（2） $\int_a^b \lambda A(t)\,dt=\lambda\int_a^b A(t)\,dt$；

（3） $A(t)$在$[a,b]$上连续时，则$\forall t\in(a,b)$

$$\frac{d}{dt}\left(\int_a^t A(\tau)\,d\tau\right)=A(t);$$

（4） $A(t)$在$[a,b]$上连续时，有

$$\int_a^b A'(\tau)\,d\tau=A(b)-A(a).$$

定义4.8.2(矩阵对矩阵的导数) 设$F(X)=(f_{ij}(X))\in\mathbb{C}^{m\times n}$是$X\in\mathbb{C}^{p\times q}$的函数矩阵,$f_{ij}(X)$, $i=1,\cdots,m,j=1,\cdots,n$作为$X\in\mathbb{C}^{p\times q}$的多元函数是可微的,令

$$\frac{dF(X)}{dX}=\left(\frac{\partial F}{\partial x_{ij}}\right)_{mp\times nq}=\begin{bmatrix}\frac{\partial F}{\partial x_{11}} & \frac{\partial F}{\partial x_{12}} & \cdots & \frac{\partial F}{\partial x_{1q}}\\ \frac{\partial F}{\partial x_{21}} & \frac{\partial F}{\partial x_{22}} & \cdots & \frac{\partial F}{\partial x_{2q}}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{\partial F}{\partial x_{p1}} & \frac{\partial F}{\partial x_{p2}} & \cdots & \frac{\partial F}{\partial x_{pq}}\end{bmatrix}$$

式中,$\frac{\partial F}{\partial x_{ij}}=(\frac{\partial f_{kl}(x)}{\partial x_{ij}}),k=1,\cdots,m,l=1,\cdots,n,$

称$\frac{dF(X)}{dX}$为$F(X)$对X的导数.

注1: 定义4.8.2既是对定义4.8.1中矩阵函数导数的扩展, 也隐含着向量对向量、向量对矩阵、矩阵对向量和矩阵对矩阵的导数定义.

例4.8.2 设$f(\boldsymbol{x})=f(x_1,\cdots,x_n)$是$n$元实可微函数，则

$$\frac{df}{d\boldsymbol{x}}=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right]$$

式中，$\boldsymbol{x}=[x_1,\cdots,x_n]$.

我们常称$\frac{df}{dx}$为f的梯度向量，常记为$\nabla(f)$.

例4.8.3 设$\boldsymbol{f}(\boldsymbol{x}) = [f_1, f_2, \cdots, f_m]^T \in F^m$是$n$元实可微向量函数,即每个$n$元函数$f_i(\boldsymbol{x}) = f_i(x_1, \cdots, x_n)$, $1 \le i \le m$均为可微函数,其中$\boldsymbol{x} = [x_1, \cdots, x_n]$为行向量.

$$\frac{d\boldsymbol{f}}{d\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

矩阵$\frac{d\boldsymbol{f}}{d\boldsymbol{x}}$常称为向量函数$\boldsymbol{f}$的Jacobian矩阵.

例4.8.4 设常矩阵$A \in \mathbb{R}^{n\times n}$和未定元向量$\boldsymbol{x} \in \mathbb{R}^{n}$，求$\frac{dA\boldsymbol{x}}{d\boldsymbol{x}^T}$.

$$\frac{d\boldsymbol{y}}{d\boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = A$$

例4.8.5 设常矩阵$A\in\mathbb{R}^{n\times n}$, $\boldsymbol{x},\boldsymbol{b}\in\mathbb{R}^n$, 求$\frac{d}{d\boldsymbol{x}}(\boldsymbol{b}^T\boldsymbol{x})$和$\frac{d}{d\boldsymbol{x}}(\boldsymbol{x}^TA\boldsymbol{x})$.

$$\frac{d}{d\boldsymbol{x}}(\boldsymbol{b}^T\boldsymbol{x})=\begin{bmatrix}\frac{\partial(\boldsymbol{b}^T\boldsymbol{x})}{\partial x_1}\\ \frac{\partial(\boldsymbol{b}^T\boldsymbol{x})}{\partial x_2}\\ \vdots\\ \frac{\partial(\boldsymbol{b}^T\boldsymbol{x})}{\partial x_n}\end{bmatrix}=\begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix}=\boldsymbol{b}$$

$$\frac{d}{d\boldsymbol{x}}(\boldsymbol{x}^TA\boldsymbol{x})=\begin{bmatrix}\sum_{i=1}^{n}a_{i1}x_i+\sum_{j=1}^{n}a_{1j}x_j\\ \sum_{i=1}^{n}a_{i2}x_i+\sum_{j=1}^{n}a_{2j}x_j\\ \vdots\\ \sum_{i=1}^{n}a_{in}x_i+\sum_{j=1}^{n}a_{nj}x_j\end{bmatrix}=(A^T+A)\boldsymbol{x}$$

设$\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$，则有

$$\frac{d\boldsymbol{u}^T\boldsymbol{v}}{d\boldsymbol{x}} = \frac{d\boldsymbol{u}^T}{d\boldsymbol{x}}\boldsymbol{v} + \frac{d\boldsymbol{v}^T}{d\boldsymbol{x}}\boldsymbol{u}$$

$$\frac{d\boldsymbol{u}\boldsymbol{v}^T}{d\boldsymbol{x}} = \frac{d\boldsymbol{u}}{d\boldsymbol{x}}\boldsymbol{v}^T + (\boldsymbol{u}\otimes I_n)\ \frac{d\boldsymbol{v}^T}{d\boldsymbol{x}}$$

网上各种公式一定要验证！

【应用】：求解最小二乘问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2$$

式中，$A\in\mathbb{R}^{n\times n}$和$\boldsymbol{b}\in\mathbb{R}^n$给定，$\boldsymbol{x}\in\mathbb{R}^n$待定向量.

引理4.8.1(无约束极值的必要条件) 设$X \in \mathbb{R}^{m\times n}$, $f(X)$是关于矩阵变量X的连续可导函数. 若X_0是$f(X)$的一个极值点,则

$$\left.\frac{df(X)}{dX}\right|_{X=X_0} = 0$$

若$f(X)$还是关于矩阵变量X的凸函数,则$f(X)$的极值点必是最值点.

补充: 考查n元函数$f(\boldsymbol{x})$在$\boldsymbol{x}_0=[x_{10},x_{20},\cdots,x_{n0}]^T$的泰勒展开.

$$f(\boldsymbol{x})=f(\boldsymbol{x}_0)+\frac{\partial f(\boldsymbol{x}_0)}{\partial x_1}(x_1-x_{10})+\cdots+\frac{\partial f(\boldsymbol{x}_0)}{\partial x_n}(x_n-x_{n0})$$

$$+\frac{1}{2!}\frac{\partial^2 f(\boldsymbol{x}_0)}{\partial {x_1}^2}(x_1-x_{10})^2$$

$$+\frac{1}{2!}\frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1\partial x_2}(x_1-x_{10})(x_2-x_{20})$$

$$+\cdots+\frac{1}{2!}\frac{\partial^2 f(\boldsymbol{x}_0)}{\partial {x_n}^2}(x_n-x_{n0})^2+\cdots$$

补充: 考查n元函数$f(\boldsymbol{x})$在$\boldsymbol{x}_0=[x_{10},x_{20},\cdots,x_{n0}]^T$的泰勒展开.

$$f(\boldsymbol{x})=f(\boldsymbol{x}_0)+\left(\frac{df(\boldsymbol{x}_0)}{d\boldsymbol{x}}\right)^T(\boldsymbol{x}-\boldsymbol{x}_0)$$

$$+\frac{1}{2!}(\boldsymbol{x}-\boldsymbol{x}_0)^T H(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)+\cdots$$

补充: 考查n元函数$f(\boldsymbol{x})$在$\boldsymbol{x}_0=[x_{10},x_{20},\cdots,x_{n0}]^T$的泰勒展开.

$$H(\boldsymbol{x}_0)=\begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial {x_1}^2} & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial {x_n}^2} \end{bmatrix}$$

$$=\frac{d}{d\boldsymbol{x}^T}\left(\frac{df(\boldsymbol{x}_0)}{d\boldsymbol{x}}\right)$$

上述矩阵称为Hessian矩阵.

【应用】: 考查一阶线性常系数微分方程组的求解问题.

$$\begin{cases} \dfrac{dx_1(t)}{dt} = a_{11}x_1(t) + \cdots + a_{1n}x_n(t) + f_1(t) \\ \vdots \\ \dfrac{dx_n(t)}{dt} = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) + f_n(t) \end{cases}$$

例4.8.6 求下列方程组的解

$$\begin{cases} \dfrac{d\boldsymbol{x}(t)}{dt} = A\boldsymbol{x}(t) + \boldsymbol{f}(t) \\ \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \end{cases}$$

其中参数分别为

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

$$\boldsymbol{f}(t) = [1, -t, t]^T, \boldsymbol{x}_0 = [1,0,-1]^T.$$

再考查n阶常系数微分方程的求解问题:

$$y^{(n)}+a_{n-1}y^{(n-1)}+\cdots+a_0y=u(t)$$

式中, $a_0,\cdots,a_{n-1}$为常数, $u(t)$为已知函数.

当$u(t)\neq 0$时, 称上述方程为非齐次微分方程, 否则称为齐次微分方程.

定义

$$x_1 = y$$
$$x_2 = \dot{y}$$
$$\vdots$$
$$x_n = y^{(n-1)}$$

由此,得到如下方程

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x})$$

上式在现代控制理论中称为状态方程,$\boldsymbol{x}(t)$称为状态向量,A为系统矩阵. 此时,将n阶常系数微分方程转化为状态方程进行求解.

例4.8.7 求解非齐次线性微分方程组

$$y^{(3)} - 6\ddot{y} + 11\dot{y} - 6y = \sin t$$

式中，$\ddot{y}(0) = \dot{y}(0) = y(0) = 0.$

【应用】:系统的可控性是指对任意给定的系统初始状态,均存在一个合适控制输入使得系统状态在有限时间转移至状态零点. 它表征了系统输入对系统状态的有效控制能力,是现代控制理论的基本概念.

例4.8.8即为线性定常系统可控的充分必要条件.

例4.8.8 n阶Hermite矩阵$W(0,t_1)$是正定的当且仅当矩阵$\mathrm{rank}(C)=n$,其中

$$W(0,t_1)=\int_0^{t_1} e^{-At}BB^Te^{-A^Tt}\,dt$$

$$C=[B,AB,\cdots,A^{n-1}B]$$

$A\in\mathbb{R}^{n\times n}$是线性定常系统的系统矩阵,$B\in\mathbb{R}^{n\times p}$是线性定常系统的输入矩阵,$t_1$是给定的正常数.

应用：主元分析法（自学）

主元分析法是一种常用的数据降维方法，其目的是在“信息损失”较小的前提下，将高维数据转换到低维. 基于此，我们给出这一问题的数学描述.

假设一组数据有n个样本，即$\boldsymbol{x}_1, \cdots \boldsymbol{x}_n$，其中，$\boldsymbol{x}_i \in \mathbb{R}^p$表示每个样本向量(或随机向量).

设降维后的样本向量为$y_i \in \mathbb{R}^q (q < p$，否则失去降维意义)，则$\mathrm{span}(\boldsymbol{y}_1, \cdots \boldsymbol{y}_n)$应与$\mathrm{span}(\boldsymbol{x}_1, \cdots \boldsymbol{x}_n)$同构.

由此，存在同构映射f使得原像$\boldsymbol{x}_i$与$\boldsymbol{y}_i$满足

$$\boldsymbol{y}_i = f(\boldsymbol{x}_i)$$

由于f是线性映射，故可等价表示为

$$\boldsymbol{y}_i = W\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{w}_1^T \boldsymbol{x}_i \\ \vdots \\ \boldsymbol{w}_p^T \boldsymbol{x}_i \end{bmatrix}$$

式中，矩阵$W \in \mathbb{R}^{q \times p}$，$\boldsymbol{w}_i \in \mathbb{R}^p$为待定向量.

考查表达式$\boldsymbol{w}_k^T\boldsymbol{x_i}$. 在欧氏空间中,

$$\boldsymbol{w}_k^T\boldsymbol{x_i}=(\boldsymbol{w}_k,\boldsymbol{x_i})$$

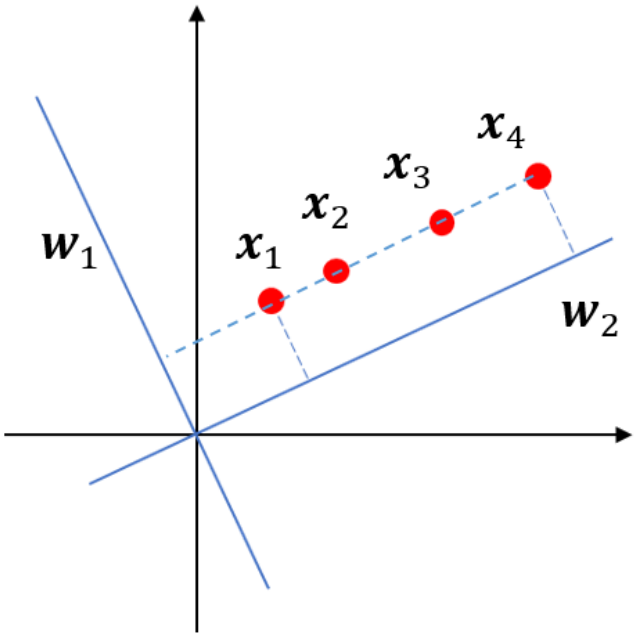
若$\|\boldsymbol{w}_k\|=1$, 则上式表示向量$\boldsymbol{x_i}$在向量$\boldsymbol{w}_k$的正交投影的长度, 即映射的$\boldsymbol{y}_i$的每一个分量是原像$\boldsymbol{x_i}$在矩阵W行向量的正交投影.

为保证$\boldsymbol{y}_i$尽量继承原始变量特征, 应选择$\boldsymbol{w}_i(i=1,\cdots,q)$是正交向量组.

所以矩阵W是行正交规范矩阵, 即向量组$\boldsymbol{w}_i$ $(i=1,\cdots,q)$是单位正交向量组.

为刻画“信息损失”较小，我们考查二维平面的一个例子.

确定待定向量（组）$\boldsymbol{w}$



在数学上,我们可以用的方差来表述样本分散的程度. 于是,优化目标可表达为

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{w}^T\boldsymbol{x_i} - \mu)^2$$

式中,μ是样本投用的均值,其定义式为

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{w}^T\boldsymbol{x_i}$$

注意到

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{w}^T\boldsymbol{x}_{\boldsymbol{i}} - \mu)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{w}^T(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{w}^T(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^T\boldsymbol{w} \\
&= \boldsymbol{w}^T(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^T)\boldsymbol{w}
\end{aligned}
$$

并定义矩阵

$$
\boldsymbol{x} = [\boldsymbol{x}_1 - \overline{\boldsymbol{x}}, \boldsymbol{x}_2 - \overline{\boldsymbol{x}}, \cdots, \boldsymbol{x}_n - \overline{\boldsymbol{x}}] \in \mathbb{R}^{p\times n}
$$

则方差σ^2可进一步改下为

$$\sigma^2 = \boldsymbol{w}^T \left(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T\right)\boldsymbol{w} = \boldsymbol{w}^T C \boldsymbol{w}$$

式中, $C = \frac{1}{n}\boldsymbol{x}\boldsymbol{x}^T \in \mathbb{R}^{p\times p}$是随机变量$\boldsymbol{x}_1, \cdots \boldsymbol{x}_n$的协方差矩阵.

由引理3.9.2知, C半正定实对称矩阵. 故

$$\sigma^2 = R(\boldsymbol{w})$$

式中, $R(\boldsymbol{w})$是协方差矩阵C的Rayleigh商.

方法一：为讨论方便，设$\lambda_1 > \lambda_2 > \cdots \lambda_m$是矩阵$C$的$m$个互异特征值，其代数重数分别为$d_1, \cdots, d_l$且有$d_1 + \cdots + d_l = p$. 根据定理4.3.4知，

$$R(\boldsymbol{w}) \leq \lambda_1$$

上式等号成立的条件为向量$\boldsymbol{w}$取为属于特征值λ_1的d_1个单位正交特征向量组中的一个向量. 由此，我们找到了d_1个$\boldsymbol{w}$向量. 那么余下的$q - d_1$个$\boldsymbol{w}$向量应如何确定呢？

我们应在特征值为$\lambda_2,\cdots,\lambda_m$的特征子空间的和空间来寻找（为什么？），即

$$\hat{\boldsymbol{w}} = \arg \max_{\boldsymbol{w} \in \left(\bigoplus_{i=2,\cdots,m} E(\lambda_i) \right)} R(\boldsymbol{w})$$

根据定理4.3.4知,

$$\max_{\boldsymbol{w} \in \left(\bigoplus_{i=2,\cdots,m} E(\lambda_i) \right)} R(\boldsymbol{w}) \le \lambda_2$$

上式等号成立的条件为向量$\boldsymbol{w}$取为属于特征值λ_2的d_2个单位正交特征向量组中的一个向量. 由此，我们找到了d_2个$\boldsymbol{w}$向量. 依次类推, 我们可以找到q个满足要求的向量$\boldsymbol{w}$. 这q个向量$\boldsymbol{w}$恰好是协方差矩阵C的属于前q最大个特征值（计重数）的单位正交特征向量即可得到.

方法二: 由Lagrange乘子法构造目标函数:

$$L(\boldsymbol{w},\lambda)=\boldsymbol{w}^T C\boldsymbol{w}+\lambda(1-\boldsymbol{w}^T\boldsymbol{w})$$

对$L(\boldsymbol{w},\mu)$求偏导数, 得

$$\frac{\partial L(\boldsymbol{w},\mu)}{\partial \boldsymbol{w}}=2C\boldsymbol{w}-2\lambda\boldsymbol{w}$$

令上式为零, 得以下极值条件

$$C\boldsymbol{w}=\lambda\boldsymbol{w}$$

上式表明,λ是协方差矩阵C的特征值, $\boldsymbol{w}$是属于特征值λ的单位特征向量.

将$C\boldsymbol{w}=\lambda\boldsymbol{w}$目标函数,得

$$\sigma^2=\lambda$$

为保证方差最大,只要我们对矩阵C的特征值进行排序,然后选取前属于前q最大个特征值(计重数)的单位正交特征向量即可得到W,上式表明投影后的方程就是协方差矩阵的特征值.