矩阵理论

王磊

自动化科学与电气工程学院

第四章 矩阵分析

- □ 向量范数
- □ 矩阵范数
- □ 相容范数
- □ 矩阵扰动分析
- □ 特征值估计
- □ 矩阵级数
- □ 矩阵函数
- □ 函数矩阵
- □ 应用：主元分析法

矩 阵 函 数

定义4.7.1（矩阵函数）设幂级数$\sum_{m=0}^{\infty} c_m z^m$的收敛半径$r$, $z \in \mathbb{C}$. 当$|z| < r$时, 幂级数收敛于函数$f(z)$, 即

$$f(z) = \sum_{m=0}^{\infty} c_m z^m, |z| < r$$

若复方阵A满足$\rho(A) < r$, 称收敛的矩阵幂级数$\sum_{m=0}^{\infty} c_m A^m$为矩阵函数, 记为$f(A)$.

常见矩阵函数有

$$e^{A}=\sum_{m=0}^{\infty}\frac{1}{m!}A^{m},\forall A\in\mathbb{C}^{n\times n}$$ (矩阵指数函数)

$$\sin A=\sum_{m=0}^{\infty}\frac{(-1)^{m}}{(2m+1)!}A^{2m+1},\forall A\in\mathbb{C}^{n\times n}$$ (矩阵正弦函数)

$$\cos A=\sum_{m=0}^{\infty}\frac{(-1)^{m}}{(2m)!}A^{2m},\forall A\in\mathbb{C}^{n\times n}$$ (矩阵余弦函数)

$$(I-A)^{-1}=\sum_{m=0}^{\infty}A^{m},\forall\rho(A)<1$$

$$\ln(I+A)=\sum_{m=0}^{\infty}\frac{(-1)^{m}}{m+1}A^{m+1},\forall\rho(A)<1$$

命题4.7.1 设$A \in \mathbb{C}^{n\times n}$, 则以下结论成立:

（1） $\cos(-A) = \cos A, \sin(-A) = -\sin A$;

（2） $e^{\mathrm{i}A} = \cos A + \sin A$;

（3） $\cos A = \frac{1}{2}(e^{\mathrm{i}A} + e^{-\mathrm{i}A})$;

（4） $\sin A = \frac{1}{2}(e^{\mathrm{i}A} - e^{-\mathrm{i}A})$.

注1: 尽管指数函数满足$e^a e^b = e^{a+b} = e^b e^a$, 但该性质对矩阵指数函数一般不成立.

例4.7.1 已知$A=\begin{bmatrix}1 & -1\\ 0 & 0\end{bmatrix}$和$B=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$, 分别计算$e^A$, e^B, e^{A+B}, e^Ae^B, e^Be^A, 并比较e^{A+B}, e^Ae^B和e^Be^A.

例4.7.1 已知$A=\begin{bmatrix}1 & -1\\ 0 & 0\end{bmatrix}$和$B=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$, 分别计算$e^A$, e^B, e^{A+B}, e^Ae^B, e^Be^A, 并比较e^{A+B}, e^Ae^B和e^Be^A.

注2: 矩阵函数的定义式提供了一种计算矩阵函数的方法. 在采用定义法计算矩阵函数时应巧妙地应用矩阵的最小多项式进行简化计算.

定理4.7.1 设$A, B \in \mathbb{C}^{n\times n}$, 若$AB = BA$, 则$e^A e^B = e^B e^A = e^{A+B}$.

推论4.7.1 设$A \in \mathbb{C}^{n\times n}$, 则$e^{A}e^{-A}=e^{-A}e^{A}=I$, 即$(e^{A})^{-1}=e^{-A}$.

注3: 推论4.7.1表明无论矩阵A是否可逆, 矩阵指数函数e^{A}必可逆, 且其逆矩阵为e^{-A}.

推论4.7.2 设$A \in \mathbb{C}^{n\times n}$, 则$\sin^2 A+\cos^2 A=I$.

若矩阵函数$f(A)$的自变量由矩阵A换成At, 其中t为标量参数,则有矩阵函数表达式为

$$f(At)=\sum_{m=0}^{\infty}c_m(At)^m=\sum_{m=0}^{\infty}c_mt^mA^m,t\rho(A)<r$$

我们称之为含参矩阵函数. 这类矩阵函数在线性常微分方程求解等应用中常会遇到.

例4.7.2 设$A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$, 求$e^{At}$.

例4.7.3 设$A \in \mathbb{C}^{4\times 4}$,其特征值分别为$\pi, -\pi, 0, 0$, 求$\sin A$和$\cos A$.

定理4.7.2 设复方阵A与B相似, 即存在可逆矩阵P使得$P^{-1}AP=B$, 若$f(A)$是矩阵函数, 则$f(A)=Pf(B)P^{-1}$.

基于定理4.7.2, 我们对矩阵A分两种情况讨论:

（1）A为单纯矩阵;（2）A不是单纯矩阵.

（1）若A为单纯矩阵, 则存在可逆矩阵P使得 $P^{-1}AP=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, 进而

$$f(A)=P\mathrm{diag}(f(\lambda_1),\cdots,f(\lambda_n))P^{-1}$$

$$f(At)=P\mathrm{diag}(f(\lambda_1 t),\cdots,f(\lambda_n t))P^{-1}$$

此时, 矩阵函数和含参矩阵函数仍是单纯矩阵.

$$e^A=P\mathrm{diag}(e^{\lambda_1},\cdots,e^{\lambda_n})P^{-1}$$

$$\sin A=P\mathrm{diag}(\sin\lambda_1,\cdots,\sin\lambda_n)P^{-1}$$

例4.7.4 设矩阵$A=\begin{bmatrix}1 & 2\\ 0 & 2\end{bmatrix}$, 求矩阵函数$\sin(A)$.

注3: 利用单纯矩阵的谱分解式也可以方便地计算矩阵函数.

（2）若A为非单纯矩阵, 则存在可逆矩阵P使得$P^{-1}AP=J$, 其中J是矩阵A的Jordan标准形, 可表示为$J=\mathrm{diag}(J_1,\cdots,J_s)$, $J_i(i=1,\cdots,s)$为Jordan块, 可表示为

$$J_i=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i\end{bmatrix}_{n_i\times n_i}$$

此时, $A^k = PJ^kP^{-1}$, 式中,

$$J^k = \mathrm{diag}(J_1^k, \cdots, J_s^k)$$

$J_i^k\ (i = 1, \cdots, s)$满足

$$J_i^k = \begin{bmatrix} \lambda_i^k & C_k^1\lambda_i^{k-1} & C_k^2\lambda_i^{k-2} & \cdots & C_k^{n_i-1}\lambda_i^{k-n_i+1} \\ & \lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{n_i-2}\lambda_i^{k-n_i+2} \\ & & \ddots & \ddots & \vdots \\ & & & \lambda_i^k & C_k^1\lambda_i^{k-1} \\ & & & & \lambda_i^k \end{bmatrix}$$

将上式代入矩阵函数$f(A)=\sum_{m=0}^{\infty}c_m A^m$, 最终得

$$f(A)=P\begin{bmatrix} f(J_1) & & 0 \\ & \ddots & \\ 0 & & f(J_s) \end{bmatrix}P^{-1}$$

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f^{n_i-1}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}_{n_i\times n_i}$$

上式称为Sylvester公式.

推论4.7.3 设矩阵$A \in \mathbb{C}^{n\times n}$的特征值为$\lambda_1, \cdots, \lambda_n$, $f(z) = \sum_{m=0}^{\infty} c_m z^m$的收敛半径为$r$, 当$\rho(A) < r$时, 矩阵函数$f(A)$的特征值为$f(\lambda_1), \cdots, f(\lambda_n)$.

例4.7.5 设矩阵$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 2 & -5 & 4\end{bmatrix}$, 求矩阵函数$e^{A}$.

例4.7.5 设矩阵$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 2 & -5 & 4\end{bmatrix}$, 求矩阵函数$e^{A}$.

所以A的Jordan标准形J及相似变换矩阵P为:

$$J=\begin{bmatrix}2 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\quad P=\begin{bmatrix}1 & 1 & 0\\ 2 & 1 & 1\\ 4 & 1 & 2\end{bmatrix}$$

推论4.7.4 设矩阵$A \in \mathbb{C}^{n\times n}$有分解式$A = PJP^{-1}$, 其中$J = \mathrm{diag}(J_1, \cdots, J_s)$, $J_i(i = 1, \cdots, s)$, 幂级数$f(z) = \sum_{m=0}^{\infty} c_m z^m$的收敛半径为$r$, 当$t\rho(A) < r$时, 含参矩阵函数

$$f(At) = P\mathrm{diag}(f(J_1(\lambda_1 t), \cdots, f(J_s(\lambda_s t))P^{-1}$$

$$f(J_i(\lambda_i t)) = \begin{bmatrix} f(\lambda_i t) & tf'(\lambda_i t) & \cdots & \frac{t^{n_i-1}}{(n_i-1)!} f^{n_i-1}(\lambda_i t) \\ & f(\lambda_i t) & \ddots & \vdots \\ & & \ddots & tf'(\lambda_i t) \\ & & & f(\lambda_i t) \end{bmatrix}_{n_i \times n_i}$$

例4.7.6 设矩阵$A=\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}$, 求矩阵函数$\cos(At)$.

例4.7.6 设矩阵$A=\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}$, 求矩阵函数$\cos(At)$.

$$\cos(At)=\begin{bmatrix}\cos 2t & t\cos' 2t\\ 0 & \cos 2t\end{bmatrix}$$

$$=\begin{bmatrix}\cos 2t & -t\sin 2t\\ 0 & \cos 2t\end{bmatrix}.$$

定理4.7.3 设矩阵$A \in \mathbb{C}^{n\times n}$的最小多项式的阶数为$l$, 幂级数$f(z) = \sum_{m=0}^{\infty} c_m z^m$的收敛半径为$r$. 若$\rho(A) < r$, 定义矩阵函数$f(A)$, 则必存在唯一的$(l-1)$次矩阵多项式$p(A) = \beta_0 I + \beta_1 A + \cdots + \beta_{l-1} A^{l-1}$使得$f(A) = p(A)$.

例4.7.8 考查$f(z), p(z)$和矩阵A

$$f(z) = e^z, \qquad p(z) = \beta_0 + \beta_1 z, \qquad A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

定义4.7.2（谱上给定）设$\lambda_1,\cdots,\lambda_s$是$n$阶复方阵$A$的$s$个互异特征值, $m_A(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}$ $(\lambda-\lambda_s)^{n_s}$ 是A的最小多项式, $\deg(m_A(\lambda))=l$. 若复函数$f(z)$及其各阶导数$f^{(i)}(z)$在$z=\lambda_i$处的n_i个值$f^{(i)}(\lambda_i)$均有界, $j=0,1,\cdots,n_i-1$, 则称$f(z)$在矩阵A的谱上给定（或谱上有定义），并称$\lambda_1,\cdots,\lambda_s$为谱点, $f^{(i)}(\lambda_i)$为$f(z)$在矩阵A上的谱值.

例4.7.7 考查函数$f(z)$在矩阵A的谱上是否有定义，其中

$$f(z)=\frac{1}{(z-3)(z-4)},\qquad A=\begin{bmatrix}2&0&0\\0&1&1\\0&0&1\end{bmatrix}$$

若$f(z)=\frac{1}{(z-1)(z-4)}$，则$f(z)$在矩阵$A$的谱上是否有定义

定义4.7.3（谱上一致） 设复方阵A的最小多项式为$m_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2}\ (\lambda - \lambda_s)^{n_s}$, $\deg(m_A(\lambda)) = l$. 若函数$f(\lambda)$和$p(\lambda)$在谱上给定且满足

$$\begin{cases} f(\lambda_i) = p(\lambda_i) \\ f'(\lambda_i) = p'(\lambda_i) \\ \vdots \\ f^{(n_i-1)}(\lambda_i) = p^{(n_i-1)}(\lambda_i) \end{cases},\ i = 1, \cdots, s$$

则称函数$f(\lambda)$和$p(\lambda)$在矩阵A的谱上一致.

定理4.7.4 设$A \in \mathbb{C}^{n\times n}$, 幂级数$f(z)=\sum_{m=0}^{\infty} c_m z^m$与多项式$p(z)=\sum_{i=0}^{k} \beta_i z$在矩阵$A$的谱上给定, 则$f(A)=p(A)$的充分必要条件是$f(z)$和$p(z)$在矩阵$A$的谱上一致.

根据定理4.7.4的谱上一致性条件, 我们可列出l个独立方程. 由此可求解出定理4.7.3中的l个系数$\beta_0,\beta_1,\cdots,\beta_{l-1}$. 这就是计算矩阵函数的第二种方法, 我们称之为**谱上一致性方法**.

例4.7.9 设矩阵$A=\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}$, 求矩阵函数$e^{A}$.

例4.7.10 设矩阵$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 2 & -5 & 4\end{bmatrix}$, 求矩阵函数$e^{A}$.

注4: 利用谱上一致性方法计算矩阵函数不仅适用于单纯矩阵, 也适用于非单纯矩阵. 相关结论如下.

设$f(At)=\sum_{m=0}^{\infty} c'_m A^m$, 其中, $c'_m=c_m t^m$. 故含参矩阵函数$f(At)$仍可以唯一地由$l-1$次矩阵多项式$p_t(A)$表示, 即

$$f(At)=\alpha_0(t)I+\alpha_1(t)A+\cdots+\alpha_{l-1}(t)A^{l-1}$$
$$=p_t(A)$$

式中, $\alpha_i(t)(i=0,1,\cdots,l-1)$为待定含参系数.

利用Sylvester公式，可得如下方程组

$$p_t^{(j)}(\lambda_i) = t^j f^{(j)}(\lambda_i t), j = 0, \cdots, n_i - 1, i = 1, \cdots, s$$

通过求解上式方程组可确定待定的l个系数$\alpha_i(t)$, $i = 0,1,\cdots,l-1.$

例4.7.11 设矩阵$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, 求矩阵函数$e^{At}$.

注5: 利用谱上一致性方法计算矩阵函数时, 可不选用矩阵的最小多项式, 而选用矩阵的任一零化多项式. 此时须把选用的零化多项式当作矩阵的最小多项式, 然后仿照上面的方法求解即可.

例4.7.12 设函数$f(z)=\frac{1}{z}$, 矩阵$A=\begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}$, 求$f(A)$.

Sylvester公式和谱上一致性方法所需条件为$f(z)$在矩阵A的谱上给定（尽管在定义或结论中我们仍要求矩阵的谱半径小于收敛半径,实际上我们可移除这一条件）．谱上给定条件比定义4.7.1的幂级数要求弱. 因此, 我们可拓宽定义4.7.1矩阵函数的定义.

定义4.7.4（矩阵函数）设复方阵A的最小多项式为$m_A(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\ (\lambda-\lambda_s)^{n_s}$, $\deg(m_A(\lambda))=l$. 若函数$f(\lambda)$在矩阵A的谱上给定，则矩阵函数$f(A)$定义为$f(A)=p(A)$, 式中, $p(A)=\beta_0 I+\beta_1 A+\cdots+\beta_{l-1}A^{l-1}$, l个系数$\beta_0,\beta_1,\cdots,\beta_{l-1}$由以下方程组决定

$$\begin{cases} f(\lambda_i)=p(\lambda_i) \\ f'(\lambda_i)=p'(\lambda_i) \\ \vdots \\ f^{(n_i-1)}(\lambda_i)=p^{(n_i-1)}(\lambda_i) \end{cases},\quad i=1,2,\cdots,s$$

注6: 矩阵函数也可以利用Sylvester公式定义, 这里就不再赘述了. 特别地, 当函数$f(z)$能够展开为“z”的幂级数时, 矩阵函数的定义4.7.1和定义4.7.4一致.

例4.7.13 考查齐次线性差分方程

$$u_{k+3}+u_{k+2}-u_{k+1}-u_k=0$$

当$k<0$时, $u_k=0$且有$u_1=0$, $u_2=1$, $u_3=0$, 求$\{u_k\}$.

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,

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

命题4.8.1（续）

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

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

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

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$\forall t \in (a,b)$

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

（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$f(\boldsymbol{x})=f(x_1,\cdots,x_m)$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$\frac{df}{d\boldsymbol{x}}$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$\frac{dA\boldsymbol{x}}{d\boldsymbol{x}^T}$.

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

【应用】:求解最小二乘问题

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

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$X$的二阶连续可导函数. 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.

【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

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

其中参数分别为

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}, \boldsymbol{f}(t) = [1, -t, t]^T, \boldsymbol{c} = [1,0,-1]^T$$

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

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

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.

式中,

$$\boldsymbol{x}(t)=\begin{bmatrix}x_1(t)\\ \vdots\\ x_n(t)\end{bmatrix}\in\Re^n,\quad A=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\ \vdots&\vdots&\vdots&\ddots&0\\0&0&0&\cdots&1\\-a_0&-a_1&-a_2&\cdots&-a_{n-1}\end{bmatrix},\quad \boldsymbol{f}(t)=\begin{bmatrix}0\\ \vdots\\0\\u(t)\end{bmatrix}$$

$\boldsymbol{x}(t)$称为状态向量, A为系统矩阵. 此时, 可将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（即采用线性常微分方程描述的系统）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

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

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

$A\in\mathbb{R}^{n\times n}$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.

第四章 矩阵分析

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）

则线性空间$\operatorname{span}(\boldsymbol{y}_1, \cdots \boldsymbol{y}_n)$应与$\operatorname{span}(\boldsymbol{x}_1, \cdots \boldsymbol{x}_n)$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

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

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,

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

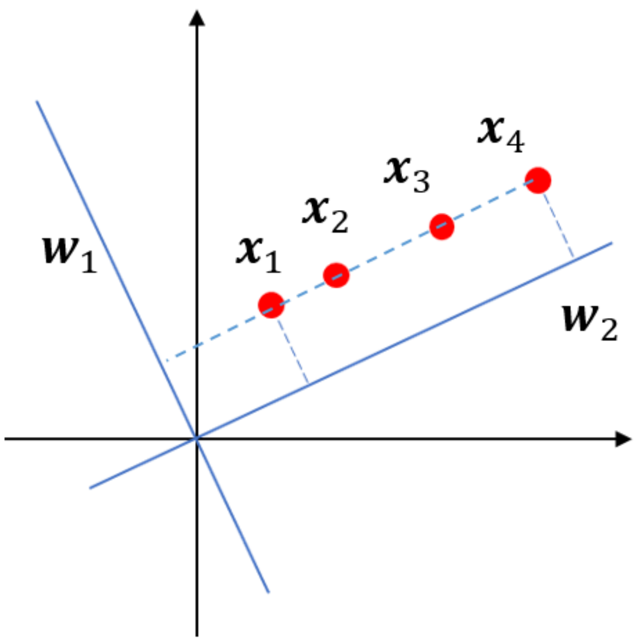
若$\|\boldsymbol{w}_k\|=1$, 则上式表示向量$\boldsymbol{x}_i$在向量$\boldsymbol{w}_k$的正交投影的长度, 即映射的$\boldsymbol{y}_i$的每一个分量是原像$\boldsymbol{x}_i$在矩阵W行向量的正交投影.

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

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

$$= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{w}^T(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{w}^T(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^T\boldsymbol{w}$$

$$= \boldsymbol{w}^T(\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\boldsymbol{i}} - \overline{\boldsymbol{x}}))^T)\boldsymbol{w}$$

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.