MATH10003.01 高等线性代数 Fall 2025

Lecture 2 矩阵运算与向量范数 — 2025.09.16

教授：邵美悦

Scribe: 老人机

你们的线性代数白学了！ —— smy

1 大纲

1. 矩阵运算
2. 向量范数

2 矩阵运算

矩阵求导 我们约定如下记号：

- 设有向量值映照 $\boldsymbol{f}:\mathbb{R}^m\to\mathbb{R}^n$，我们记 $\boldsymbol{f}'(\boldsymbol{x})$ 为其关于 $\boldsymbol{x}$ 的导数，且形状为 $\mathbb{R}^{m\times n}$，即为 Jacobi 矩阵 $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})$ 的转置 $(\boldsymbol{D}\boldsymbol{f})^{\mathrm{T}}(\boldsymbol{x})$，记其 Hessian 矩阵为 $\boldsymbol{f}''(\boldsymbol{x})$ 或 $\boldsymbol{H}_f(\boldsymbol{x})$；
- 设有多元函数 $f:\mathbb{R}^n\to\mathbb{R}$，仿照向量值映照，我们记 $\boldsymbol{f}'(\boldsymbol{x})$ 为其关于 $\boldsymbol{x}$ 的导数，且形状为 $\mathbb{R}^n$，即为梯度 $\nabla f(\boldsymbol{x})$，亦为 Jacobi 矩阵的转置 $(\boldsymbol{D}f)^{\mathrm{T}}(\boldsymbol{x})$；
- 设有矩阵函数 $f:\mathbb{R}^{m\times n}\to\mathbb{R}$，我们记 $f'(\boldsymbol{X})$ 为其关于矩阵 $\boldsymbol{X}$ 的导数，且形状为 $\mathbb{R}^{m\times n}$，即为梯度 $\nabla f(\boldsymbol{X})$.

下面我们给出一些常用的矩阵求导公式。

命题 2.1. 设有多元函数 $h:\mathbb{R}^n\to\mathbb{R}$，向量值映照 $\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^n, g:\mathbb{R}^n\to\mathbb{R}^n$，且满足

$$h(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})=\langle\boldsymbol{f}(\boldsymbol{x}),\boldsymbol{g}(\boldsymbol{x})\rangle,$$

则其导数满足

$$h'(\boldsymbol{x})=\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{g}'(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x}).$$

即我们仍可以使用积的求导法则：前导后不导，后导前不导。

证明. 我们可以将内积形式的函数 h 写成如下求和形式：

$$h(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(\boldsymbol{x}) g_i(\boldsymbol{x}).$$

然后，两边对 $\boldsymbol{x}$ 求导，得到

$$\boldsymbol{h}'(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\boldsymbol{f}_i'(\boldsymbol{x}) g_i(\boldsymbol{x}) + \boldsymbol{g}_i'(\boldsymbol{x}) f_i(\boldsymbol{x}) \right).$$

现在，我们将其还原为矩阵形式。注意到 $\boldsymbol{h}'(\boldsymbol{x}), \boldsymbol{f}_i'(\boldsymbol{x}), \boldsymbol{g}_i'(\boldsymbol{x})$ 都是向量 $\mathbb{R}^n$，而 $f_i(\boldsymbol{x}), g_i(\boldsymbol{x})$ 都是标量，因此我们可以将上式看成对 $\boldsymbol{f}_i'(\boldsymbol{x})$ 和 $\boldsymbol{g}_i'(\boldsymbol{x})$ 的线性组合。因此，我们可以将上式写成

$$\boldsymbol{h}'(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x}) \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{g}'(\boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x}).$$

而上式在形式上正好符合积的求导法则。 □

命题 2.2. 设有向量值映照 $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$，有 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$，则其导数为 $\boldsymbol{f}'(\boldsymbol{x}) = \boldsymbol{A}^{\mathrm{T}}$.

证明. 我们将向量值映照 $\boldsymbol{f}$ 看成对 $\boldsymbol{A}$ 的各个列向量 $\boldsymbol{A}_i$ 的线性组合，即

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{A}_i x_i.$$

于是，两边对 x_i 求偏导，可有

$$\frac{\partial \boldsymbol{f}}{\partial x_i}(\boldsymbol{x}) = \boldsymbol{A}_i.$$

于是，我们可以将 $\boldsymbol{f}'(\boldsymbol{x})$ 写成

$$\boldsymbol{f}'(\boldsymbol{x}) = \begin{bmatrix} \left(\frac{\partial \boldsymbol{f}}{\partial x_1}\right)^{\mathrm{T}}(\boldsymbol{x}) \\ \left(\frac{\partial \boldsymbol{f}}{\partial x_2}\right)^{\mathrm{T}}(\boldsymbol{x}) \\ \vdots \\ \left(\frac{\partial \boldsymbol{f}}{\partial x_n}\right)^{\mathrm{T}}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1^{\mathrm{T}} \\ \boldsymbol{A}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{A}_n^{\mathrm{T}} \end{bmatrix} = \boldsymbol{A}^{\mathrm{T}}.$$

□

利用上面的求导法则，我们可以验证如下结论：

- 设有 $f(\boldsymbol{x}) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$，则 $\boldsymbol{f}'(\boldsymbol{x}) = \boldsymbol{y}_{\boldsymbol{x}}'\boldsymbol{x} + \boldsymbol{x}_{\boldsymbol{x}}'\boldsymbol{y} = \boldsymbol{y}$.
- 设有 $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n$，则 $\boldsymbol{f}'(\boldsymbol{x}) = \boldsymbol{x}_{\boldsymbol{x}}'\boldsymbol{x} + \boldsymbol{x}_{\boldsymbol{x}}'\boldsymbol{x} = 2\boldsymbol{x}$.

- 设有 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{A}\in\mathbb{R}^{n\times n}$，则 $\boldsymbol{f}'(\boldsymbol{x})=\boldsymbol{x}'_{\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x})+(\boldsymbol{A}\boldsymbol{x})'_{\boldsymbol{x}}\boldsymbol{x}=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}$，二阶导数为 $\boldsymbol{f}''(\boldsymbol{x})=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}$.

下面我们补充一个重要的结论。

命题 2.3. 设 $U,V\in\mathbb{C}^{m\times n}$，则有

$$\mathrm{tr}(U^{\mathrm{T}}V)=\mathrm{vec}(U)^{\mathrm{T}}\mathrm{vec}(V)=\sum_{i,j}u_{ij}v_{ij}.$$

证明. 直接展开即可。 □

由上面这个结论我们可以获得一些矩阵函数的求导公式：

命题 2.4.

- 设有矩阵函数 $f:\mathbb{R}^{m\times n}\to\mathbb{R}$，且 $f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}),\boldsymbol{A}\in\mathbb{R}^{m\times n}$，则 $\boldsymbol{f}'(\boldsymbol{X})=\boldsymbol{A}$;
- 设有矩阵函数 $g:\mathbb{R}^{m\times n}\to\mathbb{R}$，且 $g(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$，则有 $\boldsymbol{g}'(\boldsymbol{X})=2\boldsymbol{X}$;
- 设有矩阵函数 $h:\mathbb{R}^{m\times n}\to\mathbb{R}$，且 $h(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}),\boldsymbol{A}\in\mathbb{R}^{m\times m}$，则有 $\boldsymbol{h}'(\boldsymbol{X})=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}$.

证明.

$$\begin{aligned}
f(\boldsymbol{X})&=\langle\mathrm{vec}(\boldsymbol{A}),\mathrm{vec}(\boldsymbol{X})\rangle,\quad \boldsymbol{f}'(\boldsymbol{X})=\boldsymbol{A};\\
g(\boldsymbol{X})&=\langle\mathrm{vec}(\boldsymbol{X}),\mathrm{vec}(\boldsymbol{X})\rangle,\quad \boldsymbol{g}'(\boldsymbol{X})=2\boldsymbol{X};\\
h(\boldsymbol{X})&=\langle\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}),\mathrm{vec}(\boldsymbol{X})\rangle,\quad \boldsymbol{h}'(\boldsymbol{X})=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}.
\end{aligned}$$

□

我们还可以获得多元函数、向量值映照和矩阵函数的无限小增量公式。

定理 2.5. （多元函数、向量值映照、矩阵函数的无限小增量公式）设有多元函数 $f:\mathbb{R}^n\to\mathbb{R}$，向量值映照 $\boldsymbol{g}:\mathbb{R}^m\to\mathbb{R}^n$，矩阵函数 $\boldsymbol{h}:\mathbb{R}^{m\times n}\to\mathbb{R}$，则有

- 多元函数的无限小增量公式

$$f(\boldsymbol{x}+\Delta\boldsymbol{x})=f(\boldsymbol{x})+(\boldsymbol{f}'(\boldsymbol{x}))^{\mathrm{T}}\Delta\boldsymbol{x}+\frac{1}{2}\Delta\boldsymbol{x}^{\mathrm{T}}\boldsymbol{f}''(\boldsymbol{x})\Delta\boldsymbol{x}+o(\|\Delta\boldsymbol{x}\|_2^2);$$

- 向量值映照的无限小增量公式（更高阶的结果可以参考**力学数学-谢锡麟**老师的《现代张量分析及其在连续介质力学中的应用》，或报名他的超难暑期课学习相关内容）

$$\boldsymbol{g}(\boldsymbol{x}+\Delta\boldsymbol{x})=\boldsymbol{g}(\boldsymbol{x})+(\boldsymbol{g}'(\boldsymbol{x}))^{\mathrm{T}}\Delta\boldsymbol{x}+O(\|\Delta\boldsymbol{x}\|_2);$$

- 矩阵函数的无限小增量公式

$$\boldsymbol{h}(\boldsymbol{X}+\Delta\boldsymbol{X})=\boldsymbol{h}(\boldsymbol{X})+\mathrm{tr}((\boldsymbol{h}'(\boldsymbol{X}))^{\mathrm{T}}\Delta\boldsymbol{X})+O(\|\Delta\boldsymbol{X}\|_2).$$

矩阵运算补充 我们补充一些常用的矩阵运算。

Hadamard 积 设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m\times n}$，则它们的 Hadamard 积定义为

$$\boldsymbol{A} \circ \boldsymbol{B} = (a_{ij}b_{ij})_{m\times n}.$$

容易验证，Hadamard 乘积满足交换律、结合律和分配律，并与数乘相容，存在零元与单位元。

那么，Hadamard 乘积有什么用呢？我们可以用它来离散化表示二元函数的乘积。

例 2.6. 设有二元函数 $f(x,y), g(x,y) \in \mathbb{R}, x \in \mathscr{D}_x, y \in \mathscr{D}_y$，如果我们用网格对 $\mathscr{D}_x \times \mathscr{D}_y$ 进行覆盖，取网格中的一个长方形，则对长方形上的节点 (x_i, y_i)，我们可以定义矩阵 $\boldsymbol{F} = [f(x_i, y_j)]_{m\times n}, \boldsymbol{G} = [g(x_i, y_j)]_{m\times n}$，然后节点上的函数乘积就可以用 $\boldsymbol{F}$ 和 $\boldsymbol{G}$ 的 *Hadamard* 积来表示，即

$$\boldsymbol{F} \circ \boldsymbol{G} = [f(x_i, y_j)g(x_i, y_j)]_{m\times n}.$$

另外一个用途是 Schur 乘积定理，在此我们仅仅给出定理的内容，证明留待后续课程。

定理 2.7. （*Schur* 乘积定理）设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n\times n}$，则有

- 若 $\boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0$，则 $\boldsymbol{A} \circ \boldsymbol{B} \succeq 0$;
- 若 $\boldsymbol{A} \succ 0, \boldsymbol{B} \succ 0$，则 $\boldsymbol{A} \circ \boldsymbol{B} \succ 0$.

Kronecker 积 设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{B} \in \mathbb{C}^{p\times q}$，则它们的 Kronecker 积定义为

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \in \mathbb{C}^{mp\times nq}.$$

那么它有哪些运算性质呢？

（1）结合律：$\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$.

证明. 要证明这个结论，我们需要先引入分块矩阵的 Kronecker 乘积。

设有 $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$，则 $\boldsymbol{A} \otimes \boldsymbol{B}$ 可以写成

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11} \otimes \boldsymbol{B} & \boldsymbol{A}_{12} \otimes \boldsymbol{B} \\ \boldsymbol{A}_{21} \otimes \boldsymbol{B} & \boldsymbol{A}_{22} \otimes \boldsymbol{B} \end{bmatrix}.$$

上面这个结论是显然的，把每一项都展开后就是 Kronecker 乘积的定义。

现在，我们来证明结合律。设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{B} \in \mathbb{C}^{p\times q}, \boldsymbol{C} \in \mathbb{C}^{r\times s}$，则

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}$$

如果我们将其看作是分块矩阵，利用分块矩阵的 Kronecker 乘积，我们可以得到

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \begin{bmatrix} a_{11}\boldsymbol{B}\otimes\boldsymbol{C} & a_{12}\boldsymbol{B}\otimes\boldsymbol{C} & \cdots & a_{1n}\boldsymbol{B}\otimes\boldsymbol{C} \\ a_{21}\boldsymbol{B}\otimes\boldsymbol{C} & a_{22}\boldsymbol{B}\otimes\boldsymbol{C} & \cdots & a_{2n}\boldsymbol{B}\otimes\boldsymbol{C} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B}\otimes\boldsymbol{C} & a_{m2}\boldsymbol{B}\otimes\boldsymbol{C} & \cdots & a_{mn}\boldsymbol{B}\otimes\boldsymbol{C} \end{bmatrix}.$$

现在，我们来计算 $\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$，我们将 $(\boldsymbol{B} \otimes \boldsymbol{C})$ 视为一个整体，则有

$$\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = \begin{bmatrix} a_{11}(\boldsymbol{B}\otimes\boldsymbol{C}) & a_{12}(\boldsymbol{B}\otimes\boldsymbol{C}) & \cdots & a_{1n}(\boldsymbol{B}\otimes\boldsymbol{C}) \\ a_{21}(\boldsymbol{B}\otimes\boldsymbol{C}) & a_{22}(\boldsymbol{B}\otimes\boldsymbol{C}) & \cdots & a_{2n}(\boldsymbol{B}\otimes\boldsymbol{C}) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\boldsymbol{B}\otimes\boldsymbol{C}) & a_{m2}(\boldsymbol{B}\otimes\boldsymbol{C}) & \cdots & a_{mn}(\boldsymbol{B}\otimes\boldsymbol{C}) \end{bmatrix}$$

这和上面的结果完全相同，因此我们证明了结合律。 □

（2）分配律：

$$(\boldsymbol{A}+\boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C};$$
$$\boldsymbol{A} \otimes (\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}.$$

证明. 直接展开即可。 □

（3）$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$.

证明. 利用分块矩阵的技巧。我们对 A 进行行分块，对 C 进行列分块，则有

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_m \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 & \cdots & \boldsymbol{C}_n \end{bmatrix}.$$

于是，我们可以将 $\boldsymbol{A}\otimes\boldsymbol{B}$ 和 $\boldsymbol{C}\otimes\boldsymbol{D}$ 写成

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{bmatrix}\boldsymbol{a}_1^{\mathrm{T}}\otimes\boldsymbol{B}\\ \boldsymbol{a}_2^{\mathrm{T}}\otimes\boldsymbol{B}\\ \vdots\\ \boldsymbol{a}_m^{\mathrm{T}}\otimes\boldsymbol{B}\end{bmatrix},\quad \boldsymbol{C}\otimes\boldsymbol{D}=\begin{bmatrix}\boldsymbol{c}_1\otimes\boldsymbol{D} & \boldsymbol{c}_2\otimes\boldsymbol{D} & \cdots & \boldsymbol{c}_n\otimes\boldsymbol{D}\end{bmatrix}.$$

现在，我们来计算 $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})$，有

$$\begin{aligned}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})&=\begin{bmatrix}(\boldsymbol{a}_1^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_1\otimes\boldsymbol{D}) & (\boldsymbol{a}_1^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_2\otimes\boldsymbol{D}) & \cdots & (\boldsymbol{a}_1^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_n\otimes\boldsymbol{D})\\ (\boldsymbol{a}_2^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_1\otimes\boldsymbol{D}) & (\boldsymbol{a}_2^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_2\otimes\boldsymbol{D}) & \cdots & (\boldsymbol{a}_2^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_n\otimes\boldsymbol{D})\\ \vdots & \vdots & & \vdots\\ (\boldsymbol{a}_m^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_1\otimes\boldsymbol{D}) & (\boldsymbol{a}_m^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_2\otimes\boldsymbol{D}) & \cdots & (\boldsymbol{a}_m^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_n\otimes\boldsymbol{D})\end{bmatrix}\\ &=\begin{bmatrix}(\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{c}_1)(\boldsymbol{B}\boldsymbol{D}) & (\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{c}_2)(\boldsymbol{B}\boldsymbol{D}) & \cdots & (\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{c}_n)(\boldsymbol{B}\boldsymbol{D})\\ (\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{c}_1)(\boldsymbol{B}\boldsymbol{D}) & (\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{c}_2)(\boldsymbol{B}\boldsymbol{D}) & \cdots & (\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{c}_n)(\boldsymbol{B}\boldsymbol{D})\\ \vdots & \vdots & & \vdots\\ (\boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{c}_1)(\boldsymbol{B}\boldsymbol{D}) & (\boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{c}_2)(\boldsymbol{B}\boldsymbol{D}) & \cdots & (\boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{c}_n)(\boldsymbol{B}\boldsymbol{D})\end{bmatrix}\\ &=(\boldsymbol{A}\boldsymbol{C})\otimes(\boldsymbol{B}\boldsymbol{D}).\end{aligned}$$

在倒数第二个等号中，我们将 $(\boldsymbol{a}_i^{\mathrm{T}}\otimes\boldsymbol{B})(\boldsymbol{c}_j\otimes\boldsymbol{D})$ 进行了展开，化简得到了 $(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{c}_j)(\boldsymbol{B}\boldsymbol{D})$，而在最后一个等号中，我们将上面的矩阵重新组合成了 Kronecker 乘积的形式。 □

（4）转置：$(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\otimes\boldsymbol{B}^{\mathrm{T}}$.

证明.

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{bmatrix}a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B}\\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B}\\ \vdots & \vdots & & \vdots\\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B}\end{bmatrix}$$

于是，利用分块矩阵的转置，我们可以得到

$$
\begin{aligned}
(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{T}} &= \begin{bmatrix} (a_{11}\boldsymbol{B})^{\mathrm{T}} & (a_{21}\boldsymbol{B})^{\mathrm{T}} & \cdots & (a_{m1}\boldsymbol{B})^{\mathrm{T}} \\ (a_{12}\boldsymbol{B})^{\mathrm{T}} & (a_{22}\boldsymbol{B})^{\mathrm{T}} & \cdots & (a_{m2}\boldsymbol{B})^{\mathrm{T}} \\ \vdots & \vdots & & \vdots \\ (a_{1n}\boldsymbol{B})^{\mathrm{T}} & (a_{2n}\boldsymbol{B})^{\mathrm{T}} & \cdots & (a_{mn}\boldsymbol{B})^{\mathrm{T}} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\boldsymbol{B}^{\mathrm{T}} & a_{21}\boldsymbol{B}^{\mathrm{T}} & \cdots & a_{m1}\boldsymbol{B}^{\mathrm{T}} \\ a_{12}\boldsymbol{B}^{\mathrm{T}} & a_{22}\boldsymbol{B}^{\mathrm{T}} & \cdots & a_{m2}\boldsymbol{B}^{\mathrm{T}} \\ \vdots & \vdots & & \vdots \\ a_{1n}\boldsymbol{B}^{\mathrm{T}} & a_{2n}\boldsymbol{B}^{\mathrm{T}} & \cdots & a_{mn}\boldsymbol{B}^{\mathrm{T}} \end{bmatrix} \\
&= \boldsymbol{A}^{\mathrm{T}}\otimes\boldsymbol{B}^{\mathrm{T}}.
\end{aligned}
$$

对于复矩阵，我们只需将上面的转置换成共轭转置即可，即 $(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\otimes\boldsymbol{B}^{\mathrm{H}}$. □

（5）逆：若 $\boldsymbol{A},\boldsymbol{B}$ 可逆，则 $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$.

证明. 利用 Kronecker 乘积的矩阵乘法性质，我们有

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}) = (\boldsymbol{A}\boldsymbol{A}^{-1})\otimes(\boldsymbol{B}\boldsymbol{B}^{-1}) = \boldsymbol{I}\otimes\boldsymbol{I} = \boldsymbol{I}.$$

因此，$(\boldsymbol{A}\otimes\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$. □

（5）**没有交换律**：一般情况下，$\boldsymbol{A}\otimes\boldsymbol{B}\neq\boldsymbol{B}\otimes\boldsymbol{A}$，但我们可以利用置换矩阵将二者互相转换。

下面我们展示 Kronecker 乘积的一个重要应用：它可以把矩阵方程转化为向量方程。

例 2.8. 设有 $\boldsymbol{A}\in\mathbb{C}^{m\times m},\boldsymbol{B}\in\mathbb{C}^{n\times n},\boldsymbol{C}\in\mathbb{C}^{m\times n}$，定义

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{O} & \boldsymbol{B} \end{bmatrix}.$$

若 $\operatorname{spec}(\boldsymbol{A})\cap\operatorname{spec}(\boldsymbol{B}) = \varnothing$，则 $\boldsymbol{M}$ 可相似对角化，且 $\operatorname{spec}(\boldsymbol{M}) = \operatorname{spec}(\boldsymbol{A})\cup\operatorname{spec}(\boldsymbol{B})$.

证明. 设有初等矩阵 $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{X} \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix}$，根据其初等变换的意义，我们可以轻松地得到它的逆阵

$$\boldsymbol{P}^{-1} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{X} \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix}.$$

这是因为 $\boldsymbol{P}$ 的操作是将下方的行左乘上 $\boldsymbol{X}$ 加到上方，而其逆变换 $\boldsymbol{P}^{-1}$ 的操作自然就是将下方的行左乘 $-\boldsymbol{X}$ 加到上方。

现在，我们来计算 $\boldsymbol{P}^{-1}\boldsymbol{M}\boldsymbol{P}$：

$$\boldsymbol{P}^{-1}\boldsymbol{M}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{C} \\ \boldsymbol{O} & \boldsymbol{B} \end{bmatrix}.$$

如果它能够相似对角化，则需要满足存在 $\boldsymbol{X} \in \mathbb{C}^{m\times n}$，使得

$$\boldsymbol{A}\boldsymbol{X} - \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}.$$

这是一个矩阵方程，叫做 Sylvester 方程。先给出结论：我们可以利用 Kronecker 乘积将其转化为向量方程。这样，我们就可以利用求解线性方程组的方法来求解 $\boldsymbol{X}$.

我们先一列一列地看这个矩阵方程。设 $\boldsymbol{C} = [\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n], c_i \in \mathbb{C}^m, \forall i \in \{1, ..., m\}$.

对左边，我们同样要提取出列。先看第一项 $\boldsymbol{A}\boldsymbol{X}$，如果我们设 $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$，那么它的第 i 列就是 $\boldsymbol{A}\boldsymbol{x}_i$. 对于第二项 $\boldsymbol{X}\boldsymbol{B}$，如果我们设 $\boldsymbol{B} = [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n]$，那么它的第 i 列就是 $\boldsymbol{X}\boldsymbol{b}_i$；但我们并不喜欢这个形式，我们希望将 $\boldsymbol{X}$ 的各个列向量分离开来，这样才方便我们稍后将这个矩阵方程转化为向量方程——我们可以将 $\boldsymbol{X}\boldsymbol{b}_i$ 视为 $\boldsymbol{X}$ 的各个列向量的线性组合，即

$$\boldsymbol{X}\boldsymbol{b}_i = \sum_{j=1}^{n} \boldsymbol{x}_j b_{ji}.$$

于是，我们可以将矩阵方程的第 i 列写成

$$\boldsymbol{A}\boldsymbol{x}_i - \sum_{j=1}^{n} \boldsymbol{x}_j b_{ji} = \boldsymbol{c}_i.$$

当然，这个形式是极不美观的，我们希望将其写成一个线性方程组 $\boldsymbol{P}\boldsymbol{d} = \boldsymbol{f}$ 的形式。那么这里的未知数项自然是由 $\boldsymbol{X}$ 的各列拼在一起构成的，常数项自然是由 $\boldsymbol{C}$ 的各列拼在一起构成的。为方便起见，我们定义

$$\operatorname{vec}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \in \mathbb{C}^{mn}, \quad \operatorname{vec}(\boldsymbol{C}) = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_n \end{bmatrix} \in \mathbb{C}^{mn}.$$

来表示 $\boldsymbol{X}$ 和 $\boldsymbol{C}$ 各列的拼接。现在，我们的任务就是要构造出系数矩阵，使得向量方程与原来的矩阵方程等价。

首先，我们注意到 $\operatorname{vec}(\boldsymbol{X}) \in \mathbb{C}^{mn}$，因此系数矩阵一定要是 $\mathbb{C}^{mn\times mn}$ 阶矩阵。但是我们的 $\boldsymbol{A}$ 是 m 阶矩阵，$\boldsymbol{B}$ 是 n 阶矩阵，因此我们需要对它们进行扩展。一个自然的想法就是利用 Kronecker 乘积来扩展它们，即使用 $\star \otimes \star$ 来扩展它们。

先看 $\boldsymbol{A}$，我们需要的分量方程组是 $\boldsymbol{A}\boldsymbol{x}_i$. 如果我们拿 $\star\otimes\star$ 乘上 $\operatorname{vec}(\boldsymbol{X})$，从分块矩阵的角度看，结果一定需要是

$$\begin{bmatrix} \boldsymbol{A}\boldsymbol{x}_1 \\ \boldsymbol{A}\boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{A}\boldsymbol{x}_n \end{bmatrix}.$$

试想，如果 A 是一个常数而非矩阵，x_i 是一个数而非向量，那么我们可以直接将 A 这个常数提取出来；但如果我们想要将系数写成一个矩阵呢？我们就可以将其写作 $A\boldsymbol{I}$，将其展开，就有

$$\begin{bmatrix} A & & & \\ & A & & \\ & & \ddots & \\ & & & A \end{bmatrix}.$$

现在，$\boldsymbol{A}$ 是一个矩阵，$\boldsymbol{x}_i$ 是一个向量，我们很自然地能够猜想：我们的系数矩阵会不会也长上面这样呢？答案是肯定的，这就是 Kronecker 乘积 $\boldsymbol{I}_n\otimes\boldsymbol{A}$ 的形式。现在，我们来验证一下这一形式是否符合我们的需求：

$$(\boldsymbol{I}_n\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{A} & & & \\ & \boldsymbol{A} & & \\ & & \ddots & \\ & & & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{x}_1 \\ \boldsymbol{A}\boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{A}\boldsymbol{x}_n \end{bmatrix}.$$

显然，我们的目标达成了。

接下来，我们来处理 $\boldsymbol{B}$ 这一项。我们需要的分量方程组是 $\sum_{j=1}^n \boldsymbol{x}_j b_{ji}$. 仿照上面的方法，如果我们把 x_i 看成一个数而非一个向量，那么上面这个式子就是内积的表达式，即 $\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{x}$. 现在我们把 i 去掉，写成一个完整的矩阵形式，我们就有

$$\begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}} \\ \boldsymbol{b}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{b}_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} \sum_{j=1}^n b_{j1}x_j \\ \sum_{j=1}^n b_{j2}x_j \\ \vdots \\ \sum_{j=1}^n b_{jn}x_j \end{bmatrix}$$

现在，我们把 $\boldsymbol{x}_i$ 看成一个向量而非一个数，那么我们就需要对系数矩阵进行修正，但系数矩阵的形式肯定还是类似的。我们先把 $\boldsymbol{B}^{\mathrm{T}}$ 完整展开，观察一下它的结构：

$$\boldsymbol{B}^{\mathrm{T}} = \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{n1} \\ b_{12} & b_{22} & \cdots & b_{n2} \\ \vdots & \vdots & & \vdots \\ b_{1n} & b_{2n} & \cdots & b_{nn} \end{bmatrix}.$$

如果我们单看第一行，即 $\boldsymbol{b}_1^{\mathrm{T}} = [b_{11}, \cdots, b_{n1}]$，我们最终需要的结果是

$$b_{11}\boldsymbol{x}_1 + b_{21}\boldsymbol{x}_2 + \cdots + b_{n1}\boldsymbol{x}_n.$$

在这里，我们需要每个 b_{j1} 都去乘上 $\boldsymbol{x}_j$，因此我们可以将 b_{j1} 扩展成 $b_{j1}\boldsymbol{I}_m$ 的形式，将系数矩阵的第一行重新写为

$$[b_{11}\boldsymbol{I}_m, b_{21}\boldsymbol{I}_m, \cdots, b_{n1}\boldsymbol{I}_m].$$

这样，我们拿新的系数矩阵的第一行去乘 $\mathrm{vec}(\boldsymbol{X})$，就可以得到我们想要的结果了。同理，我们可以将系数矩阵的第 i 行写成

$$[b_{1i}\boldsymbol{I}_m, b_{2i}\boldsymbol{I}_m, \cdots, b_{ni}\boldsymbol{I}_m].$$

于是，我们可以将系数矩阵写成

$$\begin{bmatrix} b_{11}\boldsymbol{I}_m & b_{21}\boldsymbol{I}_m & \cdots & b_{n1}\boldsymbol{I}_m \\ b_{12}\boldsymbol{I}_m & b_{22}\boldsymbol{I}_m & \cdots & b_{n2}\boldsymbol{I}_m \\ \vdots & \vdots & & \vdots \\ b_{1n}\boldsymbol{I}_m & b_{2n}\boldsymbol{I}_m & \cdots & b_{nn}\boldsymbol{I}_m \end{bmatrix} = \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m.$$

这样我们就构造出了第二个系数矩阵。

我们将两个系数矩阵相减，就得到了最终的系数矩阵：

$$\boldsymbol{P} = (\boldsymbol{I}_n \otimes \boldsymbol{A}) - (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m).$$

于是，矩阵方程 $\boldsymbol{A}\boldsymbol{X} - \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}$ 就等价于向量方程

$$\big(\boldsymbol{I}_n \otimes \boldsymbol{A} - \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m\big)\,\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}).$$

我们要判断 $\boldsymbol{M}$ 是否可相似对角化，其实就是要判断上面这个向量方程是否有解、有多少解。

直接将其解出肯定是不易的，但我们可以通过分析系数矩阵的秩，来判断其解的个数。由于相似变换不改变矩阵的秩，我们可以考虑先对系数矩阵进行相似变化，得到一个更简单的矩阵，然后再分析这个矩阵的秩。

首先，根据上节课的结论，我们知道 $\boldsymbol{A}$ 和 $\boldsymbol{B}^{\mathrm{T}}$ 都可以相似上三角化，即

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}_1, \quad \boldsymbol{V}^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{T}_2,$$

利用这两个式子，我们可以对系数矩阵做进一步化简：

$$\begin{aligned} \boldsymbol{I}_n \otimes \boldsymbol{A} - \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m &= \boldsymbol{I}_n \otimes (\boldsymbol{U}\boldsymbol{T}_1\boldsymbol{U}^{-1}) - (\boldsymbol{V}\boldsymbol{T}_2\boldsymbol{V}^{-1}) \otimes \boldsymbol{I}_m \\ &= (\boldsymbol{V}\boldsymbol{I}_n\boldsymbol{V}^{-1}) \otimes (\boldsymbol{U}\boldsymbol{T}_1\boldsymbol{U}^{-1}) - (\boldsymbol{V}\boldsymbol{T}_2\boldsymbol{V}^{-1}) \otimes (\boldsymbol{U}\boldsymbol{I}_m\boldsymbol{U}^{-1}) \\ &= (\boldsymbol{V} \otimes \boldsymbol{U})(\boldsymbol{I}_n \otimes \boldsymbol{T}_1 - \boldsymbol{T}_2 \otimes \boldsymbol{I}_m)(\boldsymbol{V}^{-1} \otimes \boldsymbol{U}^{-1}) \\ &= (\boldsymbol{V} \otimes \boldsymbol{U})(\boldsymbol{I}_n \otimes \boldsymbol{T}_1 - \boldsymbol{T}_2 \otimes \boldsymbol{I}_m)(\boldsymbol{V} \otimes \boldsymbol{U})^{-1}. \end{aligned}$$

在第一个等号中，我们利用了 Schur 分解 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{T}_1\boldsymbol{U}^{-1},\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{V}\boldsymbol{T}_2\boldsymbol{V}^{-1}$；在第二个等号中，我们将 $\boldsymbol{I}$ 重新写成三个矩阵的乘积，意在使用性质 Kronecker 乘积的积等于积的 Kronecker 乘积；在第三个等号中，我们正是运用了这个性质；在第四个等号中，我们利用了逆的 Kronecker 乘积等于 Kronecker 乘积的逆这个性质。

在上面的操作中，我们相当于将系数矩阵相似化了，而相似变换并不改变矩阵的秩，因此我们只需要分析中间的矩阵 $\boldsymbol{I}_n\otimes\boldsymbol{T}_1-\boldsymbol{T}_2\otimes\boldsymbol{I}_m$ 的秩就可以了。

现在，我们来分析 $\boldsymbol{I}_n\otimes\boldsymbol{T}_1-\boldsymbol{T}_2\otimes\boldsymbol{I}_m$ 的结构。注意到 $\boldsymbol{T}_1,\boldsymbol{T}_2$ 都是上三角阵，因此 $\boldsymbol{I}_n\otimes\boldsymbol{T}_1,\boldsymbol{T}_2\otimes I_m$ 都是上三角阵，其差自然也是上三角阵。对于上三角阵，其秩要等与其非零行的个数，非零行的个数也就是非零特征值的个数。所以，我们只需要关心这个矩阵有多少个非零特征值即可。由于 $\operatorname{spec}(\boldsymbol{A})\cap\operatorname{spec}(\boldsymbol{B})=\varnothing$，$\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$ 没有相同的特征值，因此矩阵 $\boldsymbol{I}_n\otimes\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2\otimes\boldsymbol{I}_m$ 的特征值没有重复，因此矩阵 $\boldsymbol{I}_n\otimes\boldsymbol{T}_1-T_2\otimes\boldsymbol{I}_m$ 没有零特征值，从而满秩，因此向量方程 $(\boldsymbol{I}_n\otimes\boldsymbol{A}-\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{I}_m)\operatorname{vec}(\boldsymbol{X})=\operatorname{vec}(\boldsymbol{C})$ 有唯一解，亦即矩阵方程 $\boldsymbol{A}\boldsymbol{X}-\boldsymbol{X}\boldsymbol{B}=\boldsymbol{C}$ 有唯一解，亦即矩阵 $\boldsymbol{M}$ 可相似对角化，且方式唯一。

由于 $\boldsymbol{M}$ 的相似矩阵的对角线块就是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，因此 $\operatorname{spec}(\boldsymbol{M})=\operatorname{spec}(\boldsymbol{A})\cup\operatorname{spec}(\boldsymbol{B})$.

□

现在，我们来总结一下这道例题。我们通过引入 Kronecker 乘积，将矩阵方程转化为了向量方程。这样，我们就可以利用线性方程组的理论来分析矩阵方程的解的情况。然后，我们通过对系数矩阵进行相似变换，将其化简为一个更容易分析的矩阵。最后，我们通过分析这个矩阵的特征值，得出了矩阵方程解的情况，从而证明了 $\boldsymbol{M}$ 可相似对角化。

其中，最困难的一步在于如何将矩阵方程转化为向量方程。下面我们给出一个结论，它能够帮我们更快地将矩阵方程转化为向量方程。

定理 2.9. *向量化操作符具有如下性质：设有* $\boldsymbol{A}\in\mathbb{C}^{m\times m},\boldsymbol{B}\in\mathbb{C}^{n\times n},\boldsymbol{X}\in\mathbb{C}^{m\times n}$，*则有*

$$\operatorname{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})=(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{X}).$$

证明. 仿照上面例题中的方法。我们先分析 $\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$ 每一列的行为，然后再考虑向量化之后会出现什么形式。设 $\boldsymbol{B}=[\boldsymbol{b}_1,\boldsymbol{b}_2,\cdots,\boldsymbol{b}_n]$，则 $\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$ 的第 i 列为

$$\boldsymbol{A}\boldsymbol{X}\boldsymbol{b}_i=\sum_{j=1}^{n}\boldsymbol{A}\boldsymbol{X}b_{ji}.$$

我们希望将其表示为 $\boldsymbol{X}$ 的各个列向量的线性组合，因此我们设 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$，则有

$$\boldsymbol{A}\boldsymbol{X}\boldsymbol{b}_i=\sum_{j=1}^{n}b_{ji}\boldsymbol{A}\boldsymbol{x}_j.$$

取 $i=1$，上式变为

$$\boldsymbol{A}\boldsymbol{X}\boldsymbol{b}_1 = b_{11}\boldsymbol{A}\boldsymbol{x}_1 + b_{21}\boldsymbol{A}\boldsymbol{x}_2 + \cdots + b_{n1}\boldsymbol{A}\boldsymbol{x}_n.$$

按照分块矩阵的观点，我们可以设新矩阵的第一行为

$$[b_{11}\boldsymbol{A}, b_{21}\boldsymbol{A}, \cdots, b_{n1}\boldsymbol{A}].$$

那么，在乘上 $\mathrm{vec}(\boldsymbol{X})$ 之后我们就可以得到 $\boldsymbol{A}\boldsymbol{X}\boldsymbol{b}_1$. 同理，我们可以将新矩阵的第 i 行写成

$$[b_{1i}\boldsymbol{A}, b_{2i}\boldsymbol{A}, \cdots, b_{ni}\boldsymbol{A}].$$

于是，我们可以将新矩阵写成

$$\begin{bmatrix} b_{11}\boldsymbol{A} & b_{21}\boldsymbol{A} & \cdots & b_{n1}\boldsymbol{A} \\ b_{12}\boldsymbol{A} & b_{22}\boldsymbol{A} & \cdots & b_{n2}\boldsymbol{A} \\ \vdots & \vdots & & \vdots \\ b_{1n}\boldsymbol{A} & b_{2n}\boldsymbol{A} & \cdots & b_{nn}\boldsymbol{A} \end{bmatrix} = \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A}.$$

这样，我们就得到了

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X}).$$

□

运用这个结论，我们可以把矩阵 $\boldsymbol{A}\boldsymbol{X}$ 写成 $\boldsymbol{A}\boldsymbol{X}\boldsymbol{I}_n$，在向量化后就是 $(\boldsymbol{I}_n \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{X})$；而对矩阵 $\boldsymbol{X}\boldsymbol{B}$，我们可以将其重新写为 $\boldsymbol{I}_m\boldsymbol{X}\boldsymbol{B}$，在向量化之后就是 $(\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m)\mathrm{vec}(\boldsymbol{X})$. 因此，我们可以直接写出矩阵方程 $\boldsymbol{A}\boldsymbol{X} - \boldsymbol{X}\boldsymbol{B} = \boldsymbol{C}$ 对应的向量方程为

$$\big((\boldsymbol{I}_n \otimes \boldsymbol{A}) - (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{I}_m)\big)\,\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{C}).$$

下面我们给出 Kronecker 乘积的另一个重要应用：它可以用于证明代数数可以构成一个数域。

首先，我们来回顾一下代数数的定义。对于 $\alpha \in \mathbb{C}$，如果存在一个首系数为 1，其余各项系数均为有理数的多项式 $f(x)$，使得 $f(\alpha)=0$，则称 α 为代数数。

例 2.10. 证明：代数数构成一个数域。

证明. 要证明这个结论，我们只需要证明代数数对加法、乘法和除法封闭即可。

我们只需要证明对任意代数数 α, β，$\alpha+\beta, -\beta, \alpha\times\beta, \frac{1}{\beta}$ 为代数数即可。

回顾上节课的内容，若 α, β 分别是多项式 $f(x), g(x)$ 的根，则 α, β 分别是 $f(x), g(x)$ 对应的 Frobenius 友阵的特征值。

我们设 α 是 $\boldsymbol{A} \in \mathbb{Q}^{m\times m}$ 的特征值，β 是 $\boldsymbol{B} \in \mathbb{Q}^{n\times n}$ 的特征值。即有

$$\begin{cases} \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}\alpha \\ \boldsymbol{B}\boldsymbol{y} = \boldsymbol{y}\beta \end{cases}.$$

注意到：$(\boldsymbol{I}_n \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{I}_m) \in \mathbb{Q}^{mn\times mn}$ 的特征值是 $\alpha+\beta$，对应的特征向量为 $\boldsymbol{y}\otimes\boldsymbol{x}$； 我并不知道是怎么注意到的，但我们可以对这个结论进行验证：

$$\begin{aligned}(\boldsymbol{I}_n \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{I}_m)(\boldsymbol{y}\otimes\boldsymbol{x}) &= (\boldsymbol{I}_n \otimes \boldsymbol{A})(\boldsymbol{y}\otimes\boldsymbol{x}) + (\boldsymbol{B} \otimes \boldsymbol{I}_m)(\boldsymbol{y}\otimes\boldsymbol{x}) \\ &= (\boldsymbol{I}_n\boldsymbol{y}) \otimes (\boldsymbol{A}\boldsymbol{x}) + (\boldsymbol{B}\boldsymbol{y}) \otimes (\boldsymbol{I}_m\boldsymbol{x}) \\ &= \boldsymbol{y} \otimes (\alpha\boldsymbol{x}) + (\beta\boldsymbol{y}) \otimes \boldsymbol{x} = (\alpha+\beta)(\boldsymbol{y}\otimes\boldsymbol{x}).\end{aligned}$$

注意到：$(-\boldsymbol{B})\boldsymbol{y} = \boldsymbol{y}(-\beta)$，因此 $-\beta$ 是 $-\boldsymbol{B}$ 的特征值，$-\beta$ 是代数数。 这个还是可以注意到的。

注意到：$\boldsymbol{A}\otimes\boldsymbol{B} \in \mathbb{Q}^{mn\times mn}$ 的特征值是 $\alpha\beta$，特征向量是 $\boldsymbol{x}\otimes\boldsymbol{y}$； 我并不知道是怎么注意到的，但我们可以对这个结论进行验证：

$$\begin{aligned}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{x}\otimes\boldsymbol{y}) &= (\boldsymbol{A}\boldsymbol{x}) \otimes (\boldsymbol{B}\boldsymbol{y}) \\ &= (\alpha\boldsymbol{x}) \otimes (\beta\boldsymbol{y}) \\ &= (\alpha\beta)(\boldsymbol{x}\otimes\boldsymbol{y}).\end{aligned}$$

注意到：β 是有理系数多项式 $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$ 的根，即有

$$\beta^n + a_{n-1}\beta^{n-1} + \cdots + a_0 = 0.$$

如果我们两边同时乘上 β^{-n}，则有

$$1 + a_{n-1}\beta^{-1} + a_{n-2}\beta^{-2} + \cdots + a_0\beta^{-n} = 0.$$

再将两边同时乘上 $(-a_0^{-1})$，则有

$$(\beta^{-1})^n + (-a_0^{-1}a_{n-1})(\beta^{-1})^{n-1} + \cdots + (-a_0^{-1}a_1)(\beta^{-1}) + (-a_0^{-1}) = 0.$$

因此 β^{-1} 也是代数数。 □

3 向量范数

在数值计算中，我们经常需要衡量向量的大小，这就需要用到向量范数的概念。

定义 3.1. 设 $\boldsymbol{x} \in \mathbb{C}^n$，如果一个映射 $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ 满足以下条件，则称其为一个向量范数：

- 非负性：$\|\boldsymbol{x}\| \geq 0$，当且仅当 $\boldsymbol{x} = \boldsymbol{0}$ 时等号成立；
- 齐次性：$\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|, \forall \alpha \in \mathbb{C}$；
- 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$.

例 3.2. 常见的向量范数有以下几种：

- ∞-范数：$\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$；
- *1*-范数：$\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$；
- *2*-范数：$\|\boldsymbol{x}\|_2 = (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}}$.

我们最常用的就是 2-范数，它也就是欧几里得距离。1-范数和 ∞-范数在某些情况下也会用到。在物流配送中，假设一个城市有多个配送点，每个配送点需要配送的货物量可视为向量的元素。1 -范数可以用来计算总的货物配送量，方便物流公司提前规划车辆的运载能力、安排配送资源等。比如，某物流公司负责向 5 个不同的超市送货，每个超市的订货量分别为 $x_1, ..., x_5$，则

$$\|\boldsymbol{x}\|_1 = |x_1| + |x_2| + |x_3| + |x_4| + |x_5|$$

就表示总的配送量。在图像识别中，当比较两幅图像的差异时，可以将图像的像素值差异构成一个向量。无穷范数表示两幅图像在所有像素点上差异的最大值。比如，在检测图像的篡改时，计算原始图像和疑似篡改图像对应像素点的差值向量，通过分析其无穷范数，能快速判断是否存在某个像素区域的差异过大，进而确定图像是否被篡改以及篡改的严重程度。

将上面三种范数进行推广，我们可以得到 p-范数的概念。

定义 3.3. 设 $\boldsymbol{x} \in \mathbb{C}^n$，$p \geq 1$，则 p-范数定义为

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}.$$

请注意，当 $p < 1$ 时，我们可以验证上面这个函数并不满足三条范数公理，因此不是范数。

现在我们总结一下向量范数的性质：

（1）（等价性）任何两种范数等价，即对任意两种范数 f_1, f_2，存在常数 $m, M > 0$，使得

$$m f_1(\boldsymbol{x}) \leq f_2(\boldsymbol{x}) \leq M f_1(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{C}^n.$$

证明. 我们只需要证明任何范数 f 都与 1-范数等价即可。设 $\mathscr{D} = \{\boldsymbol{x} : \|\boldsymbol{x}\|_1 = 1\}$，则有 $\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_1} \in \mathscr{D}$. 由于 $\mathscr{D}$ 是一个闭集，因此 $f(\boldsymbol{x})$ 在 $\mathscr{D}$ 上有最小值 m 和最大值 M. 因此，我们有

$$m \leq f\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_1}\right) = \frac{f(\boldsymbol{x})}{\|\boldsymbol{x}\|_1} \leq M, \forall \boldsymbol{x} \in \mathbb{C}^n.$$

于是，我们有

$$m\|\boldsymbol{x}\|_1 \leq f(\boldsymbol{x}) \leq M\|\boldsymbol{x}\|_1, \forall \boldsymbol{x} \in \mathbb{C}^n.$$

因此任何范数都与 1-范数等价，亦即任何两种范数等价。 □

（2）（收敛性）依范数收敛等价于依坐标收敛。

证明. 在本部分证明中，下标表示第几个向量，上标表示向量的第几个元素。

依坐标收敛 $\Rightarrow$ 依范数收敛：设有 $\boldsymbol{x}_k \to \boldsymbol{x}$，则对任意 $\varepsilon > 0$，存在 $N_\varepsilon \in \mathbb{N}$，有

$$|x_k^i - x^i| < \varepsilon, \forall i \in \{1, \ldots, n\}, k > N_\varepsilon.$$

于是，$\boldsymbol{x}_k - \boldsymbol{x}$ 的 1-范数为

$$\|\boldsymbol{x}_k - \boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_k^i - x^i| < n\varepsilon.$$

因此，$\boldsymbol{x}_k \to \boldsymbol{x}$ 依 1-范数收敛，亦即依任何范数收敛。

依范数收敛 $\Rightarrow$ 依坐标收敛：设有 $\boldsymbol{x}_k \to \boldsymbol{x}$ 依范数收敛，则对任意 $\varepsilon > 0$，存在 $N_\varepsilon \in \mathbb{N}$，有

$$\|\boldsymbol{x}_k - \boldsymbol{x}\|_\infty < \varepsilon, \forall k > N_\varepsilon.$$

亦即

$$\max_{1 \leq i \leq n} |x_k^i - x^i| < \varepsilon, \forall k > N_\varepsilon.$$

因此，$|x_k^i - x^i| < \varepsilon, \forall i \in \{1, \ldots, n\}, k > N_\varepsilon$. 亦即 $\boldsymbol{x}_k \to \boldsymbol{x}$ 依坐标收敛。 □

（3）（连续性）范数是关于坐标的连续函数。

证明. 设有 $\boldsymbol{x}_k \to \boldsymbol{x}$，则对任意 $\varepsilon > 0$，存在 $N_\varepsilon \in \mathbb{N}$，有

$$\|\boldsymbol{x}_k - \boldsymbol{x}\| < \varepsilon, \forall k > N_\varepsilon.$$

我们先来推导一个不等式：

$$\|\boldsymbol{x} - \boldsymbol{y}\| \geq |\|\boldsymbol{x}\| - \|\boldsymbol{y}\||.$$

由三角不等式，我们有

$$\begin{aligned}
\|\boldsymbol{x}\| = \|(\boldsymbol{x} - \boldsymbol{y}) + \boldsymbol{y}\| &\leq \|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{y}\| \\
\|\boldsymbol{y}\| = \|(\boldsymbol{y} - \boldsymbol{x}) + \boldsymbol{x}\| &\leq \|\boldsymbol{y} - \boldsymbol{x}\| + \|\boldsymbol{x}\| = \|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{x}\|
\end{aligned}$$

移项即得上面的不等式。

于是，我们有

$$\left|\|\boldsymbol{x}_k\| - \|\boldsymbol{x}\|\right| \leq \|\boldsymbol{x}_k - \boldsymbol{x}\| < \varepsilon, \forall k > N_\varepsilon.$$

因此，$\|\boldsymbol{x}_k\| \to \|\boldsymbol{x}\|$. $\square$