§4　矩阵微积分

一、函数矩阵的导数与积分

定义　若矩阵 $\boldsymbol{A}$ 的每个元素都是变量 t 的函数，即 $a_{ij}=a_{ij}(t)$，则称之为变量 t 的**函数矩阵**，记为 $\boldsymbol{A}(t)$，即 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$。

例　$\boldsymbol{A}(t)=\begin{pmatrix} t^2+1 & 1 & t \\ 0 & \mathrm{e}^t & 2 \\ 1 & 0 & \ln t \end{pmatrix}$ 是变量 t 的函数矩阵。

定义　若函数矩阵 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$ 的每个元素 $a_{ij}(t)$ 对变量 t 均可导（可微），则称 $\boldsymbol{A}(t)$ 为**可导**（或**可微**）**矩阵**，其**导数矩阵**为

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\ （或\ \boldsymbol{A}'(t)\ ）=(\tfrac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t))_{m\times n}$$

如果 $\boldsymbol{A}(t)$ 的每个元素 $a_{ij}(t)$ 都在区间 $[a,b]$ 上可积，则称 $\boldsymbol{A}(t)$ 是区间 $[a,b]$ 上的**可积矩阵**，其**积分矩阵**为

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

如果 $\boldsymbol{A}(t)$ 的每个元素 $a_{ij}(t)$ 在区间 $[a,b]$ 上连续，则称 $\boldsymbol{A}(t)$ 为区间 $[a,b]$ 上的**连续矩阵**。

显然，对常数矩阵 $\boldsymbol{A}$ 有 $\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}=\boldsymbol{O}$；若 $\boldsymbol{A}(t)$ 是区间 $[a,b]$ 上的连续矩阵，则它一定是 $[a,b]$ 上的可积矩阵。

对于上例的函数矩阵有

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\begin{pmatrix} 2t & 0 & 1 \\ 0 & \mathrm{e}^t & 0 \\ 0 & 0 & \frac{1}{t} \end{pmatrix},\quad \int_1^2 \boldsymbol{A}(t)\,\mathrm{d}t=\begin{pmatrix} \frac{10}{7} & 1 & \frac{3}{2} \\ 0 & \mathrm{e}^2-\mathrm{e} & 2 \\ 1 & 0 & 2\ln 2-1 \end{pmatrix}$$

函数矩阵的导数有下列性质：

性质 1　$\tfrac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\pm\boldsymbol{B}(t))=\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\pm\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$；

性质 2　$\tfrac{\mathrm{d}}{\mathrm{d}t}(\lambda(t)\boldsymbol{A}(t))=(\tfrac{\mathrm{d}}{\mathrm{d}t}\lambda(t))\boldsymbol{A}(t)+\lambda(t)\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$，

特别地，若 λ 为常数，则　$\tfrac{\mathrm{d}}{\mathrm{d}t}(\lambda\boldsymbol{A}(t))=\lambda\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$；

性质 3　$\tfrac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B}(t))=(\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\boldsymbol{B}(t)+\boldsymbol{A}(t)\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$；

证 设 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$， $\boldsymbol{B}(t)=(b_{ij}(t))_{n\times s}$，则 $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B}(t))$ 的 (i,j)-元素为

$$\frac{\mathrm{d}}{\mathrm{d}t}(\sum_{k=1}^{n}a_{ik}(t)b_{kj}(t))=\sum_{k=1}^{n}\frac{\mathrm{d}}{\mathrm{d}t}(a_{ik}(t)b_{kj}(t))=\sum_{k=1}^{n}[(\frac{\mathrm{d}}{\mathrm{d}t}(a_{ik}(t))b_{kj}(t)+a_{ik}(t)\frac{\mathrm{d}}{\mathrm{d}t}b_{kj}(t)]$$

$$=\sum_{k=1}^{n}(\frac{\mathrm{d}}{\mathrm{d}t}(a_{ik}(t))b_{kj}(t)+\sum_{k=1}^{n}a_{ik}(t)\frac{\mathrm{d}}{\mathrm{d}t}b_{kj}(t)$$

恰等于 $(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\boldsymbol{B}(t)+\boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$ 的 (i,j)-元素。**证毕**

特别地，若 $\boldsymbol{A}$ 是常数矩阵，则 $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}\boldsymbol{B}(t))=\boldsymbol{A}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$；若 $\boldsymbol{B}$ 是常数矩阵，则 $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B})=(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\boldsymbol{B}$。

性质 4 一般地，$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t))^m\neq m(\boldsymbol{A}(t))^{m-1}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$，等式成立的条件为 $(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\boldsymbol{A}(t)=\boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$。

例如，取 $\boldsymbol{A}(t)=\begin{pmatrix}t^2 & t\\ 0 & \mathrm{e}^t\end{pmatrix}$，则 $\boldsymbol{A}^2(t)=\begin{pmatrix}t^4 & t^3+t\,\mathrm{e}^t\\ 0 & \mathrm{e}^{2t}\end{pmatrix}$，于是

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^2(t)=\begin{pmatrix}4t^3 & 3t^2+(t+1)\mathrm{e}^t\\ 0 & 2\,\mathrm{e}^{2t}\end{pmatrix},\qquad 2\boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\begin{pmatrix}4t^3 & 2t^2+2t\,\mathrm{e}^t\\ 0 & 2\,\mathrm{e}^{2t}\end{pmatrix}$$

可见 $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^2(t)\neq 2\boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$。

性质 5 设 $\boldsymbol{A}(t)$ 是可导矩阵，则在 $\boldsymbol{A}^{-1}(t)$ 存在的区间中，$\boldsymbol{A}^{-1}(t)$ 也可导，且

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=-\boldsymbol{A}^{-1}(t)\,(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\,\boldsymbol{A}^{-1}(t)$$

证 对 $\boldsymbol{A}(t)\,\boldsymbol{A}^{-1}(t)=\boldsymbol{I}$ 求导得 $(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\,\boldsymbol{A}^{-1}(t)+\boldsymbol{A}(t)\,\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=\boldsymbol{O}$，解之即得。**证毕**

例 设 $\boldsymbol{A}(t)=\begin{pmatrix}1 & t\,\mathrm{e}^t\\ 0 & 2t\,\mathrm{e}^{2t}\end{pmatrix}$，求 $\boldsymbol{A}^{-1}(t)$ 的存在区间，并求 $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)$。

解 因为 $\det\boldsymbol{A}(t)=2t\,\mathrm{e}^{2t}$，仅当 $t=0$ 时，$\boldsymbol{A}(t)$ 奇异，故 $\boldsymbol{A}^{-1}(t)$ 的存在区间为 $(-\infty,0),(0,+\infty)$。

法 1. 由于 $\boldsymbol{A}^{-1}(t)=\dfrac{1}{2t\,\mathrm{e}^{2t}}\begin{pmatrix}2t\,\mathrm{e}^{2t} & -t\,\mathrm{e}^t\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & -\frac{1}{2}\mathrm{e}^{-t}\\ 0 & \frac{1}{2t}\mathrm{e}^{-2t}\end{pmatrix}$

所以 $$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=\begin{pmatrix}0 & \frac{1}{2}\mathrm{e}^{-t}\\ 0 & -\frac{1+2t}{2t^2}\mathrm{e}^{-2t}\end{pmatrix}$$（由定义）

法 2. $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)=-\boldsymbol{A}^{-1}(t)\,(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t))\,\boldsymbol{A}^{-1}(t)$

$$=-\begin{pmatrix}1 & -\frac{1}{2}\mathrm{e}^{-t}\\ 0 & \frac{1}{2t}\mathrm{e}^{-2t}\end{pmatrix}\begin{pmatrix}0 & (1+t)\mathrm{e}^{t}\\ 0 & 2(1+2t)\mathrm{e}^{2t}\end{pmatrix}\begin{pmatrix}1 & -\frac{1}{2}\mathrm{e}^{-t}\\ 0 & \frac{1}{2t}\mathrm{e}^{-2t}\end{pmatrix}=\begin{pmatrix}0 & \frac{1}{2}\mathrm{e}^{-t}\\ 0 & -\frac{1+2t}{2t^2}\mathrm{e}^{-2t}\end{pmatrix}$$

性质 6 设 $\boldsymbol{A}(x)$ 是变量 x 的函数矩阵，而 $x=f(t)$ 是 t 的函数，且 $\boldsymbol{A}(x)$ 与 $f(t)$ 均可导，则有链式法则

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(x)=f'(t)\,\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{A}(x)=(\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{A}(x))\,f'(t)$$

性质 7 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\boldsymbol{A}\,\mathrm{e}^{\boldsymbol{A}t}=\mathrm{e}^{\boldsymbol{A}t}\,\boldsymbol{A}$，$\frac{\mathrm{d}}{\mathrm{d}t}\sin\boldsymbol{A}t=\boldsymbol{A}\cos\boldsymbol{A}t=(\cos\boldsymbol{A}t)\boldsymbol{A}$，

$\frac{\mathrm{d}}{\mathrm{d}t}\cos\boldsymbol{A}t=-\boldsymbol{A}\sin\boldsymbol{A}t=-(\sin\boldsymbol{A}t)\boldsymbol{A}$；

证 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{I}+\frac{1}{1!}\boldsymbol{A}t+\frac{1}{2!}\boldsymbol{A}^2t^2+\cdots)=\boldsymbol{A}+\frac{1}{1!}\boldsymbol{A}^2t+\frac{1}{2!}\boldsymbol{A}^3t^2+\cdots=\boldsymbol{A}\,\mathrm{e}^{\boldsymbol{A}t}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sin\boldsymbol{A}t=\frac{\mathrm{d}}{\mathrm{d}t}(\frac{\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}-\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t}}{2\mathrm{i}})=\frac{\boldsymbol{A}\mathrm{e}^{\mathrm{i}\boldsymbol{A}t}+\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{A}t}}{2}=\boldsymbol{A}\cos\boldsymbol{A}t$$ **证毕**

注 由 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\boldsymbol{A}\,\mathrm{e}^{\boldsymbol{A}t}$，并注意 $\mathrm{e}^{\boldsymbol{O}}=\boldsymbol{I}$ 得 $\boldsymbol{A}=\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}\big|_{t=0}$ 。这一结果可用来检验所求的 $\mathrm{e}^{\boldsymbol{A}t}$ 是否正确，如当 $\boldsymbol{A}=\begin{pmatrix}3 & 1 & 0\\ -1 & 1 & 0\\ -1 & -1 & 4\end{pmatrix}$ 时，求得

$$\mathrm{e}^{\boldsymbol{A}t}=\begin{pmatrix}(1+t)\mathrm{e}^{2t} & t\,\mathrm{e}^{2t} & 0\\ -t\,\mathrm{e}^{2t} & (1-t)\mathrm{e}^{2t} & 0\\ \frac{1}{2}(\mathrm{e}^{2t}-\mathrm{e}^{4t}) & \frac{1}{2}(\mathrm{e}^{2t}-\mathrm{e}^{4t}) & \mathrm{e}^{4t}\end{pmatrix},$$

则 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\begin{pmatrix}(3+2t)\mathrm{e}^{2t} & (1+2t)\mathrm{e}^{2t} & 0\\ -(1+2t)\mathrm{e}^{2t} & (1-2t)\mathrm{e}^{2t} & 0\\ \mathrm{e}^{2t}-2\mathrm{e}^{4t} & \mathrm{e}^{2t}-2\mathrm{e}^{4t} & 4\mathrm{e}^{4t}\end{pmatrix}$，可见 $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}\big|_{t=0}=\boldsymbol{A}$ 。

关于函数矩阵的定积分有如下性质：

性质 8 $\int_a^b(\lambda\boldsymbol{A}(t)\pm\mu\boldsymbol{B}(t))\mathrm{d}t=\lambda\int_a^b\boldsymbol{A}(t)\mathrm{d}t\pm\mu\int_a^b\boldsymbol{B}(t)\mathrm{d}t$，$\lambda,\mu\in\mathbf{C}$

性质 9 $\int_a^b\boldsymbol{A}(t)\boldsymbol{B}\,\mathrm{d}t=(\int_a^b\boldsymbol{A}(t)\mathrm{d}t)\boldsymbol{B}$，$\int_a^b\boldsymbol{A}\boldsymbol{B}(t)\mathrm{d}t=\boldsymbol{A}\int_a^b\boldsymbol{B}(t)\mathrm{d}t$，其中第一式中 $\boldsymbol{B}$ 是常数矩阵，第二式中 $\boldsymbol{A}$ 是常数矩阵。

证 设 $A(t)=(a_{ij}(t))_{m\times n}$，$B=(b_{ij})_{n\times l}$，则 $\int_a^b(\sum_{k=1}^n a_{ik}(t)b_{kj})\mathrm{d}t=\sum_{k=1}^n(\int_a^b a_{ik}(t)\mathrm{d}t)b_{kj}$。

证毕

性质 10 $\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x A(t)\mathrm{d}t=A(x)$，其中 $A(t)$ 在 $[a,b]$ 上连续，$x\in[a,b]$。

性质 11 $\int_a^b A'(t)\mathrm{d}t=A(b)-A(a)$，要求 $a_{ij}'(t)$ 在 $[a,b]$ 上连续。

由于函数矩阵的导数仍是同阶的函数矩阵，如果它可导，仍可进行求导运算，不难给出函数矩阵的高阶导数定义

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}A(t)=\frac{\mathrm{d}}{\mathrm{d}t}(\frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}}A(t))$$

例 设 A 是可逆矩阵，则 $\int_0^1 \mathrm{e}^{At}\mathrm{d}t=A^{-1}(\mathrm{e}^A-I)$。

分析 $\int_0^1 \mathrm{e}^{At}\mathrm{d}t=A^{-1}\int_0^1 A\mathrm{e}^{At}\mathrm{d}t=A^{-1}\int_0^1\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{At})\mathrm{d}t=A^{-1}(\mathrm{e}^{At})\Big|_0^1=A^{-1}(\mathrm{e}^A-I)$。

二、数量函数对矩阵变量的导数

定义 设 $f=f(X)$ 是以矩阵 $X=(x_{ij})_{m\times n}$ 为自变量的 mn 元函数，规定 f 对于矩阵 X 的**导数**为

$$\frac{\mathrm{d}f}{\mathrm{d}X}=(\frac{\partial f}{\partial x_{ij}})_{m\times n}=\begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{pmatrix}$$ （其结果是一个 $m\times n$ 的函数矩阵）

特例，以 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为自变量的函数 $f(x)$ 的导数为

$$\frac{\mathrm{d}f}{\mathrm{d}x}=(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_n})^{\mathrm{T}}=\mathrm{grad}\, f(x)\quad \text{(即为 } f(x) \text{ 的梯度)}$$

例 已知 $f(x)=a^{\mathrm{T}}x=x^{\mathrm{T}}a$，其中 $a=(a_1,\cdots,a_n)^{\mathrm{T}}$ 是已知向量，$x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是向量变量，求 $\frac{\mathrm{d}f}{\mathrm{d}x}$。

解 $f(x)=a^{\mathrm{T}}x=a_1x_1+\cdots+a_nx_n$，因为 $\frac{\partial f}{\partial x_i}=a_i$（$i=1,2,\cdots,n$），所以

$$\frac{\mathrm{d}f}{\mathrm{d}x}=(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_n})^{\mathrm{T}}=(a_1,\cdots,a_n)^{\mathrm{T}}=a$$

$n\times m$

例 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 已知，$\boldsymbol{X}=(x_{ij})_{\cancel{m\times n}}$ 为矩阵变量，$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{X})$，求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$。

解 $$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{X})=\sum_{s=1}^{m}\sum_{t=1}^{n}a_{st}x_{ts}=a_{ji}x_{ij}+\sum_{s=1}^{m}\sum_{\substack{t=1\\ s\neq j\text{或}t\neq i}}^{n}a_{st}x_{ts}$$

因为 $\dfrac{\partial f}{\partial x_{ij}}=a_{ji}$，所以 $$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times m}=(a_{ji})_{n\times m}=\boldsymbol{A}^{\mathrm{T}}$$

例 设 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 已知，$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是向量变量，$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$。

解 $$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\sum_{s=1}^{m}\sum_{t=1}^{n}a_{st}x_sx_t$$

$$=x_1\sum_{t=1}^{n}a_{1t}x_t+x_2\sum_{t=1}^{n}a_{2t}x_t+\cdots+x_i\sum_{t=1}^{n}a_{it}x_t+\cdots+x_n\sum_{t=1}^{n}a_{nt}x_t$$

因为 $$\frac{\partial f}{\partial x_i}=a_{1i}x_1+\cdots+a_{i-1,i}x_{i-1}+(a_{ii}x_i+\sum_{t=1}^{n}a_{it}x_t)+a_{i+1,i}x_{i+1}+\cdots+a_{ni}x_n$$

$$=\sum_{s=1}^{n}a_{si}x_s+\sum_{t=1}^{n}a_{it}x_t$$

所以 $$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial f}{\partial x_1}\\ \vdots\\ \dfrac{\partial f}{\partial x_n}\end{pmatrix}=\begin{pmatrix}\sum_{s=1}^{n}a_{s1}x_s+\sum_{t=1}^{n}a_{1t}x_t\\ \vdots\\ \sum_{s=1}^{n}a_{sn}x_s+\sum_{t=1}^{n}a_{nt}x_t\end{pmatrix}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{x}=(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A})\boldsymbol{x}$$

特例，当 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$ 时，即 $\boldsymbol{A}$ 对称时，$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$。

例 设 $\boldsymbol{X}=(x_{ij})_{n\times n}$ 是矩阵变量，$f(\boldsymbol{X})=\det\boldsymbol{X}$，试求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}$。

解 设 X_{ij} 是 $\det\boldsymbol{X}$ 中元素 x_{ij} 的代数余子式，则

$$f(\boldsymbol{X})=\det\boldsymbol{X}=x_{i1}X_{i1}+\cdots+x_{ij}X_{ij}+\cdots+x_{in}X_{in}$$

因为 $\dfrac{\partial f}{\partial x_{ij}}=X_{ij}$，所以 $$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times n}=(X_{ij})_{n\times n}=(\boldsymbol{X}^{*})^{\mathrm{T}}$$

当 $\boldsymbol{X}$ 可逆时，$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=((\det\boldsymbol{X})\boldsymbol{X}^{-1})^{\mathrm{T}}=(\det\boldsymbol{X})\boldsymbol{X}^{-\mathrm{T}}$。

例 已知 $\boldsymbol{X}=\begin{pmatrix}t_1&t_2&t_3\\ t_4&t_5&t_6\end{pmatrix}$，$f(\boldsymbol{X})=t_1t_6+t_2t_5+t_3t_4$，则 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\underline{\begin{pmatrix}t_6&t_5&t_4\\ t_3&t_2&t_1\end{pmatrix}}$。

例　已知 $\boldsymbol{X}=(x_{ij})_{n\times n}$，　$f(\boldsymbol{X})=\operatorname{tr}\boldsymbol{X}$，则 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\underline{\boldsymbol{I}_n}$　。

作为应用，下面研究线性方程组的最小二乘解问题。

例　已知 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$，$\boldsymbol{b}\in\mathbf{R}^{m}$，对于矛盾方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，使得 $f(\boldsymbol{x})=\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2$ 为最小的向量 $\boldsymbol{x}^{(0)}$ 称为**最小二乘解**，试导出最小二乘解所满足的方程组。

解　$\boldsymbol{x}^{(0)}$ 使 $f(\boldsymbol{x})$ 达到极小，从而应有 $\left.\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}=\boldsymbol{0}$。因为

$$f(\boldsymbol{x})=\|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2=(\boldsymbol{Ax}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax}-\boldsymbol{b})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{Ax}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$$

由前几例得，$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$。于是 $\left.\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}^{(0)}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}^{(0)}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{0}$

即　$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}^{(0)}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$，称 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 为**法方程组**，它是最小二乘解所满足的方程组。

三、矩阵值函数对矩阵变量的导数

定义　设矩阵 $\boldsymbol{F}(\boldsymbol{X})=\begin{pmatrix} f_{11}(\boldsymbol{X}) & \cdots & f_{1s}(\boldsymbol{X}) \\ \vdots & & \vdots \\ f_{r1}(\boldsymbol{X}) & \cdots & f_{rs}(\boldsymbol{X}) \end{pmatrix}$ 的元素 $f_{ij}(\boldsymbol{X})$ 是矩阵变量 $\boldsymbol{X}=(x_{ij})_{m\times n}$ 的函数，规定

$$\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{X}}=\begin{pmatrix} \dfrac{\partial\boldsymbol{F}}{\partial x_{11}} & \cdots & \dfrac{\partial\boldsymbol{F}}{\partial x_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial\boldsymbol{F}}{\partial x_{m1}} & \cdots & \dfrac{\partial\boldsymbol{F}}{\partial x_{mn}} \end{pmatrix}，\text{其中}\quad \frac{\partial\boldsymbol{F}}{\partial x_{ij}}=\begin{pmatrix} \dfrac{\partial f_{11}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{1s}}{\partial x_{ij}} \\ \vdots & & \vdots \\ \dfrac{\partial f_{r1}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{rs}}{\partial x_{ij}} \end{pmatrix}$$

即结果是 $(mr)\times(ns)$ 矩阵。

例　已知 $\boldsymbol{A}=(a_{ij})_{m\times n}$，　$\boldsymbol{x}=(x_1,\cdots,x_m)^{\mathrm{T}}$，且 $\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}$，求 $\dfrac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{x}}$。

解　$\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}=(\sum\limits_{k=1}^{m}x_k a_{k1},\sum\limits_{k=1}^{m}x_k a_{k2},\cdots,\sum\limits_{k=1}^{m}x_k a_{kn})$，因为 $\dfrac{\partial\boldsymbol{F}}{\partial x_i}=(a_{i1},a_{i2},\cdots,a_{in})$，

所以
$$\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{x}}=\begin{pmatrix}\frac{\partial \boldsymbol{F}}{\partial x_1}\\ \vdots\\ \frac{\partial \boldsymbol{F}}{\partial x_m}\end{pmatrix}=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}=\boldsymbol{A}$$

§5 矩阵函数的应用——求解一阶常系数线性微分方程组

研究如下的一阶常系数线性微分方程组初值问题:
$$\begin{cases}\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t)\\ \boldsymbol{x}(t_0)=\boldsymbol{x}_0\end{cases}$$

其中 $\boldsymbol{A}$ 是 n 阶常数矩阵，$\boldsymbol{x}(t)$ 是 n 维函数向量，$\boldsymbol{f}(t)$ 是 n 维已知的函数向量，$\boldsymbol{x}_0$ 是 n 维初始向量。如果 $\boldsymbol{f}(t)\equiv\boldsymbol{0}$，则称该方程组是**齐次**的，否则称为**非齐次**的。

为求其解，将方程组改写为 $\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}t}-\boldsymbol{A}\boldsymbol{x}(t)=\boldsymbol{f}(t)$

两边左乘 $\mathrm{e}^{-\boldsymbol{A}t}$ 得
$$\mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}t}-\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{A}\boldsymbol{x}(t)=\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)$$

整理得
$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t))=\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{f}(t)$$

在 $[t_0,t]$ 上积分得
$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)-\mathrm{e}^{-\boldsymbol{A}t_0}\boldsymbol{x}(t_0)=\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

于是
$$\begin{aligned}\boldsymbol{x}(t)&=\mathrm{e}^{\boldsymbol{A}t}\mathrm{e}^{-\boldsymbol{A}t_0}\boldsymbol{x}_0+\mathrm{e}^{\boldsymbol{A}t}\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau\\&=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0+\mathrm{e}^{\boldsymbol{A}t}\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau\end{aligned}$$

这就是求解公式，可见其关键是求 $\mathrm{e}^{\boldsymbol{A}t}$。该公式有如下两种特殊情形:

1） $\boldsymbol{f}(t)\equiv\boldsymbol{0}$，则 $\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0$；

2） t_0=0，则 $\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_0+\mathrm{e}^{\boldsymbol{A}t}\int_{0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$。

另外,容易验证
$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{c}+\mathrm{e}^{\boldsymbol{A}t}\int_{t_0}^{t}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau$$

其中 $\boldsymbol{c}$ 为任意 n 维列向量，满足 $\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t)$，因此它是通解。

又 $\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{c}$ 是 $\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}\,t}=\boldsymbol{A}\boldsymbol{x}(t)$ 的通解，所以非齐次微分方程组的通解等于它的特解加上对应齐次微分方程组的通解。

例 用矩阵函数方法求解微分方程组 $\begin{cases}\frac{\mathrm{d}}{\mathrm{d}\,t}x_1=2x_1+x_2+4x_3+4t\\ \frac{\mathrm{d}}{\mathrm{d}\,t}x_2=2x_2\\ \frac{\mathrm{d}}{\mathrm{d}\,t}x_3=3x_2+x_3\\ x_1(0)=2,\ x_2(0)=1,\ x_3(0)=3\end{cases}$。

解 写成矩阵形式 $\begin{cases}\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}\,t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{f}(t)\\ \boldsymbol{x}(0)=\boldsymbol{x}_0\end{cases}$，其中

$$\boldsymbol{A}=\begin{pmatrix}2&1&4\\0&2&0\\0&3&1\end{pmatrix},\quad \boldsymbol{f}(t)=\begin{pmatrix}4t\\0\\0\end{pmatrix},\quad \boldsymbol{x}_0=\begin{pmatrix}2\\1\\3\end{pmatrix}$$

可求得 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-2)^2(\lambda-1)$，$\boldsymbol{A}$ 的特征值为 $\lambda_1=\lambda_2=2,\ \lambda_3=1$。设

$$r(\lambda)=b_0+b_1\lambda+b_2\lambda^2$$

由 $\begin{cases}r(2)=b_0+2b_1+4b_2=\mathrm{e}^{2t}\\ r'(2)=b_1+4b_2=t\,\mathrm{e}^{2t}\\ r(1)=b_0+b_1+b_2=\mathrm{e}^{t}\end{cases}$，解得 $\begin{cases}b_0=2t\,\mathrm{e}^{2t}-3\mathrm{e}^{2t}+4\mathrm{e}^{t}\\ b_1=-3t\,\mathrm{e}^{2t}+4\mathrm{e}^{2t}-4\mathrm{e}^{t}\\ b_2=t\,\mathrm{e}^{2t}-\mathrm{e}^{2t}+\mathrm{e}^{t}\end{cases}$，

所以 $$\mathrm{e}^{\boldsymbol{A}t}=b_0\boldsymbol{I}+b_1\boldsymbol{A}+b_2\boldsymbol{A}^2=\begin{pmatrix}\mathrm{e}^{2t}&12\mathrm{e}^{t}-12\mathrm{e}^{2t}+13t\,\mathrm{e}^{2t}&-4\mathrm{e}^{t}+4\mathrm{e}^{2t}\\0&\mathrm{e}^{2t}&0\\0&-3\mathrm{e}^{t}+3\mathrm{e}^{2t}&\mathrm{e}^{t}\end{pmatrix}$$

依次计算

$$\mathrm{e}^{\boldsymbol{A}t}\,\boldsymbol{x}_0=\begin{pmatrix}2\mathrm{e}^{2t}+13t\,\mathrm{e}^{2t}\\ \mathrm{e}^{2t}\\ 3\mathrm{e}^{2t}\end{pmatrix},\quad \int_0^t\mathrm{e}^{-\boldsymbol{A}\tau}\,\boldsymbol{f}(\tau)\,\mathrm{d}\,\tau=\int_0^t\begin{pmatrix}4\tau\,\mathrm{e}^{-2\tau}\\0\\0\end{pmatrix}\mathrm{d}\,\tau=\begin{pmatrix}-2t\,\mathrm{e}^{-2t}-\mathrm{e}^{-2t}+1\\0\\0\end{pmatrix},$$

$$\mathrm{e}^{\boldsymbol{A}t}\int_0^t\mathrm{e}^{-\boldsymbol{A}\tau}\,\boldsymbol{f}(\tau)\,\mathrm{d}\,\tau=\begin{pmatrix}\mathrm{e}^{2t}-2t-1\\0\\0\end{pmatrix},$$

故 $$\boldsymbol{x}(t)=\begin{pmatrix}2\mathrm{e}^{2t}+13t\,\mathrm{e}^{2t}\\ \mathrm{e}^{2t}\\ 3\mathrm{e}^{2t}\end{pmatrix}+\begin{pmatrix}\mathrm{e}^{2t}-2t-1\\0\\0\end{pmatrix}=\begin{pmatrix}3\mathrm{e}^{2t}+13t\,\mathrm{e}^{2t}-2t-1\\ \mathrm{e}^{2t}\\ 3\mathrm{e}^{2t}\end{pmatrix}。$$

例 已知 $\boldsymbol{A}=\begin{pmatrix}-1&2&1\\1&0&1\\1&2&-1\end{pmatrix}$，$\boldsymbol{b}(t)=\begin{pmatrix}\mathrm{e}^{-2t}\\0\\-\mathrm{e}^{-2t}\end{pmatrix}$，$\boldsymbol{x}(0)=\begin{pmatrix}0\\-1\\2\end{pmatrix}$。

1）求 $\mathrm{e}^{\boldsymbol{A}t}$；

2）用矩阵函数方法求微分方程 $\dfrac{\mathrm{d}\,\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{b}(t)$ 满足初始条件 $\boldsymbol{x}(0)$ 的解.

解 1）$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda+2)^2(\lambda-2)$。

法 1. 可求得 $\boldsymbol{P}=\begin{pmatrix}2&1&1\\-1&0&1\\0&-1&1\end{pmatrix}$，$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}-2&&\\&-2&\\&&2\end{pmatrix}$，

所以 $$\mathrm{e}^{\boldsymbol{A}t}=\boldsymbol{P}\begin{pmatrix}\mathrm{e}^{-2t}&&\\&\mathrm{e}^{-2t}&\\&&\mathrm{e}^{2t}\end{pmatrix}\boldsymbol{P}^{-1}=\frac{1}{4}\begin{pmatrix}3\mathrm{e}^{-2t}+\mathrm{e}^{2t}&-2\mathrm{e}^{-2t}+2\mathrm{e}^{2t}&-\mathrm{e}^{-2t}+\mathrm{e}^{2t}\\-\mathrm{e}^{-2t}+\mathrm{e}^{2t}&-2\mathrm{e}^{-2t}+2\mathrm{e}^{2t}&-\mathrm{e}^{-2t}+\mathrm{e}^{2t}\\-\mathrm{e}^{-2t}+\mathrm{e}^{2t}&-2\mathrm{e}^{-2t}+2\mathrm{e}^{2t}&3\mathrm{e}^{-2t}+\mathrm{e}^{2t}\end{pmatrix}$$

法 2. 可求得 $m_{\boldsymbol{A}}=(\lambda+2)(\lambda-2)$。设 $r(\lambda)=b_0+b_1\lambda$，由

$$\begin{cases}r(-2)=b_0-2b_1=\mathrm{e}^{-2t}\\ r(2)=b_0+2b_1=\mathrm{e}^{2t}\end{cases}，\quad 解得\begin{cases}b_0=\frac{1}{2}(\mathrm{e}^{-2t}+\mathrm{e}^{2t})\\ b_1=\frac{1}{4}(\mathrm{e}^{2t}-\mathrm{e}^{-2t})\end{cases}$$

于是 $$\mathrm{e}^{\boldsymbol{A}t}=b_0\boldsymbol{I}+b_1\boldsymbol{A}=\begin{pmatrix}b_0-b_1&2b_1&b_1\\b_1&b_0&b_1\\b_1&2b_1&b_0-b_1\end{pmatrix}=\cdots$$

法 3. 设 $r(\lambda)=b_0+b_1\lambda+b_2\lambda^2$，由

$$\begin{cases}r(-2)=b_0-2b_1+4b_2=\mathrm{e}^{-2t}\\ r'(-2)=b_1-4b_2=t\,\mathrm{e}^{-2t}\\ r(2)=b_0+2b_1+4b_2=\mathrm{e}^{2t}\end{cases}，\quad 解得\begin{cases}b_0=t\,\mathrm{e}^{-2t}+\frac{1}{4}\mathrm{e}^{2t}+\frac{3}{4}\mathrm{e}^{-2t}\\ b_1=\frac{1}{4}(\mathrm{e}^{2t}-\mathrm{e}^{-2t})\\ b_2=-\frac{1}{4}t\,\mathrm{e}^{-2t}+\frac{1}{16}(\mathrm{e}^{2t}-\mathrm{e}^{-2t})\end{cases}，$$

故 $$\mathrm{e}^{\boldsymbol{A}t}=b_0\boldsymbol{I}+b_1\boldsymbol{A}+b_2\boldsymbol{A}^2=\begin{pmatrix}b_0-b_1+4b_2&2b_1&b_1\\b_1&b_0+4b_2&b_1\\b_1&2b_1&b_0-b_1+4b_2\end{pmatrix}=\cdots$$

2） 计算 $\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{b}(\tau)=\begin{pmatrix}1\\0\\-1\end{pmatrix}$，$\int_0^t\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{b}(\tau)\,\mathrm{d}\tau=\begin{pmatrix}t\\0\\-t\end{pmatrix}$，

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\,\boldsymbol{x}(0)+\mathrm{e}^{\boldsymbol{A}t}\int_0^t \mathrm{e}^{-\boldsymbol{A}\tau}\,\boldsymbol{b}(\tau)\,\mathrm{d}\,\tau=\begin{pmatrix}0\\-\mathrm{e}^{-2t}\\2\mathrm{e}^{-2t}\end{pmatrix}+\begin{pmatrix}t\,\mathrm{e}^{-2t}\\0\\-t\,\mathrm{e}^{-2t}\end{pmatrix}=\mathrm{e}^{-2t}\begin{pmatrix}t\\-1\\2-t\end{pmatrix}。$$