线性代数笔记

清华大学《线性代数》课程笔记

艾颖华

目录

第一章　线性映射与矩阵

1.1　线性代数介绍

线性代数是研究线性空间与线性映射的学问. 从如下两个方面, 可以看出线性代数在数学与科学中的重要性.

(1) 线性化原理:(可微) 函数的改变在小范围内是线性的. 具体的说, 对每个 x, 当 h 很小时, f 的改变 $\Delta f = f(x+h) - f(x)$ 可以近似为 h 的线性映射. 推而广之, 对于任何一个系统, 它在小范围内的改变, 也可以近似为线性的. 这样, 为了研究一般的复杂系统, 需要先研究相对简单的线性系统.

(2) 量子物理中的态矢量叠加原理: 对于一个量子体系, 如果 Ψ_1, Ψ_2 是该体系可能处在的两个态, 则对任何复数 c_1, c_2, $c_1\Psi_1 + c_2\Psi_2$ 也是该系统可能处在的一个态. 这样, 同一个量子体系可能处在的态的全体构成一个 (复数域上的) 线性空间. 换句话说, 需要用线性空间的语言去描述量子物理. 不但如此, 假设有两个观测者 O 与 O' 来观测此量子体系, 如果观测者 O 观测到系统处于态 Ψ, 则观测者 O' 观测到系统处于某个确定的态 $T(\Psi)$, 量子物理中可以证明, 前述变换 $\Psi \to T(\Psi)$ 是一个线性映射. 这表明, 量子物理的基础层面需要用线性映射进行描述.

1.2　映射

定义 1.2.1 (映射)**.** 给定集合 X, Y, 从 X 到 Y 的一个映射 $f : X \to Y$ 是指: 对每个 $x \in X$, 都指定唯一确定的 $y \in Y$ 与之对应. 把 Y 中与 x 对应的元素记作 $f(x)$, 称为 f 在 x 处的值 (或像).

称 X 为 f 的定义域 (*domain*), 称 Y 为 f 的 *co-domain*.

例 1.2.2 (恒同映射)**.** 给定集合 X, 定义 X 的恒同映射 $id_X : X \to X$ 为

$$id_X(x) = x, \quad \forall x \in X.$$

定义 1.2.3 (映射的复合)**.** 给定映射 $f: X \to Y, g: Y \to Z$, 定义它们的复合映射 $g \circ f: X \to Z$ 为

$$g \circ f(x) = g(f(x)), \quad \forall x \in X.$$

例 1.2.4 (恒同映射是复合的单位)**.** 设 $f: X \to Y$, 则 $f \circ id_X = f = id_Y \circ f$.

命题 1.2.5 (映射的复合满足结合律)**.** 设 $f: X \to Y, g: Y \to Z, h: Z \to W$, 则

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

定义 1.2.6. (1) 设 $f: X \to Y$ 是映射. 令

$$f(X) = \{y \in Y | \exists x \in X, \text{使得} f(x) = y\},$$

称为 f 的像集 (*image*).

(2) 对于 Y 的子集 V, 令

$$f^{-1}(V) = \{x \in X | f(x) \in V\},$$

称为 V(在 f 下) 的原像集 (*pre-image*).

定义 1.2.7. 设 $f: X \to Y$ 是映射.

(1) 称 f 为单射, 如果 $\forall x \neq x'$, 有 $f(x) \neq f(x')$.

(2) 称 f 是满射, 如果 $f(X) = Y$.

(3) 称 f 为双射 (或一一对应, 可逆映射), 如果 f 既单又满.

命题 1.2.8. f 是双射的充分必要条件是存在映射 $g: Y \to X$ 使得 $g \circ f = id_X, f \circ g = id_Y$.

如果上述 g 存在的话, 则是唯一的, 称之为 f 的逆映射, 记作 $f^{-1}: Y \to X$.

命题 1.2.9. 设 $f: X \to Y$ 是映射.

(1) 如果 X 是非空集合, 则 f 是单射当且仅当存在映射 $g: Y \to X$, 使得 $g \circ f = id_X$.

(2) f 是满射当且仅当存在映射 $g: Y \to X$, 使得 $f \circ g = id_Y$.

以后, 我们会证明在线性空间与线性映射的范畴中, 命题1.2.9的相应版本也是成立的, 后者是个有用的结论.

1.3 $\mathbf{R}^n$ 上的线性结构

定义 1.3.1. 一个 n 元有序数组 $\mathbf{a}=(a_1,a_2,\cdots,a_n)$ 称为一个 n 维向量, 称 a_i 为向量 $\mathbf{a}$ 的第 i 个分量或坐标.

令 $\mathbf{R}^n$ 为所有的 n 维向量所构成的集合, 其上有如下两种运算:

(1) 向量的加法: $(a_1,a_2,\cdots,a_n)+(b_1,b_2,\cdots,b_n)=(a_1+b_1,a_2+b_2,\cdots,a_n+b_n)$.

(2) 实数与向量的数乘: $k(a_1,a_2,\cdots,a_n)=(ka_1,ka_2,\cdots,ka_n)$.

人们把这两种运算统称为向量的线性运算. 称带有上述线性运算的集合 $\mathbf{R}^n$ 为线性空间 $\mathbf{R}^n$.

命题 1.3.2. *向量的加法与数乘运算满足如下运算法则:*

(1) *加法是结合的:* $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$*;*

(2) *加法是交换的:* $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$*;*

(3) *加法有零元, 即存在零向量* $\mathbf{0}=(0,0,\cdots,0)$ *满足* $\mathbf{a}+\mathbf{0}=\mathbf{a}=\mathbf{0}+\mathbf{a}$*;*

(4) *加法有逆, 对每个向量* $\mathbf{a}=(a_1,a_2,\cdots,a_n)$*, 令* $-\mathbf{a}=(-a_1,-a_2,\cdots,-a_n)$*, 它满足* $\mathbf{a}+(-\mathbf{a})=\mathbf{0}$*;*

(5) *数乘作用有单位,* $1\mathbf{a}=\mathbf{a}$*;*

(6) *数乘作用是结合的,* $k(l\mathbf{a})=(kl)\mathbf{a}$*;*

(7) *数乘作用保持向量加法,* $k(\mathbf{a}+\mathbf{b})=k\mathbf{a}+k\mathbf{b}$*;*

(8) *数乘关于数的分配律:* $(k+l)\mathbf{a}=k\mathbf{a}+l\mathbf{a}$*.*

以后, 人们利用这八条运算法则去定义一般的线性空间: 如果一个集合 V 上有加法运算, 并有实数与 V 中成员的数乘运算, 且它们满足前述八条运算法则, 则称 V 是一个 (实数域上的) 线性空间.

定义 1.3.3. 保持线性运算的映射称为线性映射. 具体的说, 称映射 $f:\mathbf{R}^n\to\mathbf{R}^m$ 是一个线性映射, 如果对任何 $\mathbf{a},\mathbf{b}\in\mathbf{R}^n$ 以及任何 $k\in\mathbf{R}$ 都有

$$f(\mathbf{a}+\mathbf{b})=f(\mathbf{a})+f(\mathbf{b}),\quad f(k\mathbf{a})=kf(\mathbf{a}).$$

例 1.3.4. (1) 恒同映射 $id:\mathbf{R}^n\to\mathbf{R}^n$ 是线性映射.

(2) 平面上的旋转变换. 考虑将平面上的点绕原点逆时针旋转 θ 角所得的变换 R_θ,

$$R_\theta(x_1,x_2)=(x_1\cos\theta-x_2\sin\theta,x_1\sin\theta+x_2\cos\theta),$$

它是线性映射.

(3) 平面上关于直线 $l_\theta:x_2-x_1\tan\theta=0$ 的反射变换 H_θ,

$$H_\theta(x_1,x_2)=(x_1\cos2\theta+x_2\sin2\theta,x_1\sin2\theta-x_2\cos2\theta),$$

(4) 投影映射. 设 $m<n$, 定义 $P:\mathbf{R}^n\to\mathbf{R}^m$ 为

$$P(x_1,x_2,\cdots,x_n)=(x_1,x_2,\cdots,x_m),$$

它是把 $\mathbf{R}^n$ 向由前 m 个坐标轴所张成的子空间 $\mathbf{R}^m$ 做投影.

(5) 人们把线性映射 $f:\mathbf{R}^n\to\mathbf{R}$ 称为 n 元线性函数, 它形如

$$f(x_1,\cdots,x_n)=a_1x_1+\cdots+a_nx_n.$$

命题 1.3.5. 线性映射的复合是线性映射. 设 $f:\mathbf{R}^n\to\mathbf{R}^m$ 与 $g:\mathbf{R}^m\to\mathbf{R}^l$ 是线性映射, 则它们的复合映射 $g\circ f:\mathbf{R}^n\to\mathbf{R}^l$ 也是线性映射.

1.4 线性映射的表示矩阵

考虑 $\mathbf{R}^n$ 的标准坐标向量组 (标准基底)

$$\mathbf{e}_1=(1,0,\cdots,0),\quad \mathbf{e}_2=(0,1,\cdots,0),\quad \cdots,\quad \mathbf{e}_n=(0,0,\cdots,1),$$

则 $\mathbf{R}^n$ 中的每个向量 $\mathbf{x}$ 都可以唯一的表示成

$$\mathbf{x}=k_1\mathbf{e}_1+\cdots+k_n\mathbf{e}_n,\quad k_1,\cdots,k_n\in\mathbf{R},$$

事实上, 当且仅当每个 k_i 等于 $\mathbf{x}$ 的第 i 个分量时上式成立.

定义 1.4.1. 设 $\mathbf{a}_1,\cdots,\mathbf{a}_n$ 是 $\mathbf{R}^m$ 中的一组向量. 对任何一组实数 $k_1,\cdots,k_n$, 称向量 $\sum\limits_{i=1}^n k_i\mathbf{a}_i=k_1\mathbf{a}_1+\cdots+k_n\mathbf{a}_n$ 为向量组 $\mathbf{a}_1,\cdots,\mathbf{a}_n$ 的一个线性组合.

设 $\mathbf{b}\in\mathbf{R}^m$, 如果存在 $k_1,\cdots,k_n\in\mathbf{R}$ 使得 $\mathbf{b}=\sum\limits_{i=1}^n k_i\mathbf{a}_i$, 则称 $\mathbf{b}$ 可以被向量组 $\mathbf{a}_1,\cdots,\mathbf{a}_n$ 线性表示.

设 $f:\mathbf{R}^n\to\mathbf{R}^m$ 是线性映射, 对任何 $\mathbf{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n$, 有

$$f(\mathbf{x})=f(\sum_{i=1}^n x_i\mathbf{e}_i)=\sum_{i=1}^n x_if(\mathbf{e}_i),$$

这表明线性映射由它在标准基底上的值唯一确定. 记 $f(\mathbf{e}_i)=\mathbf{a}_i$, 我们可以把前述观察总结成如下命题.

命题 1.4.2. 线性映射 $f:\mathbf{R}^n\to\mathbf{R}^m$ 由它在标准基底上的值 $\{f(\mathbf{e}_i)=\mathbf{a}_i\}_{1\le i\le n}$ 唯一决定, $\{f(\mathbf{e}_i)=\mathbf{a}_i\}_{1\le i\le n}$ 可以取 $\mathbf{R}^m$ 中由 n 个向量构成的任何向量组.

设 $\mathbf{a}_i = (a_{1i}, a_{2i}, \cdots, a_{mi})$, 则可把映射 f 用坐标分量具体表示为:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix},$$

注意到上式右边向量的第 i 个分量为

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = (a_{i1}, a_{i2}, \cdots, a_{in}) \cdot (x_1, x_2, \cdots, x_n),$$

可以视为向量的内积. 对不同的指标 i, 上式有着相同的模式. 基于此观察, 人们引入矩阵的概念.

定义 1.4.3. 所谓一个 $m \times n$ 的矩阵, 是指由 mn 个数按如下形式排列而成的数表

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

其中第 i 行第 j 列交叉处的数为 a_{ij}(采用双下标 ij 标记), 称为矩阵 A 的 (i,j) 矩阵元, 也经常记为 A_{ij}. 如果 $m = n$, 则称 A 为 n 阶方阵.

矩阵 A 中第 i 行所有数依次排列构成向量 $(a_{i1}, a_{i2}, \cdots, a_{in}) \in \mathbf{R}^n$, 称之为 A 的第 i 个行向量; 矩阵 A 中第 j 列所有数依次排列构成向量 $(a_{1j}, a_{2j}, \cdots, a_{mj}) \in \mathbf{R}^m$, 称之为 A 的第 j 个列向量.

定义 1.4.4 (矩阵与列向量的乘法)**.** 设 A 是 $m \times n$ 矩阵, $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ 是 $\mathbf{R}^n$ 中的向量, 将它写成列向量的形式, 定义矩阵 A 的左乘作用为

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix},$$

即乘积向量的第 i 个分量等于 A 的第 i 个行向量与 $\mathbf{x}$ 的内积.

定义 1.4.5. 设 $f : \mathbf{R}^n \to \mathbf{R}^m$ 是线性映射, 以 $f(\mathbf{e}_j) = \mathbf{a}_j$ 为第 j 个列向量构建一个 $m \times n$ 矩阵 A, 称之为 f(在标准基底) 的表示矩阵, 则 f 在 $\mathbf{x} \in \mathbf{R}^n$ 处的值等于将 $\mathbf{x}$ 写成列向量并用矩阵 A 左乘以它所得到的列向量.

例 1.4.6. (1) 恒同映射 $id:\mathbf{R}^n\to\mathbf{R}^n$ 的表示矩阵为 n 阶单位矩阵 I_n.

(2) 平面上的旋转 R_θ 的表示矩阵为

$$R_\theta=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

(3) 平面上关于直线 $l_\theta: x_2-x_1\tan\theta=0$ 的反射变换 H_θ 的表示矩阵为

$$H_\theta=\begin{pmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{pmatrix}.$$

1.5 矩阵的乘法

问题 1.5.1. 设线性映射 $f:\mathbf{R}^n\to\mathbf{R}^m$ 与 $g:\mathbf{R}^m\to\mathbf{R}^l$ 的表示矩阵分别为 F,G. 求复合映射 $g\circ f:\mathbf{R}^n\to\mathbf{R}^l$ 的表示矩阵.

设复合映射 $g\circ f$ 的表示矩阵为 H, 则

$$\begin{aligned}H_{ij}&=g\circ f(\mathbf{e_j})\text{ 的}i\text{分量}\\&=g(F_{1j},\cdots,F_{mj})\text{ 的}i\text{分量}\\&=G\text{ 的第}i\text{个行向量与}(F_{1j},\cdots,F_{mj})\text{的内积}\\&=\sum_{k=1}^{m}G_{ik}F_{kj}.\end{aligned}$$

定义 1.5.2. 设 A 是 $l\times m$ 矩阵, B 是 $m\times n$ 矩阵, 定义 A(左) 乘以 B 所得的矩阵为 $l\times n$ 矩阵 AB, 它的矩阵元为

$$(AB)_{ij}=\sum_{k=1}^{m}A_{ik}B_{kj},$$

即乘积矩阵 AB 的 ij 矩阵元等于 A 的第 i 个行向量与 B 的第 j 个列向量的内积. 注意, 为了定义矩阵乘积 AB, 必须要求 A 的列数等于 B 的行数.

命题 1.5.3. 设线性映射 $f:\mathbf{R}^n\to\mathbf{R}^m$ 与 $g:\mathbf{R}^m\to\mathbf{R}^l$ 的表示矩阵分别为 F,G, 则复合映射 $g\circ f:\mathbf{R}^n\to\mathbf{R}^l$ 的表示矩阵为 GF.

推论 1.5.4. 矩阵乘法满足结合律. 设 A,B,C 分别是 $l\times m, m\times n, m\times s$ 矩阵, 则有 $(AB)C=A(BC)$.

例 1.5.5. (1) 平面上两个旋转的复合是旋转, 易直接验证 $R_\alpha R_\beta=R_{\alpha+\beta}$.

(2) 平面上两个反射的复合是旋转, 易直接验证 $H_\alpha H_\beta=R_{2\alpha-2\beta}$.

例 1.5.6. 向量空间 $\mathbf{R}^n$ 上的仿射变换为形如

$$T(\mathbf{x}) = A\mathbf{x} + \mathbf{a}, \quad \forall \mathbf{x} \in \mathbf{R}^n,$$

它由线性变换与平移变换复合而成, 其中 A 是 $n \times n$ 的方阵, $\mathbf{a}$ 是 $\mathbf{R}^n$ 中的向量. 我们把上述仿射变换记为 $T(A, \mathbf{a})$, 则可知

$$T(A, \mathbf{a}) \circ T(B, \mathbf{b})(\mathbf{x}) = A(B\mathbf{x} + \mathbf{b}) + \mathbf{a} = (AB)\mathbf{x} + A\mathbf{b} + \mathbf{a}, \quad \forall \mathbf{x} \in \mathbf{R}^n,$$

即得到仿射变换的复合法则

$$T(A, \mathbf{a}) \circ T(B, \mathbf{b}) = T(AB, A\mathbf{b} + \mathbf{a}).$$

命题 1.5.7. *设 k 是给定的正整数, $A_1, A_2, \cdots, A_k$ 是 n 个矩阵, 满足对每个 $1 \le i \le k-1$, 矩阵 A_i 的列数等于矩阵 A_{i+1} 的行数, 则乘积矩阵 $A_1 A_2 \cdots A_k$ 的矩阵元为:*

$$(A_1 A_2 \cdots A_k)_{ij} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_{k-1}} \left((A_1)_{ii_1} (A_2)_{i_1 i_2} \cdots (A_k)_{i_{k-1} j} \right),$$

其中 $\sum\limits_{i_t}$ 表示对指标 i_t 求和, i_t 取遍不超过矩阵 A_{t+1} 行数的所有正整数.

1.6 Gauss 消元法

求解方程是数学中最重要的思想之一: 如果要寻找具有某种特定性质的对象, 人们把描述此对象的参数设为未知元, 把对象需要满足的性质用关于这些参数的等式表示出来, 然后再用代数的方法去解出未知元, 这就是方程的思想.

最简单的方程组是线性方程组, 即如下形式的方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2; \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{cases} \tag{1.1}$$

对于此类方程组, 人们关心这样一些问题: 上述方程组是否有解? 解是否唯一? 能否写出解的一般形式?

可用线性映射或矩阵的语言重新叙述此问题. 定义线性映射 $f: \mathbf{R}^n \to \mathbf{R}^m$ 为:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

则求解方程组 (1.1) 等价于求向量 $\mathbf{b}=(b_1,b_2,\cdots,b_m)$ 在 f 下的原像集 $f^{-1}(\{\mathbf{b}\})$, 人们问此原像集合是否非空? 是否是单元集? 其一般形式是怎样的? 把 f 的表示矩阵记为 A, 称之为方程组 (1.1) 的系数矩阵, 由此可把方程组 (1.1) 简记为 $A\mathbf{x}=\mathbf{b}$, 并称 $\mathbf{b}$ 为方程组的常数项.

求解线性方程组的方法是所谓的 Gauss 消元法. 不妨设 $a_{11}\neq 0$, 则把方程组 (1.1) 中的第 $i(2\leq i\leq m)$ 个方程加上第 1 个方程的 $-\frac{a_{i1}}{a_{11}}$ 倍, 得到与原方程组同解的方程组

$$\begin{cases} a_{11}x_1 + & a_{12}x_2+\cdots+a_{1n}x_n=b_1;\\ & a'_{22}x_2+\cdots+a'_{2n}x_n=b'_2;\\ & \cdots\\ & a'_{m2}x_2+\cdots+a'_{mn}x_n=b'_m. \end{cases} \tag{1.2}$$

注意到, 方程组 (1.2) 的第 2 个到第 m 个方程构成的方程组是关于 $n-1$ 个变元 $x_2,\cdots,x_n$ 的, 共由 $m-1$ 个方程构成的方程组, 它比原方程组 (1.1) 简单, 可先解出这个更简单的方程组, 再由方程组 (1.2) 的第 1 个方程确定出 x_1:

$$x_1=\frac{b_1}{a_{11}}-\sum_{j=2}^{n}\frac{a_{1j}}{a_{11}}x_j.$$

以上是用递归的语言来描述 Gauss 消元法, 也可直接的表述成如下过程. 可以对线性方程组实施如下三类同解变形:

(1) 交换两个方程的位置;

(2) 把某个方程的若干倍加到另一个方程上;

(3) 把某个方程两边同乘以一个非零常数.

按照如下步骤对方程组进行同解变形, 使得最后所得的方程组形式最简单.

(1) 依次考虑所有方程中 $x_1,x_2,\cdots$ 的系数, 设 $j(1)$ 是使得所有方程中 $x_{j(1)}$ 系数不全为零的最小指标. 把某个 $x_{j(1)}$ 系数不为零的方程换到第一个方程的位置, 并把所有其他方程适当的加上第一个方程的若干倍, 使得其他方程中 $x_{j(1)}$ 系数都为零. 经此步骤后, 方程组同解变形为如下形式的方程组:

$$\begin{cases} a_{1j(1)}x_{j(1)} + & a_{1(j(1)+1)}x_{j(1)+1}+\cdots+a_{1n}x_n=b_1;\\ & a_{2(j(1)+1)}x_{j(1)+1}+\cdots+a_{2n}x_n=b_2;\\ & \cdots\cdots\\ & a_{m(j(1)+1)}x_{j(1)+1}+\cdots+a_{mn}x_n=b_m; \end{cases}$$

(2) 对后面 $m-1$ 个方程重复上述操作, 方程组同解变形为如下形式的方程组:

$$\begin{cases} a_{1j(1)}x_{j(1)}+\cdots+a_{1n}x_n=b_1; \\ a_{2j(2)}x_{j(2)}+\cdots+a_{2n}x_n=b_2; \\ a_{2(j(2)+1)}x_{j(2)+1}+\cdots+a_{2n}x_n=b_3; \\ \cdots\cdots \\ a_{m(j(2)+1)}x_{j(2)+1}+\cdots+a_{mn}x_n=b_m; \end{cases}$$

(3) 反复重复上述操作, 直到方程组同解变形为如下形式的方程组:

$$\begin{cases} \overline{a}_{1j(1)}x_{j(1)}+\cdots+\overline{a}_{1n}x_n=\overline{b}_1; \\ \overline{a}_{2j(2)}x_{j(2)}+\cdots+\overline{a}_{2n}x_n=\overline{b}_2; \\ \cdots\cdots \\ \overline{a}_{rj(r)}x_{j(r)}+\cdots+\overline{a}_{rn}x_n=\overline{b}_r; \\ 0=\overline{b}_{r+1}; \\ \cdots\cdots \\ 0=\overline{b}_m; \end{cases} \tag{1.3}$$

其中 $0\le r\le m$, $1\le j(1)<j(2)<\cdots<j(r)\le n$, 且 $\overline{a}_{1j(1)},\overline{a}_{2j(2)},\cdots,\overline{a}_{rj(r)}$ 都非零. 人们把最后所得的方程组称为阶梯形线性方程组.

命题 1.6.1. *(1) 任何线性方程组可以通过对换变换与倍加变换同解变形成阶梯形线性方程组.*

(2) 阶梯形线性方程组 (1.3) 有解当且仅当 $\overline{b}_{r+1}=\cdots=\overline{b}_m=0$.

(3) 当 $\overline{b}_{r+1}=\cdots=\overline{b}_m=0$ 时, 阶梯形线性方程组 (1.3) 的解为: 对 $j\ne j(1),j(2),\cdots,j(r)$, x_j 可以任意取值, $x_{j(1)},x_{j(2)},\cdots,x_{j(r)}$ 由它们唯一确定. 特别的, 如果 $r=n$, 则方程组的解唯一, 如果 $r<n$, 则方程组有无限多个解.

对 $j\ne j(1),j(2),\cdots,j(r)$ 称相应的 x_j 为阶梯形线性方程组 (1.3) 的自由变量; 称 $x_{j(1)},x_{j(2)},\cdots,x_{j(r)}$ 为阶梯形线性方程组 (1.3) 的主变量. 将主变量用自由变量表示出, 称为给出方程组的通解或一般解.

我们考察命题1.6.1的一个特殊形式. 当常数项 $\mathbf{b}=\mathbf{0}$ 时, 对应的方程组 $A\mathbf{x}=\mathbf{0}$, 称为齐次线性方程组, 它显然有解 $\mathbf{x}=\mathbf{0}$, 称为零解或平凡解. 除此之外, 若还有别的解, 则称为非零解或非平凡解.

命题 1.6.2. *如果齐次线性方程组中方程的个数小于未知元的个数, 则该方程组有非零解.*

证明: 由命题1.6.1, 原方程组可同解为阶梯形线性方程组 (1.3), 且显然 $\bar{b}_{r+1} = \cdots = \bar{b}_m = 0$. 注意到, $r \le m < n$, 由命题1.6.1中 (3) 的结论可知方程有无限多个解. □

定理 1.6.3 (De Bruijn–Erdos theorem)**.** *设 n 是大于 2 的整数, 则平面上不共线的 n 个点至少确定 n 条直线.*

证明: 用反证法, 设 n 个不共线的点 $P_1, \cdots, P_n$ 一共确定了 m 条直线 $L_1, \cdots, L_m$, 且 $m < n$. 对每个点 P_i 赋予一个未知元 x_i, 考虑如下由 m 个方程构成的齐次线性方程组

$$\sum_{i:P_i \in L_j} x_i = 0, \quad 1 \le j \le m.$$

由假设 $m < n$ 及前述命题, 可知此方程组有非零解. 对于其非零解 $x_1, \cdots, x_n$, 有

$$0 = \sum_{j=1}^{m} \left(\sum_{i:P_i \in L_j} x_i \right)^2 = \sum_{i=1}^{n} (k_i - 1) x_i^2 + (x_1 + x_2 + \cdots + x_n)^2,$$

其中 k_i 表示过 P_i 的直线条数. 由假设 $P_1, \cdots, P_n$ 不共线可知 $k_i \ge 2$, 这样上式推出每个 x_i 都必须等于零, 矛盾! □

实施 Gauss 消元法时, 只有系数与常数项参与了运算, 因此可以在如下矩阵上展现每一步具体的操作:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right)$$

称此矩阵为线性方程组 (1.2) 的增广矩阵. 这样, 前述关于方程组的三种操作可以表述成关于增广矩阵的操作.

定义 1.6.4. 称如下三种变换为矩阵的初等行变换:

(1) 对换变换: 互换某两行的位置;

(2) 倍加变换: 把某行的若干倍加到另一行上;

(3) 倍乘变换: 把某一行乘以一个非零常数.

Gauss 消元法是把一般的线性方程组通过同解变形变成相对简单的阶梯形方程组. 翻译成矩阵的语言, 可以定义阶梯形矩阵.

定义 1.6.5. 称矩阵 C 是行阶梯矩阵, 如果它满足如下两个条件:

(1) 元素全为零的行 (称为零行), 只能存在于矩阵的下方;

(2) 元素不全为零的行 (称为非零行), 从左数第一个不为零的元素 (称为主元) 的列指标随着行指标的增加而严格递增.

进一步, 称阶梯形矩阵 C 是 (行) 简化阶梯矩阵, 如果它还满足如下两个条件:

(3) 每个非零行的主元都等于 1;

(4) 每个主元所在列的其他元素都是零.

例 1.6.6. 阶梯数等于行数的简化阶梯方阵是单位矩阵.

有了这些术语后, 人们可以把 Gauss 消元法总结成如下定理.

定理 1.6.7. *任意矩阵都可以用对换与倍加两种变换化成阶梯形矩阵; 任意矩阵都可以用对换, 倍加与倍乘三种变换化成简化阶梯形矩阵.*

将线性方程的增广矩阵用对换与倍加变换化成阶梯形矩阵的过程称为 Gauss 消元法; 将线性方程的增广矩阵用对换, 倍加与倍乘变换化成简化阶梯形矩阵的过程称为 Gauss-Jordan 消元法.

注 1.6.8. 对同一个矩阵, 通过初等行变换化成阶梯形矩阵 (或简化阶梯形矩阵), 可采用的方法可能有多种, 最后所得的阶梯形矩阵依赖于具体的转化方法. 但以后我们会证明, 最后所得的阶梯形矩阵中非零行的数目不依赖于具体的转化方法, 它是初始矩阵的一个不变量, 我们暂且称之为初始矩阵的阶梯数.

把线性方程组 (1.2) 的系数矩阵记为 A, 常数项记为 B, 增广矩阵记为 $(A|B)$.

定理 1.6.9 (Rouche-Capelli)**.** (1) *如果增广矩阵的阶梯数大于系数矩阵的阶梯数, 则线性方程组无解.*

(2) *如果增广矩阵的阶梯数等于系数矩阵的阶梯数, 则线性方程组有解. 进一步, 在此条件下有:*

(2.a) *如果系数矩阵的阶梯数等于未知元个数, 则线性方程组有唯一解;*

(2.b) *如果系数矩阵的阶梯数小于未知元个数, 则线性方程组有无穷多个解.*

在上述定理中第 (2) 中情形下, 假设通过 Gauss 消元法最后获得的增广系数矩阵 $(\overline{A}|\overline{B})$ 是阶梯型矩阵, 则可写出方程组的通解: 利用最后一个主元所在的行的数据, 可以把此主元对应的主变量用自由变量表示, 之后再利用倒数第二个主元所在行的信息, 可以把对应的倒数第二个主变量用自由变量与最后那个主变量表示,......, 依次进行下去可以写出通解, 人们把这种从下至上依次求解阶梯形方程的办法称为 “回代法”.

假设通过 Gauss-Jordan 消元法最后获得的增广系数矩阵 $(\overline{A}|\overline{B})$ 是简化阶梯型矩阵, 则可更快的写下通解, 因为每个主元所在的行没有其他主元, 则该行的数据告诉我们如何把对应的主变量用自由变量表示.

1.7 初等矩阵

可用矩阵乘积的语言描述矩阵的初等行 (列) 变换. 考虑对 $m\times n$ 的矩阵 A 进行初等行变换或列变换, 将 A 视为线性映射 $f_A(\mathbf{x})=A\mathbf{x}$ 的表示矩阵. 假设经过某个初等行变换 T 后把 A 变成了矩阵 B. 将 B 视为线性映射 $f_B(\mathbf{x})=B\mathbf{x}$ 的表示矩阵, 简单的直接验证可得: $f_B(\mathbf{x})$ 等于对 $f_A(\mathbf{x})$(视为 $m\times 1$ 的矩阵) 做初等行变换 T. 具体的说, 用映射 $T:\mathbf{R}^m\to\mathbf{R}^m$ 表示把 $\mathbf{R}^m$ 中的向量视为列向量并对它做初等行变换 T 所得到的线性映射, 则有 $f_B=T\circ f_A$. 进一步, 写下线性映射 T 的表示矩阵, 不妨把它的表示矩阵仍然记为 T, 则有 $B=TA$.

对 $\mathbf{R}^m$ 中的列向量实施对换第 i,j 行的行变换所得的线性映射为 $T_{ij}:\mathbf{R}^m\to\mathbf{R}^m$, 将其表示矩阵记为 T_{ij}(《线性代数入门》 54 页上记为 P_{ij}), 则它的所有非零矩阵元为

$$(T_{ij})_{ij}=(T_{ij})_{ji}=1,\quad (T_{ij})_{kk}=1,\quad \forall k\neq i,k\neq j.$$

对 $\mathbf{R}^m$ 中的列向量实施将第 i 行的 k 倍加到第 j 行的行变换所得的线性映射为 $L_{ij}(k):\mathbf{R}^m\to\mathbf{R}^m$, 将其表示矩阵记为 $L_{ij}(k)$(《线性代数入门》 55 页上记为 $E_{ji;k}$), 则它的所有非零矩阵元为

$$(L_{ij}(k))_{ji}=k,\quad (L_{ij}(k))_{ll}=1,\quad \forall l.$$

对 $\mathbf{R}^m$ 中的列向量实施将第 i 行乘以非零常数 k 的行变换所得的线性映射为 $D_i(k):\mathbf{R}^m\to\mathbf{R}^m$, 将其表示矩阵记为 $D_i(k)$(《线性代数入门》 55 页上记为 $E_{ii;k}$), 则它的所有非零矩阵元为

$$(D_i(k))_{ii}=k,\quad (D_i(k))_{ll}=1,\quad \forall l\neq i.$$

我们把上述三种矩阵 $T_{ij},L_{ij}(k),D_i(k)$ 统称为初等矩阵.

定义 1.7.1. 设 A 是 n 阶方阵, 定义 A 的转置矩阵 A^T 为 n 阶方阵:

$$(A^T)_{ij}=A_{ji},\quad \forall 1\leq i,j\leq n,$$

即 A^T 的第 i 个行向量等于 A 的第 i 个列向量.

命题 1.7.2. (1) 将矩阵 A 的第 i,j 行互换所得的矩阵为 $T_{ij}A$, 将矩阵 A 的第 i,j 列互换所得的矩阵为 $A(T_{ij})^T$.

(2) 将矩阵 A 的第 i 行的 k 倍加到第 j 行所得的矩阵为 $L_{ij}(k)A$, 将矩阵 A 的第 i 列的 k 倍加到第 j 列所得的矩阵为 $A(L_{ji}(k))^T$.

(3) *将矩阵 A 的第 i 行乘以非零常数 k 所得的矩阵为 $D_i(k)A$, 将矩阵 A 的第 i 列乘以非零常数 k 所得的矩阵为 $A(D_i(k))^T$.*

证明*:* 用矩阵乘法的定义式直接验证. □

总结一下, 对矩阵 A 做初等行变换等价于在 A 左边乘以初等矩阵, 对 A 做初等列变换等价于在 A 右边乘以初等矩阵.

1.8 可逆矩阵

注意到, 初等矩阵都是方阵, 不但如此, 它们都是可逆矩阵.

定义 1.8.1. 称 n 阶方阵 A 是可逆矩阵 (或者非奇异矩阵), 如果存在一个 n 阶方阵 B 使得 $AB = BA = I_n$, 并称 B 为 A 的逆矩阵, 记为 A^{-1}. 不可逆的矩阵称为奇异矩阵.

如果矩阵 A 可逆的话, 则 A 的逆矩阵唯一. 这是因为, 如果 B 与 B' 都满足上述定义式中的等式, 则有

$$B = BI_n = B(AB') = (BA)B' = I_nB' = B'.$$

以后我们会证明: 如果 B 与 A 是同阶方阵, 且满足 $AB = I_n$ 或 $BA = I_n$ 中的一个, 则 B 是 A 的逆矩阵. 这个性质是矩阵代数所特有的, 在更一般的代数中并不成立.

命题 1.8.2. *设 A 是 n 阶可逆矩阵, X, B 是 $n \times l$ 矩阵, 则 $AX = B$ 成立当且仅当 $X = A^{-1}B$.*

证明*:* 若 $AX = B$, 两边同乘以 A^{-1} 可得

$$X = I_nX = (A^{-1}A)X = A^{-1}(AX) = A^{-1}B;$$

另一方面, 若 $X = A^{-1}B$, 则有

$$AX = A(A^{-1}B) = (AA^{-1})B = I_nB = B.$$

□

命题 1.8.3. *初等矩阵都是可逆矩阵, 且有 $(P_{ij})^{-1} = P_{ij}$, $(L_{ij}(k))^{-1} = L_{ij}(-k)$, $(D_i(k))^{-1} = D_i(\frac{1}{k})$.*

例 1.8.4. 称方阵 D 为对角矩阵, 如果 D 的非对角元都等于零. 可把对角矩阵记为 $D = \mathrm{diag}\{d_{11}, \cdots, d_{nn}\}$. 对角矩阵 D 左乘以 n 阶方阵 B 的结果是: 将 B 第 i 行乘以 d_{ii}; 对角矩阵 D 右乘以 n 阶方阵 B 的结果是: 将 B 第 i 列乘以 d_{ii}. 这样, 如果 $DB = I_n$, 则有 $d_{ii}b_{ii} = 1$, 由此可知 D 是可逆矩阵的必要条件是每个 d_{ii} 都不等于零; 另一方面, 在此条件下可知 $D^{-1} = \mathrm{diag}\{d_{11}^{-1}, \cdots, d_{nn}^{-1}\}$.

命题 1.8.5. 如果 A, B 是可逆矩阵, 则 AB 也是可逆矩阵, 且有 $(AB)^{-1} = B^{-1}A^{-1}$.

定理 1.8.6. 设 A 是 n 阶方阵, 则以下断言彼此等价.

(1) A 是可逆矩阵;

(2) 对任何 n 维向量 $\mathbf{b}$, 方程组 $A\mathbf{x} = \mathbf{b}$ 有唯一解 $\mathbf{x} = A^{-1}\mathbf{b}$.

(3) 齐次方程组 $A\mathbf{x} = \mathbf{0}$ 只有零解;

(4) 将 A 化成阶梯形矩阵后恰有 n 个主元;

(5) 将 A 化成简化梯形矩阵得到的矩阵是 I_n;

(6) A 是有限多个初等矩阵的乘积.

证明: (5) $\Rightarrow$ (6) 由命题1.7.2, 对 A 做初等行变换的效果等于在 A 的左边乘以初等矩阵, 这样有 (5) 的结论可知, 存在有限多个初等矩阵 $E_1, E_2, \cdots, E_t$, 使得 $E_t \cdots E_1 A = I_n$, 从而有 $A = (E_t \cdots E_1)^{-1} = E_1^{-1} \cdots E_t^{-1}$, 它是有限多个初等矩阵的乘积. □

命题 1.8.7. 设 A 是 n 阶可逆矩阵, B 是 $n \times l$ 矩阵, 令 $(A|B)$ 是把 A 作为前 n 列把 B 作为后 l 列所构成的 $n \times (n+l)$ 矩阵. 如果对矩阵 $(A|B)$ 做若干次初等行变换使得左边的 $n \times n$ 矩阵变成 I_n, 则变换结束时所得矩阵为 $(I_n|A^{-1}B)$.

上述命题的好处是: Gauss-Jordan 消元法给出一系列确定的步骤, 告诉人们怎么通过初等行变换把 A 变成简化阶梯型矩阵, 当 A 可逆时, 所得的简化阶梯型矩阵即为 I_n. 这样, 只要把 B 写在 A 的右边, 当 Gauss-Jordan 消元法对 A 做初等行变换时, 让这些初等行变换同时也作用在 B 上, 上述命题告诉我们, 当所有操作结束时, 右边的矩阵就变成了 $A^{-1}B$. 这样, 如果人们要求解复杂的矩阵方程 $AX = B$, 只需对增广矩阵 $(A|B)$ 做初等行变换, 利用 Gauss-Jordan 消元法告诉我们的步骤把左边的 $n \times n$ 变成 I_n, 一旦实现了这个目标, 则最后所得矩阵右边的 $n \times l$ 矩阵就是所要求解方程 $AX = B$ 的解! 这个方法特别适用于计算机编程运行, 因为在实际问题中人们可能关心对很多不同的矩阵 B, 求解 $AX = B$, 人们可以把 A 的 Gauss-Jordan 消元法所确定的初等行变换步骤存储下来, 对不同的常数项 B, 让计算机提取之前存储的初等行变换步骤, 用于每个 $(A|B)$ 即可很快的求解一系列方程.

称 n 阶方阵 $A = (a_{ij})_{1 \le i,j \le n}$ 是上三角矩阵, 如果对任何 $i > j$ 都有 $a_{ij} = 0$; 称 A 是严格上三角矩阵, 如果对任何 $i \ge j$ 都有 $a_{ij} = 0$. 类似的, 可定义下三角矩阵与严格下三角矩阵.

命题 1.8.8. 上三角矩阵 A 可逆当且仅当 A 的对角元素都不等于零. 此时, 其逆矩阵也是上三角矩阵, 且 $(A^{-1})_{ii} = (A_{ii})^{-1}$.

证明: 若 A 的对角元素都不等于零, 利用回代法可得 $A\mathbf{x} = \mathbf{0}$ 只有零解; 若 A 的对角元素不全为零, 设 k 是使得 $a_{kk} = 0$ 的最小指标, 则可构造 $A\mathbf{x} = \mathbf{0}$ 的非零解为 $x_k = 1, x_{k+1} = \cdots =$

$x_n = 0$, $x_1, \cdots, x_{k-1}$ 由回代法由 x_k 唯一确定. 这样, 由定理1.8.6可知 A 可逆当且仅当 A 的对角元素都不等于零.

进一步, 假设 A 的对角元素都不等于零, 记 $B = A^{-1}$, 则对每个 $1 \le k \le n$, 有

$$A \begin{pmatrix} b_{1k} \\ \vdots \\ b_{kk} \\ \vdots \\ b_{nk} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

解得 $b_{kk} = (a_{kk})^{-1}$, 且对 $i > k$ 有 $b_{ik} = 0$. □

命题 1.8.9. *设 A, B 是 n 阶方阵, 满足 $AB = I_n$, 则 A, B 互为逆矩阵.*

证明: 对任何向量 $\mathbf{v} \in \mathbf{R}^n$, 令 $\mathbf{x} = B\mathbf{v}$, 则有 $A\mathbf{v} = A(B\mathbf{v}) = (AB)\mathbf{v} = \mathbf{v}$, 即方程组 $A\mathbf{x} = \mathbf{v}$ 总有解.

我们先证明 A 可逆. 用反证法, 假设 A 不可逆. 用 Gauss 消元法对 A 实施有限次初等行变换 (记为操作 T) 化成阶梯矩阵 A', 由定理1.8.6可知 A' 的阶梯数 r 小于 n. 令 $\mathbf{b}' = \mathbf{e}_{r+1}$ 为第 $r+1$ 个坐标向量, 考虑增广矩阵 $(A'|\mathbf{b}')$, 对它实施前述有限次初等行变换 T 的逆变换 T^{-1}, 设所得的增广矩阵为 $(A|\mathbf{b})$. 这样, 方程组 $A\mathbf{x} = \mathbf{b}$ 与方程组 $A'\mathbf{x} = \mathbf{b}'$ 同解, 后者显然无解, 故 $A\mathbf{x} = \mathbf{b}$ 也无解, 与前一段的结论矛盾!

已经证明了 A 有逆, 由命题1.8.2即得 $B = A^{-1}$. □

定义 1.8.10. 称 n 阶方阵 $A = (a_{ij})_{1 \le i,j \le n}$ 是 (行) 对角占优矩阵, 如果对每个 $1 \le i \le n$ 都有 $|a_{ii}| > \sum\limits_{j \ne i} |a_{ij}|$.

定理 1.8.11. *对角占优矩阵一定是可逆矩阵.*

证明: 用反证法, 假设对角占优矩阵 A 不可逆, 则由定理1.8.6 可知齐次方程 $A\mathbf{x} = 0$ 有非零解 $\mathbf{x}$. 考虑 $|x_1|, \cdots, |x_n|$ 的最大值, 设为 $|x_i|$, 则有

$$\left(\sum_{j \ne i} |a_{ij}| \right) |x_i| < |a_{ii} x_i| = \left| -\sum_{j \ne i} a_{ij} x_j \right| \le \sum_{j \ne i} |a_{ij} x_j| \le \left(\sum_{j \ne i} |a_{ij}| \right) |x_i|,$$

矛盾! 这就完成了证明. □

1.9 矩阵运算

设 A, B 是 $m \times n$ 矩阵, k 是实数, 可定义矩阵的加法与数乘运算为

$$(A+B)_{ij} = A_{ij} + B_{ij}, \quad (kA)_{ij} = kA_{ij}, \quad \forall 1 \le i \le m, 1 \le j \le n.$$

令 $M_{m\times n}$ 为所有 $m \times n$ 矩阵所构成的集合, 则其上有加法和数乘运算, 是个线性空间.

可在所有线性映射所构成的空间上看到这个线性结构. 令 $\mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ 为所有线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$ 所构成的集合, 则对于 $f, g \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ 与 $k \in \mathbf{R}$, 可定义 $f+g, kf \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ 为:

$$(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \quad (kf)(\mathbf{x}) = kf(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{R}^n.$$

取每个线性映射的矩阵表示, 则可得到从 $L(\mathbf{R}^n, \mathbf{R}^m)$ 到 $M_{m\times n}$ 的等同, 且此等同保持加法与数乘运算, 以后我们称这种等同为线性空间之间的同构.

对于给定的正整数 n, 令 M_n 为所有 n 阶方阵所构成的集合. 这样, M_n 上有矩阵的加法, 数乘以及矩阵乘法三种运算, 其中矩阵乘法与加法满足分配律

$$A(B+C) = AB + AC, \quad (B+C)A = BA + CA;$$

矩阵乘法与数乘运算相容

$$(kA)B = A(kB) = k(AB).$$

以后, 人们称上述三种运算赋予 M_n 一个代数结构.

对于 $A \in M_n$, 对任何非负整数 p, 定义 A 的 p 次幂 A^p 为 p 个 A 的乘积; 进一步, 对于任何一元多项式 $f(x) = a_p x^p + \cdots + a_1 x + a_0$, 定义

$$f(A) = a_p A^p + \cdots + a_1 A + a_0 I_n.$$

1.10 分块矩阵

增广系数矩阵 $(A|B)$ 分成左右两块, 每块都是矩阵; 此外, 我们经常将 $m \times n$ 矩阵视为由 m 个行向量或者 n 个列向量组合而成. 将矩阵分块表示, 有些时候可以更有效的表述计算过程.

我们来考虑一般的分块方式. 设 A 是 $m \times n$ 矩阵, 正整数 $m_1, \cdots, m_p; n_1, \cdots, n_q$ 满足 $m_1 + \cdots + m_p = m$, $n_1 + \cdots + n_q = n$, 则可将 A 表示成

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \vdots \\ A_{p1} & \cdots & A_{pq} \end{pmatrix},$$

其中 A_{ij} 是 $m_i \times n_j$ 矩阵, 称上述形式为分块矩阵.

设 B 是 $n \times l$ 矩阵, 表示成如下分块矩阵的形式

$$B = \begin{pmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \vdots \\ B_{q1} & \cdots & B_{qs} \end{pmatrix},$$

其中 B_{jk} 是 $n_j \times l_k$ 矩阵, $l_1 + \cdots + l_s = l$, 则乘积矩阵可以表示成分块矩阵:

$$\begin{pmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \vdots \\ A_{p1} & \cdots & A_{pq} \end{pmatrix} \begin{pmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \vdots \\ B_{q1} & \cdots & B_{qs} \end{pmatrix} = \begin{pmatrix} \sum\limits_{j=1}^{q} A_{1j}B_{j1} & \cdots & \sum\limits_{j=1}^{q} A_{1j}B_{js} \\ \vdots & & \vdots \\ \sum\limits_{j=1}^{q} A_{pj}B_{j1} & \cdots & \sum\limits_{j=1}^{q} A_{pj}B_{js} \end{pmatrix}.$$

例 1.10.1. 分块对角矩阵是指如下形式的矩阵:

$$A = \begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{pp} \end{pmatrix}$$

其中 A_{ii} 是 $n_i \times n_i$ 矩阵, $n = n_1 + \cdots + n_p$. 可以从线性映射的角度理解分块对角矩阵, 设 A 给出的线性映射为 $f : \mathbf{R}^n \to \mathbf{R}^n$. 将 $\mathbf{R}^n$ 中的向量也分块, 表示成

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{pmatrix},$$

其中 $\mathbf{x}_i$ 是 $\mathbf{R}^{n_i}$ 中的向量, 可以视为 $n_i \times 1$ 矩阵. 这样, f 的作用为

$$f(\mathbf{x}) = \begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{pp} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{pmatrix} = \begin{pmatrix} A_{11}\mathbf{x}_1 \\ A_{22}\mathbf{x}_2 \\ \vdots \\ A_{pp}\mathbf{x}_p \end{pmatrix},$$

即 f 把 $\mathbf{R}^{n_i}$ 块映到 $\mathbf{R}^{n_i}$ 块, 在这块上的作用由 A 的第 i 个对角块 A_{ii} 给出.

除了有可能简化某些运算之外, 引入分块矩阵的理由何在? 将 $\mathbf{R}^{m+n}$ 中的向量分块表示为 $\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$ 的形式, 其中 $\mathbf{u} \in \mathbf{R}^m, \mathbf{v} \in \mathbf{R}^n$. 这样, 对于任何线性映射 $f : \mathbf{R}^s \to \mathbf{R}^{m+n}$, f 由它的两个分量完全确定:

$$f(\mathbf{z}) = \begin{pmatrix} f_1(\mathbf{z}) \\ f_2(\mathbf{z}) \end{pmatrix},$$

其中 $f_1 : \mathbf{R}^s \to \mathbf{R}^m, f_2 : \mathbf{R}^s \to \mathbf{R}^n$ 是线性映射. 假设 f_1 与 f_2 的表示矩阵分别为 A_1, A_2, 则 f 的表示矩阵为分块矩阵 $\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$. 类似的, 对于线性映射 $h : \mathbf{R}^{p+q} \to \mathbf{R}^l$, 有

$$h\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = h\begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix} + h\begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix},$$

记 $f(\mathbf{x}) = h\begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}$, $g(\mathbf{y}) = h\begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix}$, 则 $f : \mathbf{R}^p \to \mathbf{R}^l$ 与 $g : \mathbf{R}^q \to \mathbf{R}^l$ 都是线性映射, 且有

$$h\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = f(\mathbf{x}) + g(\mathbf{y}).$$

如果 f 与 g 的表示矩阵分别为 A 与 B, 则 h 的表示矩阵为分块矩阵 (A, B).

进一步, 考虑线性映射 $h : \mathbf{R}^{p+q} \to \mathbf{R}^{m+n}$. 设 $h\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = f(\mathbf{x}) + g(\mathbf{y})$, 其中 $f : \mathbf{R}^p \to \mathbf{R}^{m+n}$, $g : \mathbf{R}^q \to \mathbf{R}^{m+n}$. 设 f 与 g 的分量函数分别为 f_1, f_2 与 g_1, g_2, 对应的矩阵表示分别为 A_1, A_2 与 B_1, B_2, 则 h 为

$$h\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} f_1(\mathbf{x}) + g_1(\mathbf{y}) \\ f_2(\mathbf{x}) + g_2(\mathbf{y}) \end{pmatrix},$$

它的矩阵表示为 $\begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix}$.

设 A, B, C, D 分别是 $p \times p, p \times q, q \times p, q \times q$ 矩阵, 考虑分块矩阵

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

它是 $(p+q) \times (p+q)$ 矩阵, 我们来确定 M 的逆.

从线性映射的角度看, 将 $\mathbf{R}^{p+q}$ 中的向量表示为 $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ 的形式, 其中 $\mathbf{x} \in \mathbf{R}^p, \mathbf{y} \in \mathbf{R}^q$, 则 M 给出的线性映射为

$$f\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} A\mathbf{x} + B\mathbf{y} \\ C\mathbf{x} + D\mathbf{y} \end{pmatrix}.$$

为了确定 M 的逆矩阵, 等价于求 f 的逆映射, 即求解方程组

$$\begin{cases} A\mathbf{x}+B\mathbf{y}=\mathbf{u}, \\ C\mathbf{x}+D\mathbf{y}=\mathbf{v}. \end{cases}$$

如果 A 可逆, 则由第一个方程可得 $\mathbf{x}=A^{-1}(\mathbf{u}-B\mathbf{y})$, 代回第二个方程解得

$$\begin{cases} \mathbf{x}=A^{-1}(\mathbf{u}-B\mathbf{y}), \\ (D-CA^{-1}B)\,\mathbf{y}=\mathbf{v}-CA^{-1}\mathbf{u}, \end{cases}$$

记 $M/A=D-CA^{-1}B$, 称为 A 在 M 中的 Schur 补, 如果 M/A 可逆, 则有

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{bmatrix} A^{-1}+A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

类似的, 如果 D 是可逆矩阵, 定义 D 在 M 中的 Schur 补为 $M/D=A-BD^{-1}C$; 如果 M/D 可逆, 则有

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{bmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1}+D^{-1}C(M/D)^{-1}BD^{-1} \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

命题 1.10.2. 如果 A 与 M/A 都可逆, 则 M 可逆, 且有

$$M^{-1} = \begin{bmatrix} A^{-1}+A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{bmatrix};$$

类似的, 如果 D 与 M/D 都可逆, 则 M 可逆, 且有

$$M^{-1} = \begin{bmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1}+D^{-1}C(M/D)^{-1}BD^{-1} \end{bmatrix}.$$

推论 1.10.3. 如果 $A, D, M/A, M/D$ 都可逆, 则有

$$(M/D)^{-1}=A^{-1}+A^{-1}B(M/A)^{-1}CA^{-1}.$$

推论 1.10.4 (Scherman-Morrison)**.** 设 A 是 n 阶可逆方阵, $\mathbf{u}, \mathbf{v}$ 是 n 维向量, 则 $A+\mathbf{u}\mathbf{v}^T$ 可逆的充分必要条件是 $1+\mathbf{v}^TA^{-1}\mathbf{u}\neq 0$. 进一步, 在此条件下有

$$(A+\mathbf{u}\mathbf{v}^T)^{-1}=A^{-1}-\frac{A^{-1}\mathbf{u}\mathbf{v}^TA^{-1}}{1+\mathbf{v}^TA^{-1}\mathbf{u}}.$$

证明: 只需再证明若 $1+\mathbf{v}^TA^{-1}\mathbf{u}=0$, 则 $A+\mathbf{u}\mathbf{v}^T$ 不可逆. 这是由于

$$\left(A+\mathbf{u}\mathbf{v}^T\right)A^{-1}\mathbf{u}=\mathbf{u}+\mathbf{u}(\mathbf{v}^TA^{-1}\mathbf{u})=\mathbf{u}-\mathbf{u}=\mathbf{0}.$$

□

也可用如下方法求 M 的逆矩阵. 为了求解如下实系数线性方程组

$$\begin{cases}ax+by=u,\\ cx+dy=v,\end{cases}$$

Gauss 消元法将第一个方程的 $-ca^{-1}$ 倍加到第二个方程上, 用以消去第二个方程中 x 的系数, 这等价于对系数矩阵 $\begin{pmatrix}a & b\\ c & d\end{pmatrix}$ 做倍加操作, 在左边乘以矩阵 $\begin{pmatrix}1 & 0\\ -ca^{-1} & 1\end{pmatrix}$. 对分块矩阵作类似的操作, 假设 A 可逆, 则有

$$\begin{pmatrix}I_p & 0\\ -CA^{-1} & I_q\end{pmatrix}\begin{pmatrix}A & B\\ C & D\end{pmatrix}=\begin{pmatrix}A & B\\ 0 & D-CA^{-1}B\end{pmatrix}. \tag{1.4}$$

引理 1.10.5. (1) 分块上三角矩阵 $\begin{pmatrix}P & Q\\ 0 & R\end{pmatrix}$ 可逆的充分必要条件是 P,R 可逆. 在此条件下, 有

$$\begin{pmatrix}P & Q\\ 0 & R\end{pmatrix}^{-1}=\begin{pmatrix}P^{-1} & -P^{-1}QR^{-1}\\ 0 & R^{-1}\end{pmatrix}.$$

(2) 分块下三角矩阵 $\begin{pmatrix}P & 0\\ Q & R\end{pmatrix}$ 可逆的充分必要条件是 P,R 可逆. 在此条件下, 有

$$\begin{pmatrix}P & 0\\ Q & R\end{pmatrix}^{-1}=\begin{pmatrix}P^{-1} & 0\\ -R^{-1}QP^{-1} & R^{-1}\end{pmatrix}.$$

证明: (1) 记 $M=\begin{pmatrix}P & Q\\ 0 & R\end{pmatrix}$, 只需证明如果 M 可逆, 则 P,R 都可逆. 如果 P 不可逆, 则 $P\mathbf{x}=\mathbf{0}$ 有非零解 $\mathbf{x}_0$, 由此可得

$$\begin{pmatrix}P & Q\\ 0 & R\end{pmatrix}\begin{pmatrix}\mathbf{x}_0\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}P\mathbf{x}_0\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix},$$

与 M 可逆矛盾, 所以 P 可逆.

如果 R 不可逆, 则 $R\mathbf{y}=\mathbf{0}$ 有非零解 $\mathbf{y}_0$. 令 $\mathbf{x}_0=-P^{-1}Q\mathbf{y}_0$, 可得

$$\begin{pmatrix} P & Q \\ 0 & R \end{pmatrix}\begin{pmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{pmatrix}=\begin{pmatrix} P\mathbf{x}_0+Q\mathbf{y}_0 \\ R\mathbf{y}_0 \end{pmatrix}=\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

与 M 可逆矛盾, 所以 R 可逆. □

结合 (1.4) 式与引理1.10.5, 即可得到命题1.10.2的更强版本.

命题 1.10.6. *设 A 可逆, 则分块矩阵 $M=\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ 可逆的充分必要条件是 A 在 M 中的 Schur 补矩阵 M/A 可逆. 在此条件下有可通过 (1.4) 式与引理1.10.5算出 M^{-1}.*

类似的, 如果 D 可逆, 则分块矩阵 $M=\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ 可逆的充分必要条件是 D 在 M 中的 Schur 补矩阵 M/D 可逆.

1.11 LU 分解: Gauss 消元法的矩阵表示

定义 1.11.1. 方阵 A 的 LU 分解 (*lower-upper decomposition*) 是指将 A 表示成矩阵乘积 $A=LU$, 其中 L 是下三角矩阵, U 是上三角矩阵.

为什么关心 LU 分解? 因为如果 A 有 LU 分解 $A=LU$, 则线性方程组 $A\mathbf{x}=\mathbf{b}$ 可表述为 $LU\mathbf{x}=\mathbf{b}$. 令 $\mathbf{y}=U\mathbf{x}$ 进行换元, 则等价于求解

$$\begin{cases} L\mathbf{y}=\mathbf{b}, \\ U\mathbf{x}=\mathbf{y}. \end{cases}$$

前一个方程组为:

$$\begin{cases} L_{11}y_1=b_1, \\ L_{21}y_1+L_{22}y_2=b_2, \\ \cdots \\ L_{n1}y_1+\cdots+L_{nn}y_n=b_n, \end{cases}$$

可用前代法解出; 后一个方程组为:

$$\begin{cases} U_{11}x_1+\cdots+U_{1n}x_n=y_1, \\ \cdots \\ U_{(n-1)(n-1)}x_{n-1}+U_{(n-1)n}x_n=y_{n-1}, \\ U_{nn}x_n=y_n, \end{cases}$$

可用回代法解出. 不但如此, 已知分解 $A = LU$ 后, 上述两个方程组的系数矩阵 L, U 都是确定的, 这样对于不同的常数项 $\mathbf{b}$, 可用同样的步骤求解很多方程组 $A\mathbf{x} = \mathbf{b}$.

为了方便起见, 称倍加变换 $L_{ij}(k)$ 为 "顺序倍加变换", 如果 $i < j$. 对应到 Gauss 消元法中, 即是用排在前面的方程的若干倍加到后面的某个方程进行消元. 注意到,"顺序倍加变换"$L_{ij}(k)$ 的逆变换 $L_{ij}(-k)$ 也是 "顺序倍加变换", 且它们对应的初等行矩阵是下三角矩阵.

命题 1.11.2. *如果方阵 A 可以只用 "顺序倍加变换" 就可以化成阶梯形矩阵, 则 A 有 LU 分解.*

证明: 设 A 经过 k 次 "顺序倍加变换"$T_1, \cdots, T_k$ 化成阶梯形矩阵 U, 则有

$$T_k \cdots T_1 A = U,$$

由此可得

$$A = (T_k \cdots T_1)^{-1} U = (T_1)^{-1} \cdots (T_k)^{-1} U.$$

令 $L = (T_1)^{-1} \cdots (T_k)^{-1}$, 由于每个 T_i 是 "顺序倍加变换", 则矩阵 $(T_i)^{-1}$ 是下三角矩阵, 从而可知 L 也是下三角矩阵. □

定义 1.11.3. 设 $A = (A_{ij})_{1 \le i,j \le n}$ 是 n 阶矩阵, 对每个正整数 $k \le n$, 称矩阵 $A_k = (A_{ij})_{1 \le i,j \le k}$ 为 A 的第 k 个顺序主子阵, 即是 A 的前 k 行与前 k 列交叉处的 $k \times k$ 矩阵.

定理 1.11.4. *设 A 是 n 阶可逆矩阵, 则 A 有 LU 分解的充分必要条件是 A 的所有顺序主子阵都是可逆矩阵.*

证明: 必要性. 设 A 有 LU 分解 $A = LU$, 记 A, L, U 的第 k 个顺序主子阵为 A_k, L_k, U_k. 设 A 的逆为 B, 则有 $LUB = I_n = BLU$, 由命题1.8.9可知 L, U 都可逆. 再由命题1.8.8可得 L_k, U_k 都可逆. 直接计算分块矩阵的乘积可得 $A_k = L_k U_k$, 故 A_k 可逆.

充分性. 假设 A 的所有顺序主子阵都可逆, 由命题1.11.2, 只需证明可以只用 "顺序倍加变换" 将 A 化成阶梯形矩阵. 用反证法, 假设只用 "顺序倍加变换" 只能将 A 的前 $k-1$ 行化成阶梯形, 其中 $1 \le k \le n$. 设此时所得的矩阵为 A', 将从 A 到 A' 用到的 "顺序倍加变换" 的操作序列记为 T, 则 $(A')_{kk} = 0$. 考虑增广系数矩阵 $((A')_k | \mathbf{e}_k)$, 对它实施 T 的逆操作, 设得到的增广系数矩阵为 $(A_k | \mathbf{v})$. 由于是进行了初等行变换, 方程组 $(A')_k \mathbf{x} = \mathbf{e}_k$ 与 $A_k \mathbf{x} = \mathbf{v}$ 同解, 前者显然无解, 但后者系数矩阵可逆因而总有解, 矛盾! □

从上述定理可知, 如果可逆矩阵 A 有 LU 分解, 则可以只用 "顺序倍加变换" 将 A 化成阶梯形矩阵, 从而获得 A 的一个 LU 分解, 其中 L 是若干个 "顺序倍加变换" 的表示矩阵的乘积. 注意到,"顺序倍加变换" 的表示矩阵都是下三角矩阵, 且其对角元都是 1, 由此可知 L 的对角元也都等于 1.

定义 1.11.5. 称 L 为单位下三角矩阵, 如果 L 是下三角矩阵, 且其对角元都是 1. 类似的, 可定义单位上三角矩阵.

利用此定义, 可以把我们之前的结论总结成: 如果可逆矩阵 A 有 LU 分解, 则可以把 A 分解为 $A = LU$ 的形式, 其中 L 是单位下三角矩阵, U 是上三角矩阵.

命题 1.11.6 (LU 分解的唯一性)**.** *设 A 是可逆矩阵, 如果 $A = L_1U_1 = L_2U_2$, 其中对每个 i 有 L_i 是单位下三角矩阵且 U_i 是上三角矩阵, 则 $L_1 = L_2$ 且 $U_1 = U_2$.*

证明: 对下三角矩阵, 命题1.8.8的相应版本也成立, 由此可知 L_1, L_2 可逆. 由条件 A 可逆, 可知 $U_i = (L_i)^{-1}A$ 都可逆. 这样就有 $(L_1)^{-1}L_2 = U_1(U_2)^{-1}$, 注意到 $(L_1)^{-1}L_2$ 是单位下三角矩阵, $U_1(U_2)^{-1}$ 是上三角矩阵, 由于两者相等, 则它们都等于单位矩阵, 这就证明了 $L_1 = L_2, U_1 = U_2$. □

命题 1.11.7 (LDU 分解)**.** *设可逆矩阵 A 有 LU 分解, 则 A 可以分解为 $A = LDU$, 其中 L 是单位下三角矩阵, D 是对角矩阵, U 是单位上三角矩阵. 进一步, A 的 LDU 分解是唯一的.*

证明: 由假设, 可把 A 分解为 $A = LV$, 其中 L 是单位下三角矩阵, V 是可逆上三角矩阵. 令 $D = \mathrm{diag}\{V_{11}, \cdots, V_{nn}\}$, $U = D^{-1}V$, 即得所需要的分解 $A = LDU$.

若 $A = L_1D_1U_1 = L_2D_2U_2$ 是两个这样的分解, 则对于 $L_1(D_1U_1) = L_2(D_2U_2)$ 使用命题1.11.6的结论可得 $L_1 = L_2$ 且 $D_1U_1 = D_2U_2$. 这样就有 $(D_1)^{-1}D_2 = U_1(U_2)^{-1}$, 此等式左边是对角矩阵, 右边是单位上三角矩阵, 因而可得两者都是单位矩阵, 表明 $D_1 = D_2, U_1 = U_2$. □

推论 1.11.8. *设 A 是可逆对称矩阵, 如果 A 有 LDU 分解 $A = LDU$, 则有 $U = L^T$.*

证明: 对 $A = LDU$ 取转置可得 $A^T = U^TD^TL^T$, 由 A 是对称矩阵, 可得 $A = U^TD^TL^T$. 注意到这也是 A 的 LDU 分解, 由 LDU 分解的唯一性即可知 $U^T = L$. □

定义 1.11.9. 从单位矩阵出发, 经过有限次对换行操作所得的矩阵称为置换矩阵. 方阵 A 是置换矩阵的充分必要条件是: A 的每行每列都只有一个非零矩阵元, 且非零矩阵元都是 1.

定理 1.11.10. *设 A 是可逆矩阵, 则存在置换矩阵 P 使得乘积矩阵 PA 有 LU 分解.*

证明: 对 A 的阶数用归纳法. 由于 A 可逆, A 第一列的矩阵元不全为零, 可适当交换行使得 $A_{11} \neq 0$. 将第 $2, 3, \cdots, n$ 行加上第 1 行的适当倍, 用来消去第一列的其他矩阵元, 设得到的矩阵为 A', 则 A' 形如如下分块矩阵

$$A' = \begin{pmatrix} a_{11} & * \\ \mathbf{0} & B \end{pmatrix},$$

其中 B 是 $n-1$ 阶可逆矩阵. 利用归纳假设, 可对 B 进行若干行对换操作 (记为 Q, 这里 Q 是有关第 2 到第 n 行的置换操作), 使得 QB 的所有顺序主子阵都可逆. 可以把 Q 视为对所有 n 行的置换操作, 考虑 $QA' = \begin{pmatrix} a_{11} & * \\ \mathbf{0} & QB \end{pmatrix}$, 考虑阶梯数可知 QA' 的所有顺序主子阵都可逆.

注意到, 从 QA 变到 QA', 是通过将第 $2,3,\cdots,n$ 行加上第 1 行的适当倍这种操作得到的, 则 QA' 的每个顺序主子阵 $(QA')_k$ 等于倍加矩阵左乘 $(QA)_k$ 得到. 由此可得 QA 的所有顺序主子式都可逆, 结合定理1.11.4, 这就证明了 QA 有 LU 分解. □

例 1.11.11. 设 n 阶方阵 T_n 为

$$T_n = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix},$$

我们来考察 T_n 的 LU 分解. 命题1.11.2告诉我们, 如果能用一系列 "顺序倍加变换" 将 T_n 化成阶梯形矩阵, 则给出 LU 分解. 注意到, 实施倍加变换 $L_{12}(1)$ 后, 矩阵变成 $\begin{pmatrix} 1 & -(\mathbf{e}_1)^T \\ \mathbf{0} & T_{n-1} \end{pmatrix}$, 递归的处理左下角的 T_{n-1}, 可得

$$(L_{n-1,n}(1)) \cdot \cdots \cdot (L_{23}(1)) \cdot (L_{12}(1)) \cdot T_n = U,$$

其中 U 为如下的单位上三角矩阵:

$$U = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & \ddots & \\ & & & \ddots & -1 \\ & & & & 1 \end{bmatrix}.$$

这就得到 T_n 的 LU 分解 $T_n = LU$, 其中

$$L = (L_{12}(1))^{-1}(L_{23}(1))^{-1} \cdot \cdots \cdot (L_{n-1,n}(1))^{-1} = (L_{12}(-1))(L_{23}(-1)) \cdot \cdots \cdot (L_{n-1,n}(-1)).$$

为了具体算出 L, 可用矩阵乘法直接计算, 也可注意到在矩阵左边乘以 $L_{ij}(k)$ 等于对矩阵实施倍加变换. 这样, L 等于从 I_n 出发, 相继实施 $L_{n-1,n}(-1),\cdots,L_{23}(-1),L_{12}(-1)$ 所得到的矩

阵. 由此可得

$$L = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}.$$

第二章　子空间和维数

2.1　子空间和维数

定义 2.1.1. 设 V 是 $\mathbf{R}^n$ 的非空子集, 称 V 是 $\mathbf{R}^n$ 的线性子空间, 如果对任何 $\mathbf{x}, \mathbf{y} \in V$ 及任何实数 k, 有 $\mathbf{x}+\mathbf{y} \in V$, 且 $k\mathbf{x} \in V$, 即要求 V 关于向量加法与数乘运算都是封闭的.

注意在上述定义中, 由于要求了 V 非空, 任取 $\mathbf{x} \in V$, 则对 $k=0$, 可知 $0\mathbf{x} \in V$, 即线性子空间总包含零向量. 另外, 对 V 中任何有限多个元素 $\mathbf{x}_1, \cdots, \mathbf{x}_m$, 以及任何实数 $k_1, \cdots, k_m$, 有 $k_1\mathbf{x}_1 + \cdots + k_m\mathbf{x}_m \in V$, 即 V 关于有限线性组合是封闭的.

定义 2.1.2. 设 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 是 $\mathbf{R}^n$ 中的一组向量, 令

$$\operatorname{span}(S) = \operatorname{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_m\} = \{k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m | k_1, \cdots, k_m \in \mathbf{R}\},$$

易验证 $\operatorname{span}(S)$ 是线性子空间, 称之为由 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 生成的子空间.

这就给出了构造子空间的办法: 任取一个向量组 S, 它生成一个线性子空间 $\operatorname{span}(S)$. 很快我们会在定理2.1.12中证明, $\mathbf{R}^n$ 的所有线性子空间都是这种形式.

$\operatorname{span}(S)$ 中的每个向量 $\mathbf{x}$ 都可以表示成 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 的线性组合 $\mathbf{x} = k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m$, 这种表示是否唯一?

定义 2.1.3. 称 $\mathbf{R}^n$ 中的一组向量 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 是线性相关的, 如果存在不全为零的 $k_1, \cdots, k_m \in \mathbf{R}$, 使得 $k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$; 等价的说, 即 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 中存在一个向量可以表示成其余向量的线性组合.

如果 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 不是线性相关的, 则称它们线性无关; 等价的说, 即 $k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$ 当且仅当 $k_1, \cdots, k_m$ 全为零. 约定空向量组为线性无关的.

例 2.1.4. 单个向量的组 $\{\mathbf{v}\}$ 线性相关当且仅当 $\mathbf{v} = \mathbf{0}$; 两个向量的组 $\{\mathbf{a}, \mathbf{b}\}$ 线性相关当且仅当这两个向量成比例, 即有 $\mathbf{a} = k\mathbf{b}$ 或 $\mathbf{b} = k\mathbf{a}$, 其中 k 是适当的实数.

命题 2.1.5. *设 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 是 $\mathbf{R}^n$ 中的向量组, 以它们为列向量构成矩阵 $A = (\mathbf{a}_1, \cdots, \mathbf{a}_m)$, 则 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 线性相关的充分必要条件是齐次方程 $A\mathbf{x} = \mathbf{0}$ 有非零解; $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 线性无关的充分必要条件是: 将 A 用 Gauss 消元法化成阶梯形矩阵后所得矩阵的阶梯数等于 m.*

假设 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 是线性相关的向量组, 我们来证明可以找到 S 的一个线性无关的子集 T, 使得 T 就能生成 $\text{span}(S)$.

定义 2.1.6. 设 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 是 $\mathbf{R}^n$ 的一个向量组, T 是 S 的子集, 称 T 为 S 的极大线性无关子集, 如果 T 是线性无关的, 且不存在 S 的线性无关子集真包含 T.

命题 2.1.7. *设 S 是 $\mathbf{R}^n$ 的一个有限子集, T 为 S 的极大线性无关子集, 则有 $span(T) = span(S)$.*

证明: 设 $T = \{\mathbf{b}_1, \cdots, \mathbf{b}_l\}$, 对任何 $\mathbf{a} \in S \backslash T$, 由 T 是极大线性无关子集的定义可知 $T \cup \{\mathbf{a}\}$ 线性相关, 即存在不全为零的实数 $k_1, \cdots, k_l, \lambda$, 使得

$$k_1\mathbf{b}_1 + \cdots + k_l\mathbf{b}_l + \lambda\mathbf{a} = \mathbf{0}.$$

由 T 线性无关可知 $\lambda \neq 0$, 从而有

$$\mathbf{a} = \sum_{i=1}^{l}(-\frac{k_i}{\lambda})\mathbf{b}_i,$$

即 $S \backslash T$ 中每个向量都可以表示成 $\mathbf{b}_1, \cdots, \mathbf{b}_l$ 的线性组合. 由此可得 $\text{span}(T) = \text{span}(S)$. □

设 S 是有限向量组, 考虑它的所有子集, 其中必有线性无关的子集 (例如空子集). 设 T 是 S 的元素个数最大的线性无关子集, 则 T 是 S 的极大线性无关子集, 且由上述命题可知 $\text{span}(T) = \text{span}(S)$. 这样, 在考虑 S 生成的线性子空间时, 可以选用更小的集合 T 生成这个线性子空间.

定义 2.1.8. 设 V 是 $\mathbf{R}^n$ 的线性子空间, 称向量组 $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 为子空间 V 的一个基底 (basis), 如果它们线性无关且生成 V; 等价的说, 即 V 中的每个向量都可以唯一的表示成 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 的线性组合.

约定零子空间的基底为空集.

例 2.1.9. 设 $\mathbf{e}_i$ 是 $\mathbf{R}^n$ 的第 i 个单位坐标向量, 它的第 i 个分量等于 1, 其余分量为零, 则所有 n 个坐标向量 $\mathbf{e}_1, \cdots, \mathbf{e}_n$ 构成 $\mathbf{R}^n$ 的一个基底, 称之为 $\mathbf{R}^n$ 的标准基底.

命题 2.1.10. *设 $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 是子空间 V 的基底, 则由 V 中多于 m 个向量构成的向量组一定线性相关.*

证明: 设 $\{\mathbf{b}_1, \cdots, \mathbf{b}_l\}$ 是 V 中的一组向量, 且 $l > m$. 由基底的定义, 每个 $\mathbf{b}_i$ 可以唯一的表示成 $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 的线性组合 $\mathbf{b}_i = \sum_{j=1}^{m} c_{ij}\mathbf{a}_j$. 考虑方程 $\sum_{i=1}^{l} k_i\mathbf{b}_i = \mathbf{0}$ 的解, 它等价于

$$\mathbf{0} = \sum_{i=1}^{l} k_i \sum_{j=1}^{m} c_{ij}\mathbf{a}_j = \sum_{j=1}^{m} \left(\sum_{i=1}^{l} c_{ij}k_i\right)\mathbf{a}_j \iff \sum_{i=1}^{l} c_{ij}k_i = 0, \quad \forall 1 \le j \le m.$$

注意到, 上式右边是关于 l 个未知元的线性方程组, 且方程个数 m 小于未知元个数, 由命题1.6.2可知上述线性方程组有非零解, 从而证明了 $\{\mathbf{b}_1, \cdots, \mathbf{b}_l\}$ 线性相关. □

推论 2.1.11. *设* $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ *与* $\{\mathbf{b}_1, \cdots, \mathbf{b}_l\}$ *都是同一个子空间* V *的基底, 则* $m = l$.

证明: 由命题2.1.10可得 $l \le m$ 且 $m \le l$. □

定理 2.1.12 (基底存在性)**.** *设* V *是* $\mathbf{R}^n$ *的线性子空间, 则* V *存在一个基底.*

证明: 若 V 为零子空间, 则空集为它的一个基底. 以下假设 V 不是零子空间, 我们按如下方式归纳的构造向量组 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$, 使得 S 线性无关且生成 V, 因而是 V 的一个基底. 首先任取 $\mathbf{a}_1 \in V\backslash\{\mathbf{0}\}$. 假设已经构造好 $\mathbf{a}_1, \cdots, \mathbf{a}_t$, 如果 $\mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_t\} = V$, 则构造中止并令 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_t\}$; 如果 $\mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_t\} \subsetneqq V$, 则任取 $\mathbf{a}_{t+1} \in V\backslash\mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_t\}$. 这就给出了整个归纳构造过程. 注意到, 由于每一步有 $\mathbf{a}_{t+1} \notin \mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_t\}$, 则对每个 $k \ge 1$, 向量组 $\mathbf{a}_1, \cdots, \mathbf{a}_t\}$ 都是线性无关的.

上述构造一定会终止, 不然的话我们将得到线性无关的向量组 $\mathbf{a}_1, \cdots, \mathbf{a}_{n+1}$, 这与命题2.1.10矛盾! 设构造中止时 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$, 则 S 线性无关且 $\mathrm{span}(S) = V$, 这就完成了证明. □

上述证明过程中, 是从空集出发, 不断的加入已有向量组的 span 之外的 V 中的新向量, 直至得到 V 的一个基底. 完全类似的, 我们也可以从 V 的一个线性无关集 S_0 出发, 不断的加入已有向量组的 span 之外的 V 中的新向量, 直至得到 V 的一个基底. 这就得到如下的定理.

定理 2.1.13 (线性无关组可扩充成基底)**.** *设* V *是* $\mathbf{R}^n$ *的线性子空间,* S *是* V *的线性无关子集, 则可将* S *扩充为* V *的一个基底.*

定理 2.1.14 (基底扩充)**.** *设* V, W *都是* $\mathbf{R}^n$ *的线性子空间且* $V \subset W$*, 则* V *的任何基底都可以扩充成* W *的一个基底.*

证明: V 的基底是 W 的一个线性无关集, 对它用定理2.1.13的结论即可. □

定义 2.1.15. 设 V 是 $\mathbf{R}^n$ 的线性子空间, 定义 V 的维数为 V 的任意一组基底中向量的个数, 记为 $\dim(V)$.

命题 2.1.16. 设 V, W 都是 $\mathbf{R}^n$ 的线性子空间且 $V \subset W$, 则有 $dim(V) \leq dim(W)$. 进一步, 如果 $V \subset W$ 且 $dim(V) = dim(W)$, 则 $V = W$.

证明: 利用定理2.1.14即可. □

设 S 是 $\mathbf{R}^n$ 的有限向量组, 由命题2.1.7 及其后面的一段论述, 可知 S 有极大线性无关子集 T, 且 $\mathrm{span}(T) = \mathrm{span}(S)$. 这样, T 就是 $\mathrm{span}(S)$ 的一个基底, 由此可知 $|T| = \dim\,(\mathrm{span}(S))$.

定义 2.1.17. 设 S 是 $\mathbf{R}^n$ 的有限向量组, 定义 S 的秩为 S 的极大线性无关子集的元素个数, 记为 $\mathrm{rank}(S)$. 由前述, S 的秩等于 S 生成的线性子空间的维数.

定义 2.1.18. 设 S, T 是 $\mathbf{R}^n$ 的有限向量组, 称 S 可用 T 线性表示, 如果 S 中的每个向量都可以表示成 T 中向量的线性组合; 等价的说, 即是 $\mathrm{span}(S) \subseteq \mathrm{span}(T)$.

称有限向量组 S 与 T 线性等价, 如果 S 可被 T 线性表示, 且 T 可被 S 线性表示; 等价的说, 即是 $\mathrm{span}(S) = \mathrm{span}(T)$.

命题 2.1.19. 设 S, T 是 $\mathbf{R}^n$ 的有限向量组.

(1) 如果 S 可用 T 线性表示, 则 $rank(S) \leq rank(T)$.

(2) 如果 S 与 T 线性等价, 则 $rank(S) = rank(T)$.

证明: 利用命题2.1.16即可. □

2.2 线性映射的像与核

线性映射的像集与核是人们最关心的线性子空间.

定义 2.2.1. 设 $f : \mathbf{R}^n \to \mathbf{R}^m$ 是线性映射, 定义 f 的像集为

$$\mathrm{Im}(f) = \mathcal{R}(f) = \{\mathbf{y} \in \mathbf{R}^m | \exists \mathbf{x} \in \mathbf{R}^n \text{使得} \mathbf{y} = f(\mathbf{x})\}.$$

定义 f 的核 (kenerl) 或零空间为

$$\ker(f) = \mathcal{N}(f) = \{\mathbf{x} \in \mathbf{R}^n | f(\mathbf{x}) = \mathbf{0}\}.$$

易验证, $\mathrm{Im}(f)$ 是 $\mathbf{R}^m$ 的线性子空间, $\ker(f)$ 是 $\mathbf{R}^n$ 的线性子空间.

这样, 线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$ 是满射当且仅当 $\mathrm{Im}(f) = \mathbf{R}^m$; f 是单射当且仅当 $\ker(f) = \{\mathbf{0}\}$. 对于两个线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$ 与 $g : \mathbf{R}^m \to \mathbf{R}^l$, 它们的复合满足 $g \circ f = 0$ 当且仅当 $\mathrm{Im}(f) \subset \ker(g)$.

定理 2.2.2. 设 $f: V \to W$ 是线性映射, 其中 $V \subset \mathbf{R}^n, W \subset \mathbf{R}^m$ 是线性子空间, 则有

$$dim\ ker(f) + dim\ Im(f) = dimV.$$

特别的, $dim\ Im(f) \leq dimV$, 即线性映射像集的维数不超过定义域的维数.

证明: 由定理2.1.12, 可取出 $\ker(f)$ 的一组基底 $\mathbf{a}_1, \cdots, \mathbf{a}_s$; 再由定理2.1.14可把它扩充成 V 的基底 $\mathbf{a}_1, \cdots, \mathbf{a}_s, \mathbf{a}_{s+1}, \cdots, \mathbf{a}_{s+t}$. 由维数的定义, 有 $\dim \ker(f) = s$, $\dim V = s + t$. 我们只需证明 $\dim \mathrm{Im}(f) = t$.

一方面, 注意到 $\mathrm{Im}(f)$ 中的每个向量都形如

$$f(\sum_{i=1}^{s+t} k_i \mathbf{a}_i) = \sum_{i=1}^{s+t} k_i f(\mathbf{a}_i) = \sum_{i=s+1}^{s+t} k_i f(\mathbf{a}_i),$$

可知 $f(\mathbf{a}_{s+1}), \cdots, f(\mathbf{a}_{s+t})$ 生成 $\mathrm{Im}(f)$.

另一方面, $f(\mathbf{a}_{s+1}), \cdots, f(\mathbf{a}_{s+t})$ 是线性无关的. 因为若有 $\sum\limits_{i=s+1}^{s+t} k_i f(\mathbf{a}_i) = \mathbf{0}$, 则有 $\sum\limits_{i=s+1}^{s+t} k_i \mathbf{a}_i \in \ker(f) = \mathrm{span}(\mathbf{a}_1, \cdots, \mathbf{a}_s)$, 从而存在实数 $k_1, \cdots, k_s$ 使得

$$\sum_{i=s+1}^{s+t} k_i \mathbf{a}_i = \sum_{i=1}^{s} k_i \mathbf{a}_i.$$

注意到 $\mathbf{a}_1, \cdots, \mathbf{a}_s, \mathbf{a}_{s+1}, \cdots, \mathbf{a}_{s+t}$ 线性无关, 上式表明 $k_1, \cdots, k_{s+t}$ 都等于零, 特别的 $k_{s+1}, \cdots, k_{s+t}$ 都等于零. 这就证明了 $f(\mathbf{a}_{s+1}), \cdots, f(\mathbf{a}_{s+t})$ 线性无关.

结合以上两方面, 可知 $f(\mathbf{a}_{s+1}), \cdots, f(\mathbf{a}_{s+t})$ 是 $\mathrm{Im}(f)$ 的基底, 从而有 $\dim \mathrm{Im}(f) = t$. □

推论 2.2.3. 设 $f: \mathbf{R}^n \to \mathbf{R}^n$ 是线性映射, 则

$$f\text{是单射} \iff f\text{是满射}.$$

证明: 利用定理2.2.2与命题2.1.16, 可得

$$f\text{是单射} \iff \dim \ker(f) = 0 \iff \dim \mathrm{Im}(f) = n \iff f\text{是满射}.$$

□

可以用此推论给出命题1.8.9的另一个证明: 设 A, B 给出的线性映射分别为 f, g, 则 $AB = I_n$ 表明 $f \circ g = Id_{\mathbf{R}^n}$, 由此可知 f 是满射, g 是单射, 再利用推论2.2.3可得 f 与 g 都是可逆线性映射, 因而它们的表示矩阵都是可逆矩阵.

我们把可逆的线性映射称为线性同构.

定义 2.2.4. 设 $f : V \to W$ 是线性映射, 其中 $V \subset \mathbf{R}^n, W \subset \mathbf{R}^m$ 是线性子空间, 称 f 为线性同构, 如果存在线性映射 $g : W \to V$, 使得 $g \circ f = \mathrm{Id}_V$ 且 $f \circ g = \mathrm{Id}_W$. 容易验证, f 是线性同构当且仅当 f 是既单又满的线性映射.

称线性子空间 V 与 W 是线性同构的, 记为 $V \cong W$, 如果存在一个从 V 到 W 的线性同构.

命题 2.2.5. *设 $f : V \to W$ 是线性同构, 则 $dimV = dimW$.*

证明: 由 f 单射, 有 $\ker(f) = \{\mathbf{0}\}$; 由由 f 满射, 有 $\mathrm{Im}(f) = W$. 利用定理2.2.2即得

$$\mathrm{dim}V = \mathrm{dim}\ \ker(f) + \mathrm{dim}\ \mathrm{Im}(f) = \mathrm{dim}W.$$

□

命题 2.2.6. *$V \cong W$ 的充分必要条件是 $dimV = dimW$.*

证明: 必要性由命题2.2.5可得. 只需再证明充分性. 设 $\mathrm{dim}V = \mathrm{dim}W = n$, 由定理2.1.12可知 V 有一个基底 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$; 类似的, W 有一个基底 $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$. 定义线性映射 $f : V \to W$ 为

$$f(\sum_{i=1}^{n} k_i \mathbf{v}_i) = \sum_{i=1}^{n} k_i \mathbf{w}_i, \quad \forall k_1, \cdots, k_n \in \mathbf{R},$$

易验证 f 既单又满, 因而给出 V 与 W 的一个线性同构. □

命题 2.2.7. *线性映射 $f : \mathbf{R}^n \to \mathbf{R}^n$ 是线性同构的充分必要条件为 f 的表示矩阵是可逆矩阵.*

证明: 注意到, 若 $f, g : \mathbf{R}^n \to \mathbf{R}^n$ 的表示矩阵分别为 A, B, 则

$$g \circ f = \mathrm{Id}_{\mathbf{R}^n} \iff BA = I_n.$$

□

2.3 线性方程组解的结构

定理 2.3.1. *设 $f : \mathbf{R}^n \to \mathbf{R}^m$ 是线性映射, $\mathbf{b} \in \mathbf{R}^m$, $\mathbf{x}_0 \in f^{-1}(\{\mathbf{b}\})$, 则*

$$f^{-1}(\{\mathbf{b}\}) = \mathbf{x}_0 + ker(f),$$

其中, 对于 $\mathbf{a} \in \mathbf{R}^n$, $V \subset \mathbf{R}^n$, 约定 $\mathbf{a} + V = \{\mathbf{a} + \mathbf{v} | \mathbf{v} \in V\}$ 为把 V 平移 $\mathbf{a}$ 所得的子集.

证明*:* 这是因为

$$\mathbf{x} \in f^{-1}(\{\mathbf{b}\}) \iff f(\mathbf{x}) = f(\mathbf{x}_0) \iff f(\mathbf{x} - \mathbf{x}_0) = 0 \iff \mathbf{x} - \mathbf{x}_0 \in \ker(f).$$

□

可用方程的语言叙述上述结论. 考虑方程 $A\mathbf{x} = \mathbf{b}$ 的解, 假设已经找到一个特解 $\mathbf{x}_0$, 则上述定理说 $A\mathbf{x} = \mathbf{b}$ 的所有解等于特解 $\mathbf{x}_0$ 加上齐次方程 $A\mathbf{x} = \mathbf{0}$ 的一般解. 人们把此结论称为线性方程解的结构定理.

下面我们来描述齐次方程 $A\mathbf{x} = \mathbf{0}$ 的解集, 即对应的线性映射 f 的核. 假设 A 可以通过初等行变换化成简化阶梯形矩阵 B, 则 $A\mathbf{x} = \mathbf{0}$ 与 $B\mathbf{x} = \mathbf{0}$ 的解集相等, 对此现在有直接的看法: 存在初等矩阵 T 使得 $TA = B$, 由 T 可逆可得

$$A\mathbf{x} = \mathbf{0} \iff T(A\mathbf{x}) = \mathbf{0} \iff B\mathbf{x} = \mathbf{0}.$$

设简化阶梯形矩阵 B 共有 r 个阶梯行, 第 i 个阶梯行为

$$0, \cdots, 0, b_{ij(i)}, \cdots, b_{in},$$

其中 $j(1) < j(2) < \cdots < j(r)$, 每个 $b_{ij(i)} = 1$, 且第 $j(i)$ 列只有 $b_{ij(i)}$ 唯一一个非零元. 以前我们把 $x_{j(1)}, \cdots, x_{j(r)}$ 称为主变量, 其他 x_k 称为自由变量. 对于自由变量的任何一组取值, 它们唯一的确定 $B\mathbf{x} = \mathbf{0}$ 的一个解: 每个主变量由下式给出

$$x_{j(i)} = -\sum_{k > j(i)} b_{ik} x_k, \quad \forall 1 \le i \le r, \tag{2.1}$$

即可把上述解写成

$$\mathbf{x} = \sum_{\text{自由变量} x_k} x_k \mathbf{e}_k + \sum_{i=1}^{r} \left(-\sum_{k > j(i)} b_{ik} x_k \right) \mathbf{e}_{j(i)}. \tag{2.2}$$

对每个 $k \ne j(1), \cdots, j(r)$, 考虑只有一个非零自由变量的解

$$\mathbf{s}_k = \mathbf{e}_k + \sum_{j(i) < k} -b_{ik} \mathbf{e}_{j(i)},$$

则由 (2.2) 式可知齐次方程的每一个解 $\mathbf{x}$ 都可以唯一的表示为上述这些基本解 $\mathbf{s}_k$ 的线性组合

$$\mathbf{x} = \sum_{\text{自由变量} x_k} x_k \mathbf{s}_k,$$

这表明基本解 $\{\mathbf{s}_k\}$ 是方程 $B\mathbf{x} = \mathbf{0}$ 解集的一个基底. 这就得到如下定理.

定理 2.3.2. 设线性映射 $f:\mathbf{R}^n\to\mathbf{R}^m$ 的表示矩阵为 A, 利用初等行变化把 A 化成阶梯行矩阵后所得矩阵的阶梯数为 r, 则有 $dim\ ker(f)=n-r$, $dim\ Im(f)=r$.

证明*:* 令 $K=\{1,2,\cdots,n\}\backslash\{j(1),j(2),\cdots,j(r)\}$, 把所有自由变量的集合记为 $\{x_k\}_{k\in K}$. 由 (2.1) 式, 可将主变量 $x_{j(i)}$ 用自由变量表示, 将此表示简记为

$$x_{j(i)}=\varphi_i(\{x_k\}_{k\in K}),$$

其中 φ_i 是自由变量 $\{x_k\}_{k\in K}$ 的线性函数.

令 $V=\{\sum\limits_{k\in K}x_i\mathbf{e}_k|x_k\in\mathbf{R}\}$, 它是 $|K|=n-r$ 维线性空间. 定义线性映射 $\Phi:V\to\ker(f)$ 为

$$\Phi(\sum_{k\in K}x_k\mathbf{e}_k)=\sum_{k\in K}x_k\mathbf{e}_k+\sum_{i=1}^{r}\varphi_i(\{x_k\}_{k\in K})\mathbf{e}_{j(i)}.$$

显然 Φ 是既单又满的线性映射, 因而是从 V 到 $\ker(f)$ 之间的线性同构. 利用命题2.2.5即得 $\mathrm{dim}\ker(f)=\mathrm{dim}V=n-r$. □

2.4 秩

为了更好的叙述定理2.3.2, 我们用矩阵语言来表述线性映射的像与核. 设 f 的表示矩阵为 A, 则 $\ker(f)$ 等于方程 $A\mathbf{x}=\mathbf{0}$ 的解集.

命题 2.4.1. 设 $f:\mathbf{R}^n\to\mathbf{R}^m$ 是线性映射, 则有

$$Im(f)=span\{f(\mathbf{e}_1),\cdots,f(\mathbf{e}_n)\}.$$

如果 f 的表示矩阵为 A, 则 f 的像集等于 A 的所有列向量生成的线性子空间.

证明*:* 这是直接的, 有

$$\begin{aligned}\mathbf{y}\in\mathrm{Im}(f)&\iff\exists k_1,\cdots,k_n\in\mathbf{R}\text{使得}\mathbf{y}=f(\sum_{i=1}^{n}k_i\mathbf{e}_i)=\sum_{i=1}^{n}k_if(\mathbf{e}_i)\\&\iff\mathbf{y}\in\mathrm{span}\{f(\mathbf{e}_1),\cdots,f(\mathbf{e}_n)\}.\end{aligned}$$

□

定义 2.4.2. 设 A 是 $m\times n$ 矩阵. 称它的所有列向量生成的线性子空间, 为 A 的列向量空间, 记为 $\mathcal{R}(A)$. 称方程 $A\mathbf{x}=\mathbf{0}$ 的解集为 A 的零空间, 记为 $\mathcal{N}(A)$.

设 A 确定的线性映射 $f_A : \mathbf{R}^n \to \mathbf{R}^m$ 为

$$f_A(\mathbf{x}) = \mathbf{A}\mathbf{x}, \quad \forall \mathbf{x} \in \mathbf{R}^\mathbf{n},$$

则有 $\mathcal{R}(A) = \mathrm{Im}(f_A)$, $\mathcal{N}(A) = \ker(f_A)$. 由定理2.2.2可知, $\dim\mathcal{N}(A) + \dim\mathcal{R}(A) = n$.

定义 2.4.3. 定义 $m \times n$ 矩阵 A 的秩如下四个相等的数, 记为 $\mathrm{rank}(A)$.

(1) f_A 的像集的维数;

(2) A 的列向量空间的维数;

(3) A 的列向量组的秩;

(4) A 的列向量组的极大线性无关子集的元素个数.

由此定义, 可将定理2.3.2按如下方式重新叙述.

命题 2.4.4. *rank(A) 等于用初等行变换将 A 化成阶梯形矩阵所得矩阵的阶梯数.*

上述结果给出了计算矩阵秩的具体办法: 将矩阵 A 用初等行变换化成阶梯形矩阵, 则最后所得阶梯形矩阵的阶梯数就是 A 的秩. 此前我们是按如下方式证明此结果的: 利用核与像的维数关系, 将 $\mathrm{rank}(A)$ 的计算转化为方程 $A\mathbf{x} = \mathbf{0}$ 解空间的维数; 由于初等行变换保持方程同解, 这样只需对阶梯形矩阵写出其零空间.

实际上, 我们可以清楚的描述矩阵的列向量空间在初等行变换下的变化方式. 用 $\mathcal{R}(A)$ 表示 A 的列向量空间, 则 $\mathcal{R}(A^T)$ 表示 A 的行向量空间. 我们也把 A 的行向量空间记为 $\mathrm{Row}(A)$.

命题 2.4.5. *设矩阵 A 经过初等行变换 T 变为矩阵 $B = TA$, 则有*

(1) $\mathcal{R}(B)$ 等于 $\mathcal{R}(A)$ 在线性同构 $T : \mathbf{R}^m \to \mathbf{R}^m$ 下的像集. 特别的, $\mathcal{R}(B)$ 与 $\mathcal{R}(A)$ 维数相等, 即 $rank(B) = rank(A)$.

(2) $Row(B) = Row(A)$. 特别的, $rank(B^T) = rank(A^T)$

定理 2.4.6 (行秩等于列秩)**.** *设 A 是 $m \times n$ 矩阵, 则 $rank(A) = rank(A^T)$.*

证明: 设 A 可用初等行变换化成简化行阶梯形矩阵 B. 由命题2.4.5可知 $\mathrm{rank}(A) = \mathrm{rank}(B)$, $\mathrm{rank}(A^T) = \mathrm{rank}(B^T)$. 设 B 的阶梯数为 r, 则由命题2.4.4, 只需验证 $\mathrm{rank}(B^T) = r$, 即 B 的行向量的秩等于 B 的行阶梯数. 这是显然的, 因为 B 的所有 r 个阶梯行是 $\mathrm{Row}(B)$ 的一个基底. □

定义 2.4.7. 设 A, B 都是 $m \times n$ 矩阵, 称 A 与 B 左相抵, 如果可从 A 经过若干次初等行变换化为 B.

类似的, 称 A 与 B 右相抵, 如果可从 A 经过若干次初等列变换化为 B.

称 A 与 B 相抵, 如果可从 A 经过若干次初等行变换和初等列变换化为 B.

注意, 左相抵与右相抵都强于相抵.

命题 2.4.8. 设 A, B 都是 $m\times n$ 矩阵, 则有

(1) A 与 B 左相抵的充分必要条件是存在 m 阶可逆矩阵 P 使得 $PA=B$;

(2) A 与 B 右相抵的充分必要条件是存在 n 阶可逆矩阵 Q 使得 $AQ=B$;

(3) A 与 B 相抵的充分必要条件是存在 m 阶可逆矩阵 P 与 n 阶可逆矩阵 Q 使得 $PAQ=B$.

证明: 利用定理1.8.6即可. □

命题 2.4.9. 设 A 与 B 相抵, 则 $rank(A)=rank(B)$.

证明: 命题2.4.5说初等行变换保持矩阵的秩不变. 只需证明初等列变换也保持秩不变. 设用初等列变换将 A 变成 B, 则可用初等行变换将 A^T 变成 B^T, 利用命题2.4.5中 (2) 的结论, 有 $\mathrm{rank}((B^T)^T)=\mathrm{rank}((A^T)^T)$, 此即 $\mathrm{rank}(B)=\mathrm{rank}(A)$. □

命题 2.4.10. 设 $f:\mathbf{R}^n\to\mathbf{R}^m$ 是线性映射, $dimIm(f)=r$, 则存在线性同构 $\alpha:\mathbf{R}^n\to\mathbf{R}^n$ 与 $\beta:\mathbf{R}^m\to\mathbf{R}^m$, 使得复合映射 $g=\beta\circ f\circ\alpha:\mathbf{R}^n\to\mathbf{R}^m$ 的表示矩阵为分块矩阵 $\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$.

证明: 在定理2.2.2的证明中, 我们证明了如下结论 (这里稍做改写): 存在 $\ker(f)$ 的基底 $\{\mathbf{a}_{r+1},\cdots,\mathbf{a}_n\}$, 把它扩充为 $\mathbf{R}^n$ 的基底 $\{\mathbf{a}_1,\cdots,\mathbf{a}_r,\mathbf{a}_{r+1},\cdots,\mathbf{a}_n\}$, 则 $\{f(\mathbf{a}_1),\cdots,f(\mathbf{a}_r)\}$ 是 $\mathrm{Im}(f)$ 的基底. 利用定理2.1.14 可把后者扩充为 $\mathbf{R}^m$ 的基底 $\{f(\mathbf{a}_1),\cdots,f(\mathbf{a}_r);\mathbf{b}_{r+1},\cdots,\mathbf{b}_m\}$.

定义 $\alpha:\mathbf{R}^n\to\mathbf{R}^n$ 为

$$\alpha(x_1,\cdots,x_n)=\sum_{i=1}^n x_i\mathbf{a}_i.$$

定义 $\gamma:\mathbf{R}^m\to\mathbf{R}^m$ 为

$$\gamma(y_1,\cdots,y_m)=\sum_{j=1}^r y_jf(\mathbf{a}_j)+\sum_{j=r+1}^m y_j\mathbf{b}_j.$$

显然 α,β 都是线性同构, 令 $\beta=\gamma^{-1}$, 易验证 $g=\beta\circ f\circ\alpha$ 的表示矩阵即为命题中所述的分块矩阵. □

设 f,α,β 的表示矩阵分别为 A,Q,P, 则可把上述命题叙述成如下的命题.

命题 2.4.11. 设 $m\times n$ 矩阵的秩为 r, 则存在 m 阶可逆矩阵 P 与 n 阶可逆矩阵 Q, 使得 $PAQ=\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$.

推论 2.4.12. 设 A,B 都是 $m\times n$ 矩阵, 则 A 与 B 相抵的充分必要条件是 $rank(A)=rank(B)$.

命题 2.4.13. 设 A, B 分别是 $l \times m$ 与 $m \times n$ 矩阵, 则有

$$rank(AB) \le \min\{rank(A), rank(B)\},$$

即矩阵乘法不增加秩.

证明: 只需注意到, 对于线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$ 与 $g : \mathbf{R}^m \to \mathbf{R}^l$, 有 $\mathrm{Im}(g \circ f) \subset \mathrm{Im}(g)$, 且 $\mathrm{Im}(g \circ f)$ 等于 $\mathrm{Im}(f)$ 在线性映射 $g : \mathrm{Im}(f) \to \mathbf{R}^l$ 下的像. □

定义 2.4.14. 设 A 是 $m \times n$ 矩阵, 称 A 是行满秩, 如果 $\mathrm{rank}(A)$ 等于 A 的行数; 称 A 是列满秩, 如果 $\mathrm{rank}(A)$ 等于 A 的列数.

称 A 是满秩矩阵, 如果 A 既是行满秩又是列满秩, 即有 $\mathrm{rank}(A) = m = n$.

命题 2.4.15. 设线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$ 的表示矩阵为 A, 则有

(1) f 是满射当且仅当 A 是行满秩矩阵;

(2) f 是单射当且仅当 A 是列满秩矩阵;

(3) f 是线性同构当且仅当 A 是满秩矩阵.

证明: 利用定理2.2.2即可. □

例 2.4.16. (1) 设 A, B 是同规模的矩阵, 则有 $\mathrm{rank}(A+B) \le \mathrm{rank}(A) + \mathrm{rank}(B)$.

(2) 设 C, D 是行数相同的矩阵, 则有

$$\max\{\mathrm{rank}(C), \mathrm{rank}(D)\} \le \mathrm{rank}(C|D) \le \mathrm{rank}(C) + \mathrm{rank}(D).$$

例 2.4.17. 反对称矩阵的秩是偶数.

证明: 设 A 是 n 阶反对称实矩阵, 将它给出的映射记为 $A : \mathbf{R}^n \to \mathbf{R}^n$, 令 $V = \mathrm{Im}(A)$, 则 $\mathrm{rank}(A) = \mathrm{dim} V$.

(1) $V \cap \ker(A) = \{\mathbf{0}\}$. 这是由于, 若 $\mathbf{x} \in V \cap \ker(A)$, 则存在 $\mathbf{y}$ 使得 $x = A\mathbf{y}$, 从而可得

$$\langle \mathbf{x}, \mathbf{x} \rangle = \langle A\mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{y}, A^T\mathbf{x} \rangle = \langle \mathbf{y}, -A\mathbf{x} \rangle = 0,$$

故 $\mathbf{x} = \mathbf{0}$.

(2) 考虑 $A : V \to V$, 由 (1) 的结论可知 A 作为从 V 到 V 的线性映射是单射, 因为是线性同构. 选取 V 的一个单位正交基底 $\{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$, 在此基底下设线性同构 $A : V \to V$ 的表示矩阵为 $M_{r \times r}$, 则 $\det M \ne 0$. 注意到

$$A\mathbf{e}_j = \sum_{i=1}^{r} M_{ij} \mathbf{e_i},$$

则有 $M_{ij} = \langle A\mathbf{e}_j, \mathbf{e}_i \rangle$. 利用 A 是反对称的可知

$$M_{ij} = \langle A\mathbf{e}_j, \mathbf{e}_i \rangle = \langle \mathbf{e}_j, A^T\mathbf{e}_i \rangle = \langle \mathbf{e}_j, -A\mathbf{e}_i \rangle = -\langle A\mathbf{e}_i, \mathbf{e}_j \rangle = -M_{ji},$$

即有 $M^T = -M$. 这样就有

$$\det M = \det M^T = \det(-M) = (-1)^r \det M,$$

结合 $\det M \neq 0$ 可知 r 是偶数. □

第三章　内积与正交性

3.1　内积空间

定义 3.1.1. 设 $\mathbf{a} = (a_1, \cdots, a_n), \mathbf{b} = (b_1, \cdots, b_n)$ 是 $\mathbf{R}^n$ 中的向量, 定义它们的内积为

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{n} a_i b_i.$$

如果将 $\mathbf{a}, \mathbf{b}$ 视为列向量, 则也可把内积用矩阵乘法表示

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}.$$

向量内积满足如下性质:

(1) 内积是双线性函数, $\langle \mathbf{a}, \mathbf{b} \rangle$ 关于 $\mathbf{a}$ 与 $\mathbf{b}$ 都是线性的, 即对任何 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2 \in \mathbf{R}^n$ 与任何 $k_1, k_2 \in \mathbf{R}$ 有

$$\langle k_1\mathbf{a}_1 + k_2\mathbf{a}_2, \mathbf{b} \rangle = k_1\langle \mathbf{a}_1, \mathbf{b} \rangle + k_2\langle \mathbf{a}_2, \mathbf{b} \rangle, \quad \langle \mathbf{a}, k_1\mathbf{b}_1 + k_2\mathbf{b}_2 \rangle = k_1\langle \mathbf{a}, \mathbf{b}_1 \rangle + k_2\langle \mathbf{a}, \mathbf{b}_2 \rangle.$$

(2) 对称性, 对任何 $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$ 有 $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$.

(3) 正定性, 对任何 $\mathbf{a} \in \mathbf{R}^n$ 有 $\langle \mathbf{a}, \mathbf{a} \rangle \geq 0$, 且 $\langle \mathbf{a}, \mathbf{a} \rangle = 0$ 当且仅当 $\mathbf{a} = \mathbf{0}$.

利用上述三条性质, 可定义一般的内积空间.

定义 3.1.2. 设 V 是线性空间, 所谓 V 上的一个内积是指一个对称的, 正定的双线性映射 $\langle \cdot, \cdot \rangle : V \times V \to \mathbf{R}$. 称赋予了内积 $\langle \cdot, \cdot \rangle$ 的线性空间 V 为一个内积空间.

命题 3.1.3 (Cauchy-Schwarz 不等式)**.** *设 V 是内积空间, 则对任何 $\mathbf{a}, \mathbf{b} \in V$ 有*

$$(\langle \mathbf{a}, \mathbf{b} \rangle)^2 \leq \langle \mathbf{a}, \mathbf{a} \rangle \cdot \langle \mathbf{b}, \mathbf{b} \rangle,$$

等号成立当且仅当 $\mathbf{a}, \mathbf{b}$ 线性相关.

证明: 只需考虑 $\mathbf{a} \neq \mathbf{0}$ 的情形. 对任何实数 t, 有

$$0 \leq \langle t\mathbf{a} + \mathbf{b}, t\mathbf{a} + \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{a} \rangle t^2 + 2\langle \mathbf{a}, \mathbf{b} \rangle t + \langle \mathbf{b}, \mathbf{b} \rangle.$$

将上式右边视为 t 的二次函数, 它是开口向上的, 且取值非负, 则其判别式小于等于零, 即有

$$\Delta = (2\langle \mathbf{a}, \mathbf{b} \rangle)^2 - 4\langle \mathbf{a}, \mathbf{a} \rangle \cdot \langle \mathbf{b}, \mathbf{b} \rangle \leq 0,$$

这就是 Cauchy-Schwarz 不等式. □

利用 Cauchy-Schwarz 不等式, 可定义 V 中两个向量 $\mathbf{a}, \mathbf{b}$ 的距离为

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{\langle \mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{b} \rangle}.$$

特别的, 记

$$||\mathbf{a}|| = d(\mathbf{a}, \mathbf{0}) = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle},$$

称为 $\mathbf{a}$ 的长度或模长.

命题 3.1.4 (三角不等式). 设 $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$, 则有

$$d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}) \geq d(\mathbf{a}, \mathbf{c}).$$

证明: 记 $\mathbf{a} - \mathbf{b} = \mathbf{x}$, $\mathbf{b} - \mathbf{c} = \mathbf{y}$, 则 $\mathbf{a} - \mathbf{c} = \mathbf{x} + \mathbf{y}$, 由此知要证明的不等式为

$$\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} + \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle} \geq \sqrt{\langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle},$$

平方后可知上述不等式等价于

$$\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \cdot \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle} \geq \langle \mathbf{x}, \mathbf{y} \rangle,$$

此即 Cauchy-Schwarz 不等式. □

定义 3.1.5. 集合 X 上的一个度量是指一个满足如下三个条件的映射 $d : X \times X \to \mathbf{R}$, 要求 d 满足:

(1) 对称性, 对任何 $x, y \in X$ 有 $d(x, y) = d(y, x)$.

(2) 正定性, 对任何 $x \in X$ 有 $d(x, y) \geq 0$, 且 $d(x, y) = 0$ 当且仅当 $x = y$.

(3) 三角不等式, 对任何 $x, y, z \in X$ 有 $d(x, y) + d(y, z) \geq d(x, z)$.

称对子 (X, d) 为一个度量空间, 在不引起混淆的时候也简记为度量空间 X.

利用此定义, 可把之前的结果总结为: 内积给出度量.

回到内积空间 $\mathbf{R}^n$. 设向量 $\vec{OA},\vec{OB}$ 分别为 $\mathbf{a},\mathbf{b}$, 角 $\angle AOB=\theta$, 由余弦定理可得

$$\begin{aligned}\cos\theta &= \frac{|OA|^2+|OB|^2-|AB|^2}{2|OA|\cdot|OB|}\\ &= \frac{\langle\mathbf{a},\mathbf{a}\rangle+\langle\mathbf{b},\mathbf{b}\rangle-\langle\mathbf{a}-\mathbf{b},\mathbf{a}-\mathbf{b}\rangle}{2||\mathbf{a}||\cdot||\mathbf{b}||}\\ &= \frac{\langle\mathbf{a},\mathbf{b}\rangle}{||\mathbf{a}||\cdot||\mathbf{b}||}.\end{aligned}$$

这就启发人们在一般的内积空间 V 中, 定义两个非零向量 $\mathbf{a},\mathbf{b}$ 的夹角 θ 为

$$\theta=\arccos\frac{\langle\mathbf{a},\mathbf{b}\rangle}{||\mathbf{a}||\cdot||\mathbf{b}||}\in[0,\pi],$$

Cauchy-Schwarz 不等式保证上式是有定义的. 进一步, 称 $\mathbf{a},\mathbf{b}$ 正交或垂直, 记作 $\mathbf{a}\perp\mathbf{b}$, 如果它们的夹角为 $\frac{\pi}{2}$, 或等价的说如果它们的内积为零. 这样, 勾股定理可以叙述为: 如果 $\mathbf{a}\perp\mathbf{b}$, 则有

$$||\mathbf{a}+\mathbf{b}||^2=||\mathbf{a}||^2+||\mathbf{b}||^2.$$

3.2 正交基底

定义 3.2.1. 设 V 是内积空间, $S=\{\mathbf{a}_1,\cdots,\mathbf{a}_r\}$ 为 V 的一个向量组. 如果 S 中的向量都非零且两两正交, 则称 S 为正交向量组; 进一步, 如果正交向量组中每个向量都是单位长度的, 则称 S 为单位正交向量组.

命题 3.2.2. *正交向量组是线性无关的.*

证明: 设 $S=\{\mathbf{a}_1,\cdots,\mathbf{a}_r\}$ 是正交向量组. 假设实数 $x_1,\cdots,x_r$ 使得 $\sum\limits_{j=1}^{r}x_j\mathbf{a}_j=\mathbf{0}$, 则对每个 $1\le i\le r$ 有

$$0=\langle\mathbf{a}_i,\sum_{j=1}^{r}x_j\mathbf{a}_j\rangle=\sum_{j=1}^{n}x_j\langle\mathbf{a}_i,\mathbf{a}_j\rangle=x_i\langle\mathbf{a}_i,\mathbf{a}_i\rangle,$$

结合 $\langle\mathbf{a}_i,\mathbf{a}_i\rangle\neq0$ 可得 $x_i=0$. 这就证明了 S 线性无关. □

定理 3.2.3 (Gram-Schmidt 正交化)**.** *设 $\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$ 是内积空间 V 的一个线性无关向量组, 则存在单位正交向量组 $\{\mathbf{e}_1,\cdots,\mathbf{e}_r\}$ 使得 $span(\mathbf{e}_1,\cdots,\mathbf{e}_r)=span(\mathbf{v}_1,\cdots,\mathbf{v}_r)$.*

证明: 我们递归的构造单位正交向量组 $\{\mathbf{e}_1,\cdots,\mathbf{e}_r\}$, 使得对每个 $1\le t\le r$ 都有 $\mathrm{span}(\mathbf{e}_1,\cdots,\mathbf{e}_t)=\mathrm{span}(\mathbf{v}_1,\cdots,\mathbf{v}_t)$. 为此, 取 $\mathbf{e}_1=\frac{\mathbf{v}_1}{||\mathbf{v}_1||}$. 假设已经构造好 $\mathbf{e}_1,\cdots,\mathbf{e}_t$, 考虑如下形式的向量

$$\mathbf{w}=\mathbf{v}_{t+1}-x_1\mathbf{e}_1-\cdots-x_t\mathbf{e}_t.$$

对每个 $1 \le i \le t$, 有

$$\langle \mathbf{w}, \mathbf{e}_i \rangle = \langle \mathbf{v}_{t+1}, \mathbf{e}_i \rangle - x_i.$$

这样, 取 $x_i = \langle \mathbf{v}_{t+1}, \mathbf{e}_i \rangle, 1 \le i \le t$, 所得的向量为

$$\mathbf{w} = \mathbf{v}_{t+1} - \sum_{i=1}^{t} \langle \mathbf{v}_{t+1}, \mathbf{e}_i \rangle \mathbf{e}_i,$$

它与 $\mathbf{e}_1, \cdots, \mathbf{e}_t$ 都正交. 易验证 $\mathbf{w} \neq 0$, 且

$$\mathrm{span}(\mathbf{e}_1, \cdots, \mathbf{e}_t, \mathbf{w}) = \mathrm{span}(\mathbf{v}_1, \cdots, \mathbf{v}_t, \mathbf{v}_{t+1}),$$

令 $\mathbf{e}_{t+1} = \frac{\mathbf{w}}{||\mathbf{w}||}$ 即可. 这就完成了归纳构造. □

定义 3.2.4. 设 V 是内积空间. 称 S 为 V 的一个正交基底, 如果 S 是正交向量组且是 V 的一个基底. 称 S 为 V 的一个单位正交基底, 如果 S 是单位正交向量组且是 V 的一个基底.

推论 3.2.5. *设 V 是内积空间, 则它有一个单位正交基底. 进一步, 设 W 是 V 的子空间, 则 W 的任何单位正交基底可扩充为 V 的单位正交基底.*

设 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ 是内积空间 V 的一个单位正交基底, 则任何 $\mathbf{x}$ 可以表示为

$$\mathbf{x} = \mathbf{e}_1 \langle \mathbf{e}_1, \mathbf{x} \rangle + \mathbf{e}_2 \langle \mathbf{e}_2, \mathbf{x} \rangle + \cdots + \mathbf{e}_n \langle \mathbf{e}_n, \mathbf{x} \rangle,$$

可以把上式用分块矩阵乘法表示

$$[\mathbf{x}] = [\mathbf{e}_1, \cdots, \mathbf{e}_n] \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{e}_n, \mathbf{x} \rangle \end{pmatrix}.$$

在此写法下, 如果 $\{\mathbf{a}_1, \cdots, \mathbf{a}_r\}$ 是任何一组向量, 则有

$$[\mathbf{a}_1, \cdots, \mathbf{a}_r] = [\mathbf{e}_1, \cdots, \mathbf{e}_n] \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_r \rangle \\ \vdots & & \vdots \\ \langle \mathbf{e}_n, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{a}_r \rangle \end{pmatrix},$$

这个表示对之后理解 QR 分解特别有用.

命题 3.2.6. *设 $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ 是内积空间 V 的一个线性无关向量组, 生成的子空间为 $W = span(\mathbf{v}_1, \cdots, \mathbf{v}_r)$. 设 $\mathbf{a} \in V$ 是给定的向量, 则距离函数*

$$f(\mathbf{x}) = d(\mathbf{a}, \mathbf{x})$$

在 W 有唯一的最小值点 $\mathbf{x}_0$. 进一步, 最小值点 $\mathbf{x}_0$ 满足 $(\mathbf{a} - \mathbf{x}_0) \perp W$.

证明: 由定理8.1.1可用 Gram-Schmidt 正交化方法给出 W 的单位正交基底 $\{\mathbf{e}_1,\cdots,\mathbf{e}_r\}$, 这样每个 $\mathbf{x}\in W$ 可以唯一的表示成 $\mathbf{x}=\sum\limits_{i=1}^{r}x_i\mathbf{e}_i$, 由此可得

$$\begin{aligned}
f(\mathbf{x})^2 &= \langle \mathbf{a}-\sum_{i=1}^{r}x_i\mathbf{e}_i, \mathbf{a}-\sum_{i=1}^{r}x_i\mathbf{e}_i\rangle \\
&= \langle \mathbf{a},\mathbf{a}\rangle - 2\sum_{i=1}^{r}x_i\langle \mathbf{a},\mathbf{e}_i\rangle + \sum_{i=1}^{r}x_i^2 \\
&= \langle \mathbf{a},\mathbf{a}\rangle - \sum_{i=1}^{r}(\langle \mathbf{a},\mathbf{e}_i\rangle)^2 + \sum_{i=1}^{r}(x_i-\langle \mathbf{a},\mathbf{e}_i\rangle)^2 \\
&\geq \langle \mathbf{a},\mathbf{a}\rangle - \sum_{i=1}^{r}(\langle \mathbf{a},\mathbf{e}_i\rangle)^2,
\end{aligned}$$

它的最小值为

$$\min f(\mathbf{x})^2 = \langle \mathbf{a},\mathbf{a}\rangle - \sum_{i=1}^{r}(\langle \mathbf{a},\mathbf{e}_i\rangle)^2,$$

在唯一的最小值点

$$\mathbf{x}_0 = \sum_{i=1}^{r}\langle \mathbf{a},\mathbf{e}_i\rangle\mathbf{e}_i$$

处取得. 显然对每个 $1\leq i\leq r$ 有 $\langle \mathbf{a}-\mathbf{x}_0,\mathbf{e}_i\rangle=0$, 所以有 $(\mathbf{a}-\mathbf{x}_0)\perp W$. □

在几何上看, 上述命题证明了 $\mathbf{a}$ 到 W 有唯一的最小距离点 $\mathbf{x}_0$, 该点与 $\mathbf{a}$ 的连线垂直于 W, 这就是所谓 $\mathbf{a}$ 向 W 作垂线的垂足. 称 $\mathbf{x}_0$ 代表的向量为 $\mathbf{a}$ 向 W 的正交投影, 记作 $\mathbf{a}_W$; 令 $\mathbf{a}_\perp=\mathbf{a}-\mathbf{x}_0$, 称为 $\mathbf{a}$ 在 W 垂直方向的分量, 这样就得到分解

$$\mathbf{a}=\mathbf{a}_W+\mathbf{a}_\perp,$$

将任何一个向量 $\mathbf{a}$ 分解为它向 W 的正交投影向量与它在 W 垂直方向分量的和.

如果 $\{\mathbf{e}_1,\cdots,\mathbf{e}_r\}$ 是 W 的单位正交基底, 则有

$$\mathbf{a}_W=\sum_{i=1}^{r}\langle \mathbf{a},\mathbf{e}_i\rangle\mathbf{e}_i, \quad \mathbf{a}_\perp=\mathbf{a}-\sum_{i=1}^{r}\langle \mathbf{a},\mathbf{e}_i\rangle\mathbf{e}_i.$$

这样, Gram-Schimdt 正交化方法就变得直接: 递归的找 $W=\mathrm{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_t\}$ 的单位正交基 $\{\mathbf{e}_1,\cdots,\mathbf{e}_t\}$; 下一步将 $\mathbf{v}_{t+1}$ 向 W 做正交投影, 令 $\mathbf{w}$ 为 $\mathbf{v}_{t+1}$ 在 W 正交方向上的分量

$$\mathbf{w}=(\mathbf{v}_{t+1})_\perp,$$

取 $\mathbf{e}_{t+1}=\frac{\mathbf{w}}{|\mathbf{w}|}$ 即可.

如果对 W 只取出一个一般的基底 $W = \mathrm{span}(\mathbf{v}_1, \cdots, \mathbf{v}_r)$, 则前述计算就会变得复杂. 比如, 为了确定 $\mathbf{a}$ 向 W 的正交投影, 设 $\mathbf{a}_W = \sum\limits_{i=1}^{r} x_i \mathbf{v}_i$, 则由 $\mathbf{a} - \mathbf{a}_W \perp \mathbf{v}_k$ 可得

$$\sum_{i=1}^{r} \langle \mathbf{v}_k, \mathbf{v}_i \rangle x_i = \langle \mathbf{v}_k, \mathbf{a} \rangle.$$

这样就等价于求解线性方程组

$$\begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_r \rangle \\ \vdots & & \vdots \\ \langle \mathbf{v}_r, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_r, \mathbf{v}_r \rangle \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} \langle \mathbf{a}, \mathbf{v}_1 \rangle \\ \vdots \\ \langle \mathbf{a}, \mathbf{v}_r \rangle \end{pmatrix},$$

这并不是容易的事情, 由此看出单位正交基的用处. 由于正交投影的存在性, 我们得到如下结论.

推论 3.2.7. 设 $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ 线性无关, 则矩阵

$$\begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_r \rangle \\ \vdots & & \vdots \\ \langle \mathbf{v}_r, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_r, \mathbf{v}_r \rangle \end{pmatrix}$$

是可逆矩阵.

3.3 正交矩阵

现在我们对线性空间 $\mathbf{R}^n$ 赋予了内积结构, 诱导出度量结构. 我们来研究那些保持度量的线性映射.

定义 3.3.1. 设 (X, d_X) 与 (Y, d_Y) 是度量空间, 称映射 $f : X \to Y$ 是保距映射, 如果对任何 $x_1, x_2 \in X$, 有

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2).$$

显然保距映射都是单射.

设 V, W 是内积空间, 内积给出度量结构. 具体的说, 对 $\mathbf{v}_1, \mathbf{v}_2 \in V$, 定义

$$d_V(\mathbf{v}_1, \mathbf{v}_2) = \langle \mathbf{v}_1 - \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle^{1/2},$$

在不引起混淆的情况下, 也把 d_V 简记为 d.

定义 3.3.2. 设 V, W 是内积空间, 称线性映射 $f: V \to W$ 是保持内积的, 如果对任何 $\mathbf{v}_1, \mathbf{v}_2 \in V$, 有

$$\langle f(\mathbf{v}_1), f(\mathbf{v}_2)\rangle = \langle \mathbf{v}_1, \mathbf{v}_2\rangle.$$

命题 3.3.3. 设 V, W 是内积空间, $f: V \to W$ 是线性映射, 则有

$$f \text{ 保持内积} \iff f\text{保距}.$$

证明: "⇒" 设 f 保持内积, 则有

$$\begin{aligned} d(f(\mathbf{v}_1), f(\mathbf{v}_2))^2 &= \langle f(\mathbf{v}_1) - f(\mathbf{v}_2), f(\mathbf{v}_1) - f(\mathbf{v}_2)\rangle \\ &= \langle f(\mathbf{v}_1), f(\mathbf{v}_1)\rangle - 2\langle f(\mathbf{v}_1), f(\mathbf{v}_2)\rangle + \langle f(\mathbf{v}_2), f(\mathbf{v}_2)\rangle \\ &= \langle \mathbf{v}_1, \mathbf{v}_1\rangle - 2\langle \mathbf{v}_1, \mathbf{v}_2\rangle + \langle \mathbf{v}_2, \mathbf{v}_2\rangle \\ &= d(\mathbf{v}_1, \mathbf{v}_2)^2, \end{aligned}$$

即有 f 保距.

"⇐" 设 f 保持距离, 注意到对任何 $\mathbf{v}_1, \mathbf{v}_2 \in V$, 有

$$\begin{aligned} \langle \mathbf{v}_1, \mathbf{v}_2\rangle &= \frac{1}{2}\left(\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2\rangle - \langle \mathbf{v}_1, \mathbf{v}_1\rangle - \langle \mathbf{v}_2, \mathbf{v}_2\rangle\right) \\ &= \frac{1}{2}\left(d(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{0})^2 - d(\mathbf{v}_1, \mathbf{0})^2 - d(\mathbf{v}_2, \mathbf{0})^2\right), \end{aligned}$$

由此可得

$$\begin{aligned} &\langle f(\mathbf{v}_1), f(\mathbf{v}_2)\rangle \\ =&\frac{1}{2}\left(d(f(\mathbf{v}_1) + f(\mathbf{v}_2), \mathbf{0})^2 - d(f(\mathbf{v}_1), \mathbf{0})^2 - d(f(\mathbf{v}_2), \mathbf{0})^2\right) \\ =&\frac{1}{2}\left(d(f(\mathbf{v}_1 + \mathbf{v}_2), f(\mathbf{0}))^2 - d(f(\mathbf{v}_1), f(\mathbf{0}))^2 - d(f(\mathbf{v}_2), f(\mathbf{0}))^2\right) \\ =&\frac{1}{2}\left(d(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{0})^2 - d(\mathbf{v}_1, \mathbf{0})^2 - d(\mathbf{v}_2, \mathbf{0})^2\right) \\ =&\langle \mathbf{v}_1, \mathbf{v}_2\rangle, \end{aligned}$$

即有 f 保内积. □

先研究 $\mathbf{R}^n$ 到自身的保距线性映射.

命题 3.3.4. 设 A 是 n 阶方阵, 则对任何 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ 有

$$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^T\mathbf{y}\rangle.$$

证明: 这是直接的, 把 $\mathbf{x}, \mathbf{y}$ 视为列向量, 则

$$\langle A\mathbf{x}, \mathbf{y}\rangle = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \langle \mathbf{x}, A^T\mathbf{y}\rangle.$$

□

命题 3.3.5. 设 A 是 n 阶方阵, 它给出的线性映射为 $f(\mathbf{x}) = A\mathbf{x}$, 则 f 是保距映射的充分必要条件是如下四个彼此等价的断言:

(1) $A^T A = I_n$;

(2) A 的列向量构成 $\mathbf{R}^n$ 的单位正交基;

(3) $AA^T = I_n$;

(4) A 的行向量构成 $\mathbf{R}^n$ 的单位正交基.

证明: 由命题3.3.3, f 保距当且仅当 f 保持内积. 利用命题3.3.4可得

$$\begin{aligned} f\text{保持内积} &\iff \langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle, \forall \mathbf{x}, \mathbf{y} \\ &\iff \langle \mathbf{x}, A^T A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle, \forall \mathbf{x}, \mathbf{y} \\ &\iff \langle \mathbf{x}, A^T A\mathbf{y} - \mathbf{y}\rangle = 0, \forall \mathbf{x}, \mathbf{y} \\ &\iff A^T A\mathbf{y} - \mathbf{y} = \mathbf{0}, \forall \mathbf{y} \\ &\iff A^T A = I_n, \end{aligned}$$

最后, 利用命题1.8.9可知 $A^T A = I_n$ 当且仅当 A 与 A^T 互为逆, 当且仅当 $AA^T = I_n$.

□

定义 3.3.6. 称方阵 A 为正交矩阵, 如果它满足命题3.3.5中所述的四个彼此等价的断言.

推论 3.3.7. 正交矩阵的乘积是正交矩阵.

对正整数 n, 令 $O(n) = \{A \in M_{n\times n} | AA^T = I_n\}$ 为由所有 n 阶正交矩阵所构成的集合, 则其上有矩阵乘法, 单位矩阵 $I_n \in O(n)$ 是乘法的单位元, 且每一个 $A \in O(n)$ 都有乘法的逆 $A^{-1} = A^T$. 人们称 $O(n)$ 是 n 阶正交群.

例 3.3.8. 显然 $O(1) = \{\pm 1\}$, 下面我们来确定 $O(2)$. 设 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O(2)$, 则有

$$a^2 + c^2 = 1, \quad b^2 + d^2 = 1, \quad ac + bd = 0,$$

设 $(a, c) = (\cos\alpha, \sin\alpha)$, $(b, d) = (\sin\beta, \cos\beta)$, 则有 $0 = ac + bd = \sin(\alpha+\beta)$, 可知 $\alpha + \beta \in \pi\mathbf{Z}$. 由此可得

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \text{ 或 } \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}.$$

注意到, 前者为旋转 R_α, 后者为反射 $H_{\frac{\alpha}{2}}$.

高维内积空间中也有旋转与反射. 设 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ 是 $\mathbf{R}^n$ 的单位正交基底, 可考虑 $\mathbf{e}_i, \mathbf{e}_j (i < j)$ 平面的旋转所确定的变换, 设旋转角度为 θ, 将相应的变换记为 $G(i, j, \theta)$, 称为 Givens 变换, 它的表示矩阵为将单位矩阵 I_n 中第 i, j 行 i, j 列交叉处的 2×2 单位矩阵用 $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 代替.

对于每个单位向量 $\mathbf{v} \in \mathbf{R}^n$, 以它为法向量确定一个 $(n-1)$ 维的子空间

$$W = \{\mathbf{u} \in \mathbf{R}^n | \langle \mathbf{u}, \mathbf{v} \rangle = 0\},$$

称关于 W 的反射为 Householder 变换, 记为 $H_\mathbf{v}$. 具体的说, 对每个 $\mathbf{x}$, 设它向 W 的正交投影为 $\mathbf{x}_W$, 它在 W 垂直方向的分量为 $\mathbf{x}_\perp$, 则有

$$\mathbf{x} = \mathbf{x}_W + \mathbf{x}_\perp, \quad H_\mathbf{v}(\mathbf{x}) = \mathbf{x}_W - \mathbf{x}_\perp,$$

注意到 $\mathbf{x}_\perp$ 与 $\mathbf{v}$ 同方向, 可知 $\mathbf{x}_\perp = \langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v}$, 由此可得

$$H_\mathbf{v}(\mathbf{x}) = \mathbf{x} - 2\langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v} = \mathbf{x} - 2\mathbf{v}\mathbf{v}^T\mathbf{x} = (I_n - 2\mathbf{v}\mathbf{v}^T)\mathbf{x}.$$

特别的, $H_\mathbf{v}$ 的表示矩阵为 $H_\mathbf{v} = I_n - 2\mathbf{v}\mathbf{v}^T$.

3.4 QR 分解

定理 3.4.1. *设 A 是 n 阶可逆矩阵, 则存在一个正交矩阵 Q 与一个对角元为正数的上三角矩阵 R, 使得 $A = QR$. 进一步, 满足上述条件的 Q, R 都是唯一的.*

证明: 存在性. 将 A 表示成列向量分块矩阵 $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$, 则 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 线性无关. 利用 Gram-Schmidt 正交化 (定理8.1.1), 存在 $\mathbf{R}^n$ 单位正交基底 $\mathbf{e}_1, \cdots, \mathbf{e}_n$, 使得对每个 $k \le n$ 都有 $\mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_k\} = \mathrm{span}\{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$. 特别的, $\mathbf{a}_k \in \mathrm{span}\{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$, 从而有

$$\mathbf{a}_k = \mathbf{e}_1\langle \mathbf{e}_1, \mathbf{a}_k \rangle + \mathbf{e}_2\langle \mathbf{e}_2, \mathbf{a}_k \rangle + \cdots + \mathbf{e}_k\langle \mathbf{e}_k, \mathbf{a}_k \rangle.$$

写成矩阵的形式, 记 $Q = [\mathbf{e}_1, \cdots, \mathbf{e}_n]$, 令

$$R = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{e}_n, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{a}_n \rangle \end{pmatrix}.$$

易知 Q 是正交矩阵, R 是上三角矩阵, 若 R 的某个对角元 $\langle \mathbf{e}_i, \mathbf{a}_i \rangle$ 为负的, 则用 $-\mathbf{e}_i$ 代替 $\mathbf{e}_i$, 可把 R 的 ii 矩阵元改成正数. 这样, 可不妨设 R 的对角元都是正数, 从而得到 A 的 QR 分解 $A = QR$.

唯一性. 如果 $Q_1R_1 = Q_2R_2$, 则 $(Q_2)^{-1}Q_1 = R_2(R_1)^{-1}$ 既是正交矩阵, 又是对角元为正的上三角矩阵, 可直接验证它只能是单位矩阵. □

定理 3.4.2. *设 A 是 $m \times n$ 矩阵, 秩为 r, 则存在 $m \times r$ 矩阵 Q 与 $r \times n$ 矩阵 R, 满足如下条件:*

(1) *$A = QR$;*

(2) *$Q^TQ = I_r$, 即 Q 的列向量是单位正交向量组;*

(3) *R 是上三角矩阵且对角元非负.*

证明: 仿照 Gram-Schmidt 正交化, 先构造 $\mathbf{R}^m$ 的正交单位向量组 $E = \{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$. 初始时, 令 $E_0 = \emptyset$. 依次对 A 对列向量 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 做如下考察: 如果 $\mathbf{a}_k \in \mathrm{span}(E_{k-1})$, 则令 $E_k = E_{k-1}$, 转而考虑下一个 $\mathbf{a}_{k+1}$; 如果此时 $\mathbf{a}_k \notin \mathrm{span}(E_{k-1})$, 则令 $\mathbf{u}$ 为 $\mathbf{a}_k$ 在 $W_{k-1} = \mathrm{span}(E_{k-1})$ 垂直方向上的分量, 令 $E_k = E_{k-1} \cup \{\frac{\mathbf{u}}{|\mathbf{u}|}\}$, 并转到考虑下一个 $\mathbf{a}_{k+1}$. 当所有 n 步考察结束时, 把所得的 E_n 记为 E, 则 E 是由 r 个单位正交向量构成, 且对每个 k 都有 $|E_k| \leq k$. 按照加入 E 的先后顺序, 将 E 的成员依次编号为 $\mathbf{e}_1, \cdots, \mathbf{e}_r$, 令 $Q = [\mathbf{e}_1, \cdots, \mathbf{e}_r]$, 则有

$$A = Q \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{e}_r, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{e}_r, \mathbf{a}_n \rangle \end{pmatrix}.$$

这就给出了 A 的 QR 分解. 注意到, 由构造方式可知对每个 $k \leq r$, 有

$$\mathbf{a}_k \in \mathrm{span}(E_k) \subseteq \mathrm{span}(\mathbf{e}_1, \cdots, \mathbf{e}_k),$$

由此可知当 $i > k$ 时, $R_{ik} = \langle \mathbf{e}_i, \mathbf{a}_k \rangle = 0$, 即有 R 是上三角矩阵. 最后, 若 R 的某个对角元 $\langle \mathbf{e}_i, \mathbf{a}_i \rangle$ 为负的, 则用 $-\mathbf{e}_i$ 代替 $\mathbf{e}_i$, 可把 R 的 ii 矩阵元改成正数. 这样, 可不妨设 R 的对角元都是非负数.

□

3.5 正交补

定义 3.5.1. 设 W 是 $\mathbf{R}^n$ 的子空间, 称 $\mathbf{a}$ 与 W 正交, 记作 $\mathbf{a} \perp W$, 如果 $\mathbf{a}$ 与 W 中任何向量都正交.

设 V, W 是 $\mathbf{R}^n$ 的子空间, 称 V 与 W 正交, 记作 $V \perp W$, 如果对任意 $\mathbf{v} \in V$ 与任意 $\mathbf{w} \in W$ 都有 $\mathbf{v} \perp \mathbf{w}$.

定义 3.5.2. 设 V 是 $\mathbf{R}^n$ 的子空间, 定义 $V^\perp = \{\mathbf{a} \in \mathbf{R}^n | \mathbf{a} \perp V\}$, 称为 V 的正交补, 易验证它是 $\mathbf{R}^n$ 的子空间.

定义 3.5.3. 设 V, W 是 $\mathbf{R}^n$ 的子空间, 令

$$V + W = \{\mathbf{v} + \mathbf{w} | \mathbf{v} \in V, \mathbf{w} \in W\},$$

它是 $\mathbf{R}^n$ 的子空间, 称之为 V 与 W 的和. 进一步, 如果 $V \cap W = \{\mathbf{0}\}$, 则称 $V + W$ 为 V 与 W 的直和, 记作 $V \oplus W$.

定理 3.5.4. 设 V, W 是 $\mathbf{R}^n$ 的子空间, 则有

$$dim(V + W) = dim(V) + dim(W) - dim(V \cap W).$$

证明: 取 $V \cap W$ 的一个基底 $\{\mathbf{a}_i\}_{1 \le i \le r}$, 将它扩充为 V 的一个基底 $\{\mathbf{a}_i\}_{1 \le i \le r} \cup \{\mathbf{b}_j\}_{1 \le j \le s}$; 将 $\{\mathbf{a}_i\}_{1 \le i \le r}$ 扩充为 W 的一个基底 $\{\mathbf{a}_i\}_{1 \le i \le r} \cup \{\mathbf{c}_k\}_{1 \le k \le t}$.

易验证

$$\{\mathbf{a}_i\}_{1 \le i \le r} \cup \{\mathbf{b}_j\}_{1 \le j \le s} \cup \{\mathbf{c}_k\}_{1 \le k \le t}$$

是 $V + W$ 的一个基底, 由此可得定理中所述有关维数的等式成立.

□

定理 3.5.5. 设 V 是 $\mathbf{R}^n$ 的子空间, 则 $\mathbf{R}^n = V \oplus V^\perp$.

证明: 如果 $\mathbf{a} \in V \cap V^\perp$, 则 $\mathbf{a} \in V$ 且 $\mathbf{a} \perp V$, 从而有 $\langle \mathbf{a}, \mathbf{a} \rangle = 0$, 因而 $\mathbf{a} = \mathbf{0}$. 这说明 $V \cap V^\perp = \{\mathbf{0}\}$.

利用我们在命题3.2.6之后所得的结论, 任何 $\mathbf{a} \in \mathbf{R}$ 都可以表示成

$$\mathbf{a} = \mathbf{a}_V + \mathbf{a}_\perp, \quad \mathbf{a}_V \in V, \quad \mathbf{a}_\perp \in V^\perp,$$

这说明 $\mathbf{R}^n = V + V^\perp$.

结合这两部分, 即有 $\mathbf{R}^n = V \oplus V^\perp$. □

人们也把上述定理的内容表述成如下形式. 每个 $\mathbf{a} \in \mathbf{R}$ 由唯一的表示

$$\mathbf{a} = \mathbf{a}_V + \mathbf{a}_\perp, \quad \mathbf{a}_V \in V, \quad \mathbf{a}_\perp \in V^\perp, \tag{3.1}$$

从而可以定义映射 $P_V : \mathbf{R}^n \to V$ 为 $P_V(\mathbf{a}) = \mathbf{a}_V$, 易验证它是线性映射, 称之为向 V 的正交投影, 也称 $P_V(\mathbf{a})$ 为 $\mathbf{a}$ 在 V 上的正交投影. 由此可把等式3.1 写成

$$\mathrm{Id}_{\mathbf{R}^n} = P_V + P_{V^\perp}.$$

在实际的计算中, 如果为 V 选取了一个正交单位基底 $\{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$, 则

$$P_V = \sum_{i=1}^{r} \mathbf{e}_i \langle \mathbf{e}_i, \cdot \rangle.$$

可以从更加一般的角度来理解投影. 假设可以把 $\mathbf{R}^n$ 表示成两个子空间的直和 $\mathbf{R}^n = V \oplus W$, 则可定义线性映射 $P : \mathbf{R}^n \to \mathbf{R}^n$: 对任何 $\mathbf{a} \in \mathbf{R}^n$, 它有唯一的表示 $\mathbf{a} = \mathbf{v} + \mathbf{w}$, 其中 $\mathbf{v} \in V, \mathbf{w} \in W$, 则定义 $P(\mathbf{a})$ 的值为前述唯一的 $\mathbf{v}$. 称这样的映射 P 为向 V 的投影映射. 注意, 如果只给出子空间 V, 是没法确定这个投影映射 P 的, 必须要给出直和分解 $\mathbf{R}^n = V \oplus W$ 才能定义 P. 因此, 严格的说 P 是由直和分解 $\mathbf{R}^n = V \oplus W$ 所确定的向 V 的投影映射.

进一步, 如果前述直和分解中 V 与 W 是正交的, 则对于 $\mathbf{a} = \mathbf{v} + \mathbf{w} (\mathbf{v} \in V, \mathbf{w} \in W)$, 有

$$\mathbf{a} \in V^{\perp} \iff (\mathbf{v} + \mathbf{w}) \perp V \iff \mathbf{v} \perp V \iff \mathbf{v} = \mathbf{0} \iff \mathbf{a} \in W,$$

因而有 $W = V^{\perp}$, 即直和分解为正交补直和分解 $\mathbf{R}^n = V \oplus V^{\perp}$. 此直和分解给出的向 V 的投影即为向 V 的正交投影 P_V.

定义 3.5.6. 称线性映射 $P : \mathbf{R}^n \to \mathbf{R}^n$ 为投影映射, 如果它来自于某个直和分解 $\mathbf{R}^n = V \oplus W$ 给出的投影映射; 称线性映射 $P : \mathbf{R}^n \to \mathbf{R}^n$ 为正交投影映射, 如果它来自于某个正交直和分解 $\mathbf{R}^n = V \oplus V^{\perp}$ 给出的投影映射.

将投影映射的表示矩阵称为投影矩阵, 将正交投影映射的表示矩阵称为正交投影矩阵.

命题 3.5.7. *设 $P : \mathbf{R}^n \to \mathbf{R}^n$ 是线性映射.*

(1) P 是投影映射的充分必要条件是 $P \circ P = P$.

(2) P 是正交投影映射的充分必要条件是 $P \circ P = P$ 且

$$\langle P(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, P(\mathbf{y}) \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^n.$$

证明: (1) 必要性是显然的, 只需验证充分性. 设 $P \circ P = P$, 令 $V = \mathrm{Im}(P)$, $W = \ker(P)$. 对任何 $\mathbf{x} \in V \cap W$, 存在 $\mathbf{y}$ 使得 $\mathbf{x} = P(\mathbf{y})$ 且 $P(\mathbf{x}) = \mathbf{0}$, 则有

$$\mathbf{0} = P(\mathbf{x}) = (P \circ P)(\mathbf{y}) = P(\mathbf{y}) = \mathbf{x},$$

这表明 $V \cap W = \{\mathbf{0}\}$. 利用维数公式 (定理2.2.2) 与子空间和的维数公式 (定理7.1.8) 可得

$$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W) = \dim\mathrm{Im}(P) + \dim\ker(P) = n,$$

这样子空间 $V + W$ 与全空间 $\mathbf{R}^n$ 维数一样, 因而有 $V + W = \mathbf{R}^n$, 结合 $V \cap W = \{\mathbf{0}\}$ 可得直和分解 $\mathbf{R}^n = V \oplus W$, 将此分解给出的投影映射记为 Q. 对任何 $\mathbf{a}$, 它有唯一分解 $\mathbf{a} = P(\mathbf{x}) + \mathbf{w}$, 其中 $P(\mathbf{x}) \in V$, $\mathbf{w} \in W = \ker(P)$, 可得

$$Q(\mathbf{a}) = P(\mathbf{x}) = P(\mathbf{x} + \mathbf{w}) = P(\mathbf{a}),$$

这表明 $P = Q$ 是投影映射.

(2) 必要性直接验证即可. 我们来证充分性, 由 (1) 的证明过程, 已有直和分解

$$\mathbf{R}^n = \mathrm{Im}(P) \oplus \ker(P),$$

只需要进一步证明 $\mathrm{Im}(P) \perp \ker(P)$ 即可. 这是直接的, 对任何 $P(\mathbf{x}) \in \mathrm{Im}(P)$ 与任何 $\mathbf{y} \in \ker(P)$, 有

$$\langle P(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, P(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{0} \rangle = 0.$$

□

推论 3.5.8. *(1) 方阵 A 是投影矩阵的充分必要条件是 $A^2 = A$.*

(2) 方阵 A 是正交投影矩阵的充分必要条件是 $A^2 = A = A^T$.

命题 3.5.9. *设 V 是 $\mathbf{R}^n$ 的子空间, 则 $(V^\perp)^\perp = V$.*

证明: 取 V 的一个单位正交基底 $\{\mathbf{e}_i\}_{1 \le i \le r}$, 将它扩充为 $\mathbf{R}^n$ 的单位正交基底 $\{\mathbf{e}_i\}_{1 \le i \le n}$, 则有 $V^\perp = \mathrm{span}\{\mathbf{e}_i\}_{r+1 \le i \le n}$. 由此进一步可得

$$(V^\perp)^\perp = \mathrm{span}\{\mathbf{e}_i\}_{1 \le i \le r} = V.$$

□

定理 3.5.10. *设 A 是 $m \times n$ 矩阵, 则有*

(1) $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$;

(2) $\mathcal{R}(A^T A) = \mathcal{R}(A^T)$, $\mathcal{N}(A^T A) = \mathcal{N}(A)$.

证明: (1) 对 $\mathbf{y} \in \mathbf{R}^m$, 有

$$\begin{aligned} \mathbf{y} \in \mathcal{R}(A)^\perp &\iff \langle A\mathbf{x}, \mathbf{y} \rangle = 0 \quad \forall \mathbf{x} \\ &\iff \langle \mathbf{x}, A^T\mathbf{y} \rangle = 0 \quad \forall \mathbf{x} \\ &\iff A^T\mathbf{y} = \mathbf{0} \\ &\iff \mathbf{y} \in \mathcal{N}(A^T). \end{aligned}$$

(2) 利用 (1) 以及定理3.5.5的结论, 有

$$\mathbf{R}^m = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp = \mathcal{R}(A) \oplus \mathcal{N}(A^T),$$

由此即得

$$\mathcal{R}(A^T) = A^T(\mathbf{R}^m) = A^T(\mathcal{R}(A) \oplus \mathcal{N}(A^T)) = A^T(\mathcal{R}(A)) = \mathcal{R}(A^T A).$$

考虑 $\mathcal{N}(A^TA)$. 一方面显然有 $\mathcal{N}(A) \subseteq \mathcal{N}(A^TA)$; 另一方面, 如果 $\mathbf{x} \in \mathcal{N}(A^TA)$, 则有

$$0 = \langle A^TA\mathbf{x}, \mathbf{x}\rangle = \langle A\mathbf{x}, A\mathbf{x}\rangle,$$

因而有 $A\mathbf{x} = \mathbf{0}$, 说明 $\mathcal{N}(A^TA) \subseteq \mathcal{N}(A)$. 结合这两方面, 即有 $\mathcal{N}(A^TA) = \mathcal{N}(A)$. □

3.6 线性最小二乘法

假设在某个科学系统中, 理论预言变量 y 按如下方式依赖于自变量 t:

$$y = f(t; x_1, \ldots, x_n),$$

其中 $x_1, \ldots, x_n$ 是未知的参数. 设有大量的实验数据 $(t_1, y_1), ..., (t_m, y_m)$. 如何选取参数 $x_1, \ldots, x_n$, 使得理论预言与实验结果拟合的最好? 人们将此问题称为数据拟合, 在各个学科中都会经常遇见.

最小二乘法考虑所有误差的平方和

$$E(x_1, \ldots, x_n) = \sum_{i=1}^{m} \left(f(t_i; x_1, \ldots, x_n) - y_i\right)^2.$$

计算 $E(x_1, \ldots, x_n)$ 的最小值, 可以认为最小值点给出最佳拟合方案.

在微积分中, 可按照如下步骤求最小值.

(1) 存在性. 利用最值定理或者不等式估计证明 E 能取到最小值, 设最小值点为 $\mathbf{a} = (a_1, \ldots, a_n)$.

(2) 最小值点是临界点, 它满足临界点的方程

$$\begin{cases} \frac{\partial E}{\partial x_1} = \sum_{i=1}^{m} 2\left(f(t_i; x_1, \ldots, x_n) - y_i\right) \frac{\partial f}{\partial x_1} = 0, \\ ..., \\ \frac{\partial E}{\partial x_n} = \sum_{i=1}^{m} 2\left(f(x_i; x_1, \ldots, x_n) - y_i\right) \frac{\partial f}{\partial x_n} = 0, \end{cases}$$

求解这个方程组, 找出所有临界点. 比较临界点处 $E(x_1, \ldots, x_n)$ 的值, 它们的最小值就是 E 在 $\mathbf{R}^n$ 上的最小值.

我们来考虑一个特例, 假设 y 线性的依赖于参数 $x_1, \ldots, x_n$, 即有

$$y = x_1 f_1(t) + \cdots + x_n f_n(t),$$

其中 $f_1(t), \cdots, f_n(t)$ 是给定的函数. 在此假设下, 最小二乘法要求出如下函数

$$E(x_1, \cdots, x_n) = \sum_{i=1}^{m} (x_1 f_1(t_i) + \cdots + x_n f_n(t_i) - y_i)^2$$

的最小值, 由此确定出最佳拟合.

记 $A = (f_j(t_i))_{1\le i\le m, 1\le j\le n}$, 它是 $m \times n$ 矩阵, 记 $\mathbf{x} = (x_1, \cdots, x_n)^T$, $\mathbf{b} = (y_1, \cdots, y_m)^T$, 则最小二乘法可叙述为求解如下最小值问题:

$$\min_{\mathbf{x}\in\mathbf{R}^n} ||A\mathbf{x} - \mathbf{b}||.$$

令 $V = \mathcal{R}(A)$, 则上述最小值问题恰为我们在命题3.2.6中解决过的问题: 求 $\mathbf{b}$ 到子空间 V 的最短距离. 利用命题3.2.6的结论, $\mathbf{b}$ 到 V 的最小距离点 $\mathbf{v}_0$ 是 $\mathbf{b}$ 向 V 的正交投影, 它由条件 $(\mathbf{b} - \mathbf{v}_0) \perp V$ 唯一确定.

接下来, 人们有两种方法进行求解.

(1) 正则化方法.

设 $\mathbf{v}_0 = A\mathbf{x}_0$, 此条件并没有完全确定 $\mathbf{x}_0$, 因为总可以在 $\mathbf{x}_0$ 上加上一个 $\mathcal{N}(A)$ 中的向量. 由于 $A\mathbf{x}_0$ 是 $\mathbf{b}$ 向 V 的正交投影, 则有

$$\mathbf{b} - A\mathbf{x}_0 \in \mathcal{R}(A)^{\perp} = \mathcal{N}(A^T),$$

即有 $A^T A\mathbf{x}_0 = A^T\mathbf{b}$.

命题 3.6.1 (正则化方法求解线性最小二乘问题)**.** *线性最小二乘法问题 $\min_{\mathbf{x}\in\mathbf{R}^n} ||A\mathbf{x} - \mathbf{b}||$ 有解. 进一步, $\mathbf{x}_0$ 是其最小值点的充分必要条件是 $A^T A\mathbf{x}_0 = A^T\mathbf{b}$.*

2 正交化方法.

选取 $V = \mathcal{R}(A)$ 的单位正交基底 $\{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$, 则 $\mathbf{b}$ 向 V 的正交投影为

$$\mathbf{v}_0 = P_V(\mathbf{b}) = \sum_{i=1}^{r} \mathbf{e}_i \langle \mathbf{e}_i, \mathbf{b}\rangle = [\mathbf{e}_1, \cdots, \mathbf{e}_r] \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{b}\rangle \\ \vdots \\ \langle \mathbf{e}_r, \mathbf{b}\rangle \end{pmatrix},$$

之后再求解方程 $A\mathbf{x}_0 = \mathbf{v}_0$.

注意到, 定理8.1.5告诉我们: 为 $V = \mathcal{R}(A)$ 构造单位正交基底的过程恰好就是对 A 做 QR 分解的过程. 按照定理8.1.5证明中的办法构造出 $V = \mathcal{R}(A)$ 的单位正交基底 $\{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$,

则它给出 A 的 QR 分解:

$$A=[\mathbf{e}_1,\cdots,\mathbf{e}_r]\begin{pmatrix}\langle\mathbf{e}_1,\mathbf{a}_1\rangle & \cdots & \langle\mathbf{e}_1,\mathbf{a}_n\rangle\\ \vdots & & \vdots\\ \langle\mathbf{e}_r,\mathbf{a}_1\rangle & \cdots & \langle\mathbf{e}_r,\mathbf{a}_n\rangle\end{pmatrix}=QR.$$

这样, 求解方程 $A\mathbf{x}_0=\mathbf{v}_0$ 就等价于

$$QR\mathbf{x}_0=\mathbf{v}_0=Q\begin{pmatrix}\langle\mathbf{e}_1,\mathbf{b}\rangle\\ \vdots\\ \langle\mathbf{e}_r,\mathbf{b}\rangle\end{pmatrix}=QQ^T\mathbf{b}. \tag{3.2}$$

注意到, 由于 $\{\mathbf{e}_1,\cdots,\mathbf{e}_r\}$ 单位正交, 有 $Q^TQ=I_r$, 方程3.2 两边用 Q^T 左乘, 可得

$$QR\mathbf{x}_0=QQ^T\mathbf{b}\Rightarrow R\mathbf{x}_0=Q^T\mathbf{b};$$

反之, 显然有

$$R\mathbf{x}_0=Q^T\mathbf{b}\Rightarrow QR\mathbf{x}_0=QQ^T\mathbf{b}.$$

结合这两方面, 可知只需求解简单的方程 $R\mathbf{x}_0=Q^T\mathbf{b}$. 可以把上述正交化方法叙述为如下的命题.

命题 3.6.2 (正交化方法求解线性最小二乘问题)**.** 设 A 的 QR 分解为 $A=QR$, 则 $\mathbf{x}_0$ 是线性最小二乘法问题 $\min\limits_{\mathbf{x}\in\mathbf{R}^n}||A\mathbf{x}-\mathbf{b}||$ 的最小值点的充分必要条件是 $R\mathbf{x}_0=Q^T\mathbf{b}$.

第四章　行列式

4.1　平行高维体的体积

设 $\mathbf{a}, \mathbf{b}$ 是 $\mathbf{R}^n$ 中的向量, 我们来计算由它们张成的平行四边形的面积 S. 设这两个向量的夹角为 θ, 则有

$$\cos\theta = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{|\mathbf{a}| \cdot |\mathbf{b}|},$$

由此可得所求的平行四边形面积为

$$\begin{aligned} S &= |\mathbf{a}| \cdot |\mathbf{b}| \sin\theta \\ &= |\mathbf{a}| \cdot |\mathbf{b}| \sqrt{1 - (\frac{\langle \mathbf{a}, \mathbf{b} \rangle}{|\mathbf{a}| \cdot |\mathbf{b}|})^2} \\ &= \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle \langle \mathbf{b}, \mathbf{b} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle^2}. \end{aligned} \tag{4.1}$$

例 4.1.1. 设 $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2)$ 是平面向量, 由面积公式4.1可得它们张成的平行四边形的面积为

$$\begin{aligned} S &= \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle \langle \mathbf{b}, \mathbf{b} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle^2} \\ &= |a_1 b_2 - a_2 b_1|. \end{aligned}$$

例 4.1.2. 设 $\mathbf{a} = (a_1, a_2, a_3), \mathbf{b} = (b_1, b_2, b_3)$ 是空间向量, 由面积公式4.1可得它们张成的平行四边形的面积为

$$\begin{aligned} S &= \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle \langle \mathbf{b}, \mathbf{b} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle^2} \\ &= \sqrt{(a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2}, \end{aligned}$$

上式启发人们定义向量 $\mathbf{a}$ 与 $\mathbf{b}$ 的叉乘为

$$\mathbf{a} \times \mathbf{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1),$$

则三维向量 $\mathbf{a}, \mathbf{b}$ 张成的平行四边形面积公式可以写成

$$S = |\mathbf{a} \times \mathbf{b}|.$$

为了确定 $\mathbf{a} \times \mathbf{b}$ 的方向, 直接计算有

$$\langle \mathbf{a} \times \mathbf{b}, \mathbf{a} \rangle = \langle \mathbf{a} \times \mathbf{b}, \mathbf{b} \rangle = 0,$$

可知 $\mathbf{a} \times \mathbf{b}$ 与 $\mathbf{a}, \mathbf{b}$ 所在的平面正交.

例 4.1.3. 设 $\mathbf{a} = (a_1, a_2, a_3), \mathbf{b} = (b_1, b_2, b_3), \mathbf{c} = (c_1, c_2, c_3)$ 是空间向量, 我们来计算由 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 张成的平行六面体的体积 V.

设 $\mathbf{a}, \mathbf{b}$ 张成的平面为 W, 由定理3.5.5知 $\mathbf{c}$ 有分解 $\mathbf{c} = \mathbf{c}_W + \mathbf{c}_\perp$, 其中 $\mathbf{c}_W$ 是 $\mathbf{c}$ 向 W 的正交投影, $\mathbf{c}_\perp$ 正交于 W. 由上例的结论可知 $\mathbf{c}_\perp$ 正比于 $\mathbf{a} \times \mathbf{b}$, 设 $\mathbf{c}_\perp = \lambda(\mathbf{a} \times \mathbf{b})$, 则有

$$\langle \mathbf{a} \times \mathbf{b}, \mathbf{c} \rangle = \langle \mathbf{a} \times \mathbf{b}, \mathbf{c}_\perp \rangle = \lambda \cdot |\mathbf{a} \times \mathbf{b}|^2,$$

由此可得

$$\mathbf{c}_\perp = \frac{\langle \mathbf{a} \times \mathbf{b}, \mathbf{c} \rangle}{|\mathbf{a} \times \mathbf{b}|^2} (\mathbf{a} \times \mathbf{b}).$$

由 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 张成的平行六面体的底面为 $\mathbf{a}, \mathbf{b}$ 张成的平行四边形, 面积为 $S = |\mathbf{a} \times \mathbf{b}|$, 高为 $h = |\mathbf{c}_\perp|$, 所以其体积为

$$\begin{aligned} V = S \cdot h &= \Big| \langle \mathbf{a} \times \mathbf{b}, \mathbf{c} \rangle \Big| \\ &= \Big| a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1 - a_2 b_1 c_3 - a_1 b_3 c_2 \Big|. \end{aligned}$$

由例4.1.1与例4.1.3, 人们定义 2×2 与 3×3 矩阵的行列式为

$$\det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = a_1 b_2 - a_2 b_1,$$

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1 - a_2 b_1 c_3 - a_1 b_3 c_2.$$

这样, 平行四边形的面积与平行六面体的体积都可以表述成行列式的绝对值. 进一步, 可把行列式的定义推广至高阶方阵.

4.2 行列式的定义

定义 4.2.1. 集合 $\{1,2,\cdots,n\}$ 的一个置换是指一个双射 $\sigma:\{1,2,\cdots,n\}\to\{1,2,\cdots,n\}$. 将集合 $\{1,2,\cdots,n\}$ 的所有置换构成的集合记为 S_n, 称为 n 个字母的置换群. 可以将置换 σ 具体表示为

$$\sigma(1),\cdots,\sigma(n).$$

对于置换 σ, 定义它的逆序对集为

$$\mathrm{Inv}(\sigma)=\{(i,j)|1\le i<j\le n,\sigma(i)>\sigma(j)\}.$$

定义置换 σ 的符号为

$$\mathrm{sign}(\sigma)=(-1)^{|\mathrm{Inv}(\sigma)|}.$$

命题 4.2.2. *对任何置换 σ, 有*

$$sign(\sigma)=\prod_{1\le i<j\le n}\frac{\sigma(i)-\sigma(j)}{i-j}=\prod_{l>k}\frac{\sigma(l)-\sigma(k)}{l-k}.$$

证明: 注意到, 如下 C_n^2 个差

$$\sigma(l)-\sigma(k),\quad l>k$$

中恰好 $|\mathrm{Inv}(\sigma)|$ 个为负, 由此可得

$$\prod_{l>k}(\sigma(l)-\sigma(k))=(-1)^{|\mathrm{Inv}(\sigma)|}\cdot\prod_{l>k}|\sigma(l)-\sigma(k)|=(-1)^{|\mathrm{Inv}(\sigma)|}\cdot\prod_{l>k}(l-k)|,$$

命题由此得证. □

定义 4.2.3. 设 $\sigma,\sigma'\in S_n$, 称 σ' 是由 σ 经过一次对换得到, 如果存在 $i<j\in\{1,2,\cdots,n\}$ 使得

$$\sigma'(k)=\begin{cases}\sigma(j), & 如果k=i;\\ \sigma(i), & 如果k=j;\\ \sigma(k), & 如果k\ne i,j.\end{cases}$$

即有 σ' 形如

$$\sigma(1),\cdots,\sigma(j),\cdots,\sigma(i),\cdots,\sigma(n),$$

并称 σ' 是由 σ 经对换 $\sigma(i),\sigma(j)$ 得到.

命题 4.2.4. (1) 如果 σ' 是由 σ 经过一次对换得到, 则 $sign(\sigma') = -sign(\sigma)$.

(2) 如果 σ 能通过奇数次对换变成 $1, 2, \cdots, n$, 则 $sign(\sigma) = -1$; 如果 σ 能通过偶数次对换变成 $1, 2, \cdots, n$, 则 $sign(\sigma) = 1$.

(3) 设 σ^{-1} 是 σ 的逆映射, 则 $sign(\sigma^{-1}) = sign(\sigma)$.

证明: (1) 设 σ' 是由 σ 经过对换 $\sigma(i), \sigma(j) (i < j)$ 得到, 则由命题4.2.2可得

$$\frac{\mathrm{sign}(\sigma')}{\mathrm{sign}(\sigma)} = \frac{\sigma(i) - \sigma(j)}{\sigma(j) - \sigma(i)} \cdot \frac{\prod_{k=i+1}^{j-1}(\sigma(i) - \sigma(k))}{\prod_{k=i+1}^{j-1}(\sigma(k) - \sigma(i))} \cdot \frac{\prod_{k=i+1}^{j-1}(\sigma(k) - \sigma(j))}{\prod_{k=i+1}^{j-1}(\sigma(j) - \sigma(k))} = -1.$$

(2) 由 (1) 的结论即得.

(3) 注意到

$$\{(i, j) | i < j \text{ 且} \sigma^{-1}(i) > \sigma^{-1}(j)\} = \{(k, l) | \sigma(k) < \sigma(l) \text{ 且} k > l\},$$

即有 $|\mathrm{Inv}(\sigma^{-1})| = |\mathrm{Inv}(\sigma)|$. □

定义 4.2.5. 设 $A = (a_{ij})_{1 \le i,j \le n}$ 是 n 阶方阵, 定义 A 的行列式为

$$\det A = \sum_{\sigma \in S_n} \big(\mathrm{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}\big).$$

命题 4.2.6. 行列式可以按列展开, 即有

$$\det A = \sum_{\sigma \in S_n} \left(sign(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}\right).$$

证明: 利用行列式的定义, 结合命题4.2.4, 可得

$$\begin{aligned}
\det A &= \sum_{\tau \in S_n} \big(\mathrm{sign}(\tau) a_{1\tau(1)} a_{2\tau(2)} \cdots a_{n\tau(n)}\big) \\
&= \sum_{\tau \in S_n} \big(\mathrm{sign}(\tau) a_{\tau^{-1}(1)1} a_{\tau^{-1}(2)2} \cdots a_{\tau^{-1}(n)n}\big) \\
&= \sum_{\sigma \in S_n} \big(\mathrm{sign}(\sigma^{-1}) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}\big) \\
&= \sum_{\sigma \in S_n} \left(\mathrm{sign}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}\right).
\end{aligned}$$

□

推论 4.2.7. 设 A 是 n 阶方阵, 则 $\det(A^T) = \det A$.

证明*:* 利用定义8.3.1与命题4.2.6, 可得

$$\begin{aligned}\det(A^T) &= \sum_{\sigma\in S_n}\big(\mathrm{sign}(\sigma)A^T_{1\sigma(1)}A^T_{2\sigma(2)}\cdots A^T_{n\sigma(n)}\big)\\ &= \sum_{\sigma\in S_n}\big(\mathrm{sign}(\sigma)A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n}\big)\\ &= \det A.\end{aligned}$$

□

将方阵 A 表示成列分块矩阵的形式 $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$, 则可将行列式 $\det A$ 视为 n 个列向量的函数 $\det[\mathbf{a}_1, \cdots, \mathbf{a}_n]$, 称之为行列式函数.

命题 4.2.8. *行列式函数满足如下性质:*

(1) *反对称性: 矩阵中两列交换后行列式变成相反数, 即对任何 $i < j$ 有*

$$\det[\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n] = -\det[\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n];$$

特别的, 如果矩阵中有两个列向量相等, 则行列式等于零;

(2) *行列式关于每一列都是线性的, 即有*

$$\det[\mathbf{a}_1, \cdots, k\mathbf{a}_i + l\mathbf{v}, \cdots, \mathbf{a}_n] = k\cdot\det[\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n] + l\cdot\det[\mathbf{a}_1, \cdots, \mathbf{v}, \cdots, \mathbf{a}_n];$$

(3) *类似的, 行列式关于每行都是线性的; 交换矩阵中某两行后行列式变成相反数; 如果矩阵中有两个行向量相等则行列式等于零.*

证明*:* (1) 利用命题4.2.6与命题4.2.4中 (1) 的结论, 可得

$$\begin{aligned}&\det[\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n]\\ =&\sum_{\sigma\in S_n}\big(\mathrm{sign}(\sigma)a_{\sigma(1)1}\cdots a_{\sigma(i)j}\cdots a_{\sigma(j)i}\cdots a_{\sigma(n)n}\big)\\ =&\sum_{\sigma\in S_n}\big(\mathrm{sign}(\sigma(1)\cdots\sigma(i)\cdots\sigma(j)\cdots\sigma(n))a_{\sigma(1)1}\cdots a_{\sigma(i)j}\cdots a_{\sigma(j)i}\cdots a_{\sigma(n)n}\big)\\ =&\sum_{\sigma\in S_n}\big((-1)\cdot\mathrm{sign}(\sigma(1)\cdots\sigma(j)\cdots\sigma(i)\cdots\sigma(n))a_{\sigma(1)1}\cdots a_{\sigma(j)i}\cdots a_{\sigma(i)j}\cdots a_{\sigma(n)n}\big)\\ =&-\det[\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n].\end{aligned}$$

□

由上述命题可知初等行变换对行列式的影响为:

(1) 对换两行后行列式变成相反数;

(2) 倍加变换不改变行列式;

(3) 将某行倍乘 k 之后行列式也乘以 k.

初等列变换对行列式的影响完全类似. 利用这些性质, 人们可以计算高阶矩阵的行列式.

考虑三种初等行变换的表示矩阵 $T_{ij}, L_{ij}(k)$ 与 $D_i(k)$, 利用定义8.3.1可直接算出它们的行列式

$$\det(T_{ij}) = -1, \quad \det(L_{ij}(k)) = 1, \quad \det(D_i(k)) = k.$$

这样, 前述初等行变换对行列式的影响可以叙述为

$$\det(T_{ij}B) = \det(T_{ij})\cdot\det(B), \quad \det(L_{ij}(k)B) = \det(L_{ij}(k))\cdot\det(B), \quad \det(D_iB) = \det(D_i)\cdot\det(B).$$

定理 4.2.9. 设 A 是 n 阶方阵, 则 A 是可逆矩阵的充分必要条件是 $\det A \neq 0$.

证明: 可通过初等行变换 $T_1, \cdots, T_s$ 将 A 化成简化阶梯形矩阵 B, 即 $T_s \cdots T_1 A = B$, 由前述初等行变换对行列式的影响可得

$$\det(B) = \det(T_s) \cdots \det(T_1) \det(A).$$

如果 A 不可逆, 则 B 有零行, 可得 $\det(B) = 0$, 因而 $\det(A) = 0$; 如果 A 可逆, 则 B 是单位矩阵, 有 $\det(B) = 1$, 因而 $\det(A) \neq 0$.

□

定理 4.2.10. 设 A, B 是 n 阶方阵, 则有 $\det(AB) = \det(A) \cdot \det(B)$.

证明: 如果 A 不可逆, 则由 $\operatorname{rank}(AB) \leq \operatorname{rank}(A) < n$ 可知 AB 也不可逆, 此时 $\det(AB) = \det(A) \cdot \det(B) = 0$.

如果 A 可逆, 则 A 可以表示为初等矩阵的乘积 $A = T_1 \cdots T_s$, 由前述初等行变换对行列式的影响可得

$$\det(A) = \det(T_1) \cdots \det(T_s).$$

进一步, $AB = T_1 \cdots T_s B$, 可得

$$\det(AB) = \det(T_1 \cdots T_s B) = \det(T_1) \cdots \det(T_s) \det(B).$$

结合这两方面就完成了定理的证明.

□

推论 4.2.11. 设 A 是可逆矩阵, 则有 $\det(A^{-1}) = \frac{1}{\det(A)}$.

4.3 拉普拉斯展开

定义 4.3.1. 设 $A=(a_{ij})_{1\le i,j\le n}$ 是 n 阶方阵, 令 $M_{i,j}$ 为将 A 的第 i 行与第 j 列删去所得 $(n-1)\times(n-1)$ 矩阵的行列式, 称之为矩阵元 a_{ij} 的余子式; 称 $(-1)^{i+j}M_{ij}$ 为 a_{ij} 的代数余子式.

定理 4.3.2 (Laplace's expansion)**.** 设 A 是 n 阶矩阵, 则有

$$\begin{aligned}\det(A)&=\sum_{j=1}^{n}(-1)^{i+j}a_{ij}M_{i,j}\\&=\sum_{i=1}^{n}(-1)^{i+j}a_{ij}M_{i,j}.\end{aligned}$$

证明: 一方面, 利用命题4.2.6可得

$$\begin{aligned}\det(A)&=\sum_{\sigma\in S_n}\operatorname{sign}(\sigma)a_{1\sigma(1)}\cdots a_{i\sigma(i)}\cdots a_{n\sigma(n)}\\&=\sum_{j=1}^{n}\sum_{\sigma\in S_n\text{满足}\sigma(i)=j}\operatorname{sign}(\sigma)a_{1\sigma(1)}\cdots a_{ij}\cdots a_{n\sigma(n)}\\&=\sum_{j=1}^{n}a_{ij}\sum_{\text{双射}\sigma:[n]\backslash\{i\}\to[n]\backslash\{j\}}\operatorname{sign}(\sigma)a_{1\sigma(1)}\cdots a_{n\sigma(n)}.\end{aligned}\tag{4.2}$$

另一方面, 我们来直接计算 $M_{i,j}$

$$M_{i,j}=\sum_{j_1\cdots j_{i-1}j_{i+1}\cdots j_n\text{是}[n]\backslash\{j\}\text{的排列}}\operatorname{sign}(j_1\cdots j_{i-1}j_{i+1}\cdots j_n)a_{1j_1}\cdots a_{(i-1)j_{i-1}}a_{(i+1)j_{i+1}}\cdots a_{nj_n},\tag{4.3}$$

其中 $\operatorname{sign}(j_1\cdots j_{i-1}j_{i+1}\cdots j_n)=1$ 当且仅当从 $j_1\cdots j_{i-1}j_{i+1}\cdots j_n$ 变到 $1\cdots(j-1)(j+1)\cdots n$ 一共需要偶数次对换.

下面我们来比较上述两个式子中 sign 函数的关系, 注意到对于满足 $\sigma(i)=j$ 的置换 σ, 可先用 $(i-1)$ 次对换变为 $j\sigma(1)\cdots\sigma(i-1)\sigma(i+1)\sigma(n)$; 再用若干次对换 (设为 m 次) 将 $\sigma(1)\cdots\sigma(i-1)\sigma(i+1)\sigma(n)$ 变为 $1\cdots(j-1)(j+1)\cdots n$; 最后再用 $j-1$ 次对换把 $j1\cdots(j-1)(j+1)\cdots n$ 变成 $12\cdots n$. 这样, 由 sign 函数的定义可知

$$\operatorname{sign}(\sigma)=(-1)^{i-1+m+j-1},\quad \operatorname{sign}(\sigma(1)\cdots\sigma(i-1)\sigma(i+1)\sigma(n))=(-1)^m,$$

即有

$$\operatorname{sign}(\sigma)=(-1)^{i+j}\operatorname{sign}(\sigma(1)\cdots\sigma(i-1)\sigma(i+1)\sigma(n)).$$

结合(4.2)与(4.3), 就有

$$\begin{aligned}\det(A) &= \sum_{j=1}^{n} a_{ij} \sum_{\text{双射}\sigma:[n]\setminus\{i\}\to[n]\setminus\{j\}} \operatorname{sign}(\sigma) a_{1\sigma(1)}\cdots a_{n\sigma(n)} \\ &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \sum_{\text{双射}\sigma:[n]\setminus\{i\}\to[n]\setminus\{j\}} \operatorname{sign}(\sigma(1)\cdots\sigma(i-1)\sigma(i+1)\sigma(n)) a_{1\sigma(1)}\cdots a_{n\sigma(n)} \\ &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j} M_{i,j}.\end{aligned}$$

□

推论 4.3.3. 设 $A=(a_{ij})$ 是 n 阶方阵, 矩阵元 a_{ij} 的余子式为 M_{ij}. 对任何实数 $b_1,\cdots,b_n$, 设将 A 的第 i 个列向量换成 $\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ 所得的方阵为 A_i, 则有

$$\det(A_i) = \sum_{k=1}^{n} (-1)^{k+i} b_k M_{ki}.$$

证明: 此即 A_i 对其第 i 列的 Laplace 展开式. □

定义 4.3.4. 设 A 是 n 阶方阵, 定义它的代数余子式矩阵 C 为

$$C_{ij} = (-1)^{i+j} M_{ij},$$

即 C_{ij} 为 a_{ij} 的代数余子式.

称 C^T 为 A 的伴随矩阵, 记作 $\operatorname{adj}(A) = C^T$.

命题 4.3.5. 设 A 是可逆方阵, 则 A 的逆矩阵为

$$A^{-1} = \frac{1}{\det(A)} adj(A).$$

证明: 直接计算乘积矩阵 $P = \operatorname{adj}(A)\cdot A$, 对任何 $i,j\le n$, 有

$$P_{ij} = \sum_{k=1}^{n} C_{ik}^T a_{kj} = \sum_{k=1}^{n} (-1)^{i+k} M_{ki} a_{kj}.$$

由命题4.3.3, 这是将 A 的第 i 列替换成 A 的第 j 列所得矩阵的行列式. 当 $i=j$ 时, 替换前后矩阵不变; 当 $i\neq j$ 时, 替换后所得矩阵有两列相同. 由此可得

$$P_{ij} = \begin{cases} \det(A), & \text{如果} i=j, \\ 0, & \text{如果} i\neq j, \end{cases}$$

这就证明了 $\operatorname{adj}(A)\cdot A=\det(A)I_n$. □

例 4.3.6. 设 $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 是 2×2 矩阵, 则 A 可逆当且仅当 $\det(A)=ad-bc\neq 0$. 进一步, A 的逆矩阵为

$$A^{-1}=\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

推论 4.3.7 (Cramer 法则)**.** 设 A 是可逆的 n 阶方阵, 则方程组

$$A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

有唯一解

$$x_i=\frac{\det(A_i)}{\det(A)},\quad i=1,\ldots,n,$$

其中 A_i 是由将 A 的第 i 个列向量换成 $\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ 所得的方阵.

证明*:* 利用命题8.3.6, 方程组的解为

$$\mathbf{x}=A^{-1}\mathbf{b}=\frac{1}{\det(A)}\operatorname{adj}(A)\cdot\mathbf{b},$$

由此可得

$$x_i=\frac{1}{\det(A)}\sum_{k=1}^{n}\operatorname{adj}(A)_{ik}b_k=\frac{1}{\det(A)}\sum_{k=1}^{n}(-1)^{i+k}M_{ki}b_k.$$

结合命题4.3.3即有 $x_i=\frac{\det(A_i)}{\det(A)}$.

□

4.4 行列式的例子与应用

例 4.4.1. (1) 上三角方阵的行列式等于所有对角元的乘积.

(2) 设 $X=\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ 是分块上三角矩阵, 其中 A,C 是方阵, 则 $\det(X)=\det(A)\det(C)$.

(3) 正交矩阵的行列式等于 ± 1. 定义 $SO(n)=\{A\in O(n)|\det(A)=1\}$, 称之为 n 阶特殊正交群.

(4) 如果 S 是可逆矩阵, $B=S^{-1}AS$, 则有 $\det(B)=\det(A)$.

例 4.4.2 (Vandermonde 行列式)**.** 范德蒙矩阵是指如下形式的矩阵

$$V=\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix},$$

称其行列式为范德蒙行列式, 有

$$\det(V)=\prod_{1\le i<j\le n}(x_j-x_i).$$

证明: 将 $\det(V)$ 记为 $V(x_1,\cdots,x_n)$, 如果存在两个 x_i,x_j 相等, 则显然行列式为零, 以下假设 $x_1,\ldots,x_n$ 两两不同.

对最后一列做 Laplace 展开, 结合归纳假设可得

$$\begin{aligned}\det(V)&=\sum_{i=1}^{n}(-1)^{i+n}x_i^{n-1}V(x_1,\cdots,\widehat{x_i},\cdots,x_n)\\ &=\left(\prod_{k<l}(x_l-x_k)\right)\cdot\left(\sum_{i=1}^{n}x_i^{n-1}\prod_{j\ne i}\frac{1}{x_i-x_j}\right)\\ &=\prod_{k<l}(x_l-x_k),\end{aligned}$$

最后我们用到了熟知的恒等式

$$\sum_{i=1}^{n}\left(x_i^{n-1}\prod_{j\ne i}\frac{1}{x_i-x_j}\right)=1.$$

□

例 4.4.3 (平行高维体的 Volume)**.** 设 $\mathbf{a}_1,\cdots,\mathbf{a}_k$ 是 $\mathbf{R}^n$ 中 k 个线性无关的向量, 定义由它们张成的平行 k 维体为

$$P=\{\sum_{i=1}^{k}\lambda_i\mathbf{a}_i|0\le\lambda_i\le 1,\quad\forall 1\le i\le k\}.$$

证明: P 的 k 维 Volume 为

$$\mathrm{Volume}(P)=\sqrt{\det(A^TA)},$$

其中 $A=[\mathbf{a}_1,\cdots,\mathbf{a}_k]$ 是以 $\mathbf{a}_1,\cdots,\mathbf{a}_k$ 为列向量的矩阵.

特别的, 当 $k=n$ 时, 有 $\mathrm{Volume}(P)=|\det(A)|$.

证明: 对 k 归纳, 假设 k 版本命题成立, 考虑 $(k+1)$ 的情形. 记 $V = \mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_k\})$, 设 $\mathbf{a}_{k+1}$ 向 V 的正交投影为 $\mathbf{b}$, 它可以唯一的表示为

$$\mathbf{b} = \sum_{i=1}^{k} \lambda_i \mathbf{a}_i, \quad \lambda_i \in \mathbf{R},$$

且对每个 $1 \le i \le k$ 都有 $\langle \mathbf{a}_i, \mathbf{a}_{k+1} - \mathbf{b} \rangle = 0$.

我们来计算 $\det(A^T A)$, 即如下 $(k+1) \times (k+1)$ 矩阵的行列式:

$$A^T A = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_1, \mathbf{a}_k \rangle & \langle \mathbf{a}_1, \mathbf{a}_{k+1} \rangle \\ \vdots & \ddots & \vdots & \vdots \\ \langle \mathbf{a}_{k+1}, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_{k+1}, \mathbf{a}_k \rangle & \langle \mathbf{a}_{k+1}, \mathbf{a}_{k+1} \rangle \end{bmatrix}.$$

将第 $k+1$ 列减去第 1 列的 λ_1 倍, 减去第 2 列的 λ_2 倍,..., 减去第 k 列的 λ_k 倍, 矩阵行列式不变, 即有

$$\det(A^T A) = \det \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_1, \mathbf{a}_k \rangle & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \langle \mathbf{a}_k, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_k, \mathbf{a}_k \rangle & 0 \\ \langle \mathbf{a}_{k+1}, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_{k+1}, \mathbf{a}_k \rangle & \langle \mathbf{a}_{k+1}, \mathbf{a}_{k+1} - \mathbf{b} \rangle \end{bmatrix},$$

这里用到了每个 $1 \le i \le k$ 都有 $\langle \mathbf{a}_i, \mathbf{a}_{k+1} - \mathbf{b} \rangle = 0$. 注意到

$$\langle \mathbf{a}_{k+1}, \mathbf{a}_{k+1} - \mathbf{b} \rangle = \langle \mathbf{a}_{k+1} - \mathbf{b}, \mathbf{a}_{k+1} - \mathbf{b} \rangle + \langle \mathbf{b}, \mathbf{a}_{k+1} - \mathbf{b} \rangle = ||\mathbf{a}_{k+1} - \mathbf{b}||^2,$$

结合前述列变换所得矩阵行列式的 Laplace 展开可得

$$\det(A^T A) = ||\mathbf{a}_{k+1} - \mathbf{b}||^2 \cdot \det\left([\langle \mathbf{a}_i, \mathbf{a}_j \rangle]_{i,j \le k}\right). \tag{4.4}$$

另一方面, 可将由 $\mathbf{a}_1, \cdots, \mathbf{a}_{k+1}$ 张成的平行 $k+1$ 维体 P 的底视为由 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 张成的平行 k 维体, 高恰好为 $||\mathbf{a}_{k+1} - \mathbf{b}||$. 这样, P 的 $k+1$ 维 Volume 等于底的 k 维 Volume 乘以高, 利用归纳假设即得

$$\mathrm{Volume}(P) = \sqrt{\det\left([\langle \mathbf{a}_i, \mathbf{a}_j \rangle]_{i,j \le k}\right)} \cdot ||\mathbf{a}_{k+1} - \mathbf{b}||,$$

对比(4.4), 这就证明了 $\mathrm{Volume}(P) = \sqrt{\det(A^T A)}$, 从而完成了归纳证明.

当 $k = n$ 时, 有

$$\mathrm{Volume}(P) = \sqrt{\det(A^T A)} = \sqrt{\det(A^T) \cdot \det(A)} = |\det(A)|.$$

□

注 4.4.4. 设 n 阶方阵 A 给出的线性映射为 $f: \mathbf{R}^n \to \mathbf{R}^n$, 则 f 把 $\mathbf{R}^n$ 的一个基底 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 映为 $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_n\}$. 设 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 与 $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_n\}$ 张成的平行 n 维体的 Volume 分别为 P, Q, 则有

$$\mathrm{Volume}(Q) = \big|\det([A\mathbf{v}_1, \cdots, A\mathbf{v}_n])\big| = \big|\det(A[\mathbf{v}_1, \cdots, \mathbf{v}_n])\big| = |\det(A)| \cdot \mathrm{Volume}(P),$$

即映射 f 变换前后 n 维平行体的 Volume 扩大 $|\det(A)|$ 倍, 基于这个原因, 可把 $\det(A)$ 称为体积变化倍数, 这也是在多重积分换元公式中出现 $|\det|$ 的缘故.

例 4.4.5 (Jacobi 公式)**.** 设 n 阶方阵 $A(t)$ 的矩阵元 $a_{ij}(t)$ 都是关于 t 的可导函数, 则有

$$\frac{d}{dt}\left(\det(A(t))\right) = \sum_{1 \le i,j \le n} (-1)^{i+j} M_{ij} \frac{da_{ij}}{dt} = \mathrm{trace}\left(\mathrm{adj}(A) \cdot \frac{dA}{dt}\right).$$

第五章　谱理论

5.1 Motivation

设观测者 O 关心某个线性映射 $f : \mathbf{R}^n \to \mathbf{R}^m$. 另一个观测者 O' 对定义域 $\mathbf{R}^n$ 与值域 $\mathbf{R}^m$ 都采用了新的描述方式, 则 O' 会将前述线性映射描述成新的形式. 具体的说, 设 O' 定义域中的向量 $\mathbf{x}'$ 对应于 O 定义域中的向量 $\mathbf{x} = s(\mathbf{x}')$, O' 值域中的向量 $\mathbf{y}'$ 对应于 O 值域中的向量 $\mathbf{y} = t(\mathbf{y}')$, 其中 $s : \mathbf{R}^n \to \mathbf{R}^n$ 与 $t : \mathbf{R}^m \to \mathbf{R}^m$ 都是线性同构, 则 O' 会将 O 的线性映射描述成 $f' = t^{-1} \circ f \circ s$. 这样, f 与 f' 的差异完全来自于对定义域与值域采用了不同的描述方式而已, 可以认为这两个线性映射有着相同的内涵. 如果用矩阵语言叙述, 设 f, f', s, t 的矩阵表示分别为 A, B, S, T, 则有 $B = T^{-1}AS$. 在第二章我们称满足这种关系的 A, B 互为相抵的矩阵, 并在命题2.4.11中证明了每个矩阵 A 可以相抵于一个具有最简形式的矩阵 $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

进一步, 如果观测者 O 关心 $\mathbf{R}^n$ 到自身的线性映射 $f : \mathbf{R}^n \to \mathbf{R}^n$, 设观测者 O' 对 $\mathbf{R}^n$ 采用新的描述方式, 即 O' 的向量 $\mathbf{x}'$ 对应于 O 的向量 $\mathbf{x} = s(\mathbf{x}')$, 则 O' 会将 O 的线性映射描述成 $f' = s^{-1} \circ f \circ s$, 相应的它们的表示矩阵满足 $B = S^{-1}AS$. 人们称满足这种关系的 A, B 互为相似的矩阵, 并希望将每个矩阵 A 相似于一个具有最简形式的矩阵.

对角矩阵是最简单的矩阵. 对于方阵 A, 我们希望找到可逆矩阵 S, 使得 $S^{-1}AS = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$ 是对角矩阵, 称之为 A 的对角化表示. 令 $\{\mathbf{e}_i\}$ 为 $\mathbf{R}^n$ 的标准基, 则有 $S^{-1}AS\mathbf{e}_i = \lambda_i \mathbf{e}_i$. 记 $\mathbf{x}_i = S\mathbf{e}_i$, 可得 $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$, 称这样的向量 $\mathbf{x}_i$ 为 A 的特征向量, 称 λ_i 为对应的特征值.

5.2 特征值与特征向量

定义 5.2.1. 设 A 是 n 阶实矩阵, 称实数 λ 为 A 的一个特征值 (*eigenvalue*), 如果存在非零向量 $\mathbf{x} \in \mathbf{R}^n$ 使得 $A\mathbf{x} = \lambda \mathbf{x}$. 称满足此条件的非零向量 $\mathbf{x}$ 为 A 的一个属于特征值 λ 的特征向量 (*eigenvector*).

命题 5.2.2. 设 A 是 n 阶实矩阵, 则实数 λ 是 A 的一个特征值的充分必要条件是 $\det(\lambda I_n -$

$A) = 0$.

证明: λ 是 A 的特征值, 当且仅当齐次线性方程组 $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$ 有非零解, 这等价于方阵 $\lambda I_n - A$ 不可逆, 也等价于其行列式等于零. □

定义 5.2.3. 对 n 阶方阵 A, 令

$$p_A(\lambda) = \det(\lambda I_n - A) = \det \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix},$$

称之为矩阵 A 的特征多项式, 它是 λ 的 n 次多项式, 首项系数为 1.

例 5.2.4. 考虑平面的旋转 $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, 其特征多项式为

$$p_A(\lambda) = \det \begin{bmatrix} \lambda - \cos\theta & \sin\theta \\ -\sin\theta & \lambda - \cos\theta \end{bmatrix} = \lambda^2 - 2(\cos\theta)\lambda + 1.$$

当 $\cos\theta \neq \pm 1$ 时, 特征多项式的根为复根 $\cos\theta \pm i\sin\theta$; 当 $\cos\theta = \pm 1$ 时 (对应的线性映射为恒同或角度 π 的旋转), 特征多项式才有实根.

从上述例子可以看出, 对于实方阵 A, 其特征多项式的根不一定都是实数. 另一方面, 由代数基本定理, 任何非常值的复系数多项式都有复根, 进而可知 $n(n \geq 1)$ 次复系数多项式共有 n 个复根 (计重数). 这样, 在复数域中考虑特征多项式的根更为简单, 换句话说需要考虑矩阵 A 在复数域上的特征值与特征向量.

令 $M_n(\mathbf{R})$ 为所有 n 阶实矩阵所构成的集合, $M_n(\mathbf{C})$ 为所有 n 阶复矩阵所构成的集合, 则 $M_n(\mathbf{R}) \subset M_n(\mathbf{C})$.

定义 5.2.5. 设 $A \in M_n(\mathbf{C})$, 称 $\lambda \in \mathbf{C}$ 是 A 的 (复) 特征值, 如果存在非零向量 $\mathbf{x} \in \mathbf{C}^n$ 使得 $A\mathbf{x} = \lambda\mathbf{x}$. 称满足此条件的非零向量 $\mathbf{x}$ 为 A 的一个属于特征值 λ 的 (复) 特征向量 (*eigenvector*).

命题 5.2.6. *设 $A \in M_n(\mathbf{C})$, 则 $\lambda \in \mathbf{C}$ 是 A 的 (复) 特征值的充分必要条件是 $\det(\lambda I_n - A) = 0$, 即 λ 是 A 的特征多项式的根.*

证明: 复数域上可以做 Gauss 消元法, 由此可知 $\mathbf{C}$ 上的齐次线性方程组 $B\mathbf{x} = \mathbf{0}$ 有非零解 $\mathbf{x} \in \mathbf{C}^n$ 的充分必要条件是 $\det(B) = 0$. □

总结一下, 对于 $A \in M_n(\mathbf{C})$, A 的复特征值就是特征多项式 $p_A(\lambda)$ 的复根; 对于 $A \in M_n(\mathbf{R})$, A 的实特征值就是特征多项式 $p_A(\lambda)$ 的实根.

例 5.2.7. 上三角矩阵的特征值就是其所有对角元素, 这是因为对上三角矩阵 A, 它的特征多项式为 $p_A(\lambda) = \prod\limits_{i=1}^{n}(\lambda - a_{ii})$.

定理 5.2.8 (Gershgorin 圆盘定理)**.** *设 $A = [a_{ij}] \in M_n(\mathbf{C})$, 定义如下 n 个圆盘*

$$G_i(A) = \left\{z \in \mathbf{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\right\},$$

则 A 的任何特征值都属于上述 n 个圆盘的并集.

证明: 用反证法, 假设 A 的特征值 z 不属于 $\bigcup\limits_{i=1}^{n} G_i$, 则有

$$|z - a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad \forall 1 \leq i \leq n.$$

上式表明 $zI_n - A$ 是对角占优矩阵, 由定理1.8.11可知它是可逆矩阵, 因而有 $\det(zI_n - A) \neq 0$, 这与 z 是特征值矛盾! □

命题 5.2.9. *设 A 是 n 阶方阵, 则它的特征多项式形如*

$$p_A(\lambda) = \lambda^n - \mathit{trace}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A).$$

进一步, A 的所有 n 个特征值 (计重数) 的和等于 $\mathit{trace}(A)$, 所有 n 个特征值的乘积等于 $\det(A)$.

证明: 记 $C = \lambda I_n - A$, 则只有对角元 $C_{11}, \ldots, C_{nn}$ 含有 λ. 考虑 $\det(C)$ 的完全展开式

$$\det(C) = (\lambda - a_{11}) \cdots (\lambda - a_{nn}) + \sum_{\sigma \neq (12\ldots n) \in S_n} \mathrm{sign}(\sigma) C_{1\sigma(1)} \cdots C_{n\sigma(n)}, \tag{5.1}$$

注意到, 对于 $\sigma \neq (12\ldots n)$, 至少有两个不同指标 i 使得 $\sigma(i) \neq i$, 这是因为由 $\sigma \neq (12\ldots n)$ 知存在 $\sigma(k) = l \neq k$, 则 $\sigma(l) \neq \sigma(k) = l$. 这样, (5.1)式中后面求和式中每一项关于 λ 的次数都不超过 $(n-2)$, 由此可得 $p_A(\lambda)$ 的 λ^n, λ^{n-1} 项与 $(\lambda - a_{11}) \cdots (\lambda - a_{nn})$ 的一样, 即有

$$p_A(\lambda) = \lambda^n - (a_{11} + \cdots + a_{nn})\lambda^{n-1} + \text{更低次项}.$$

另一方面, 由 $p_A(0) = \det(-A) = (-1)^n \det(A)$ 可知 $p_A(\lambda)$ 的常数项为 $(-1)^n \det(A)$.

进一步, 假设 A 的所有 n 个特征值为 $\lambda_1, \ldots, \lambda_n$, 则有

$$p_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

考虑 λ^{n-1} 项的系数与常数项的系数, 即有

$$\lambda_1+\cdots+\lambda_n=\mathrm{trace}(A),\quad \lambda_1\cdots\lambda_n=\det(A).$$

□

5.3 对角化与谱分解

定义 5.3.1. 称 $A\in M_n(\mathbf{C})$ 是可对角化的, 如果存在可逆复矩阵 S 使得 $S^{-1}AS$ 是对角矩阵. 进一步, 称满足如上条件的矩阵 S 对角化 A.

命题 5.3.2. *$A\in M_n(\mathbf{C})$ 可对角化的充分必要条件是 A 有 n 个线性无关的特征向量.*

证明: "⇒" 设 $S^{-1}AS=\mathrm{diag}\{d_1,\ldots,d_n\}$, 取 $\mathbf{C}^n$ 的标准基 $\{\mathbf{e}_i\}_{1\le i\le n}$, 有 $S^{-1}AS\mathbf{e}_i=d_i\mathbf{e}_i$. 令 $\mathbf{v}_i=S\mathbf{e}_i$, 则有 $A\mathbf{v}_i=d_i\mathbf{v_i}$. 结合 S 可逆, 可知 $\mathbf{v}_i$ 是 A 的特征向量, 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 线性无关.

"⇐" 设 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 A 的 n 个线性无关的特征向量, 且对每个 i 有 $A\mathbf{v_i}=\lambda_i\mathbf{v}_i$. 定义线性同构 $s:\mathbf{C}^n\to\mathbf{C}^n$ 为

$$s(\mathbf{e}_i)=\mathbf{v}_i,\quad \forall 1\le i\le n,$$

易知 $s^{-1}\circ A\circ s(\mathbf{e}_i)=\lambda_i\mathbf{e}_i$. 设 s 的表示矩阵为 S(即有 $S=[\mathbf{v}_1,\ldots,\mathbf{v}_n]$), 则有 $S^{-1}AS=\mathrm{diag}\{\lambda_1,\ldots,\lambda_n\}$. □

命题 5.3.3. *属于不同特征值的特征向量线性无关. 具体的说, 设 $A\in M_n(\mathbf{C})$, $\mathbf{v}_1,\ldots,\mathbf{v}_m$ 是 A 的特征向量, 且属于不同的特征值, 则 $\mathbf{v}_1,\ldots,\mathbf{v}_m$ 线性无关.*

证明: 用反证法, 假设 $\mathbf{v}_1,\ldots,\mathbf{v}_m$ 线性相关, 则其极大线性无关组的向量个数小于 m. 不妨设 $\mathbf{v}_1,\ldots,\mathbf{v}_s$ 是极大线性无关组, 则有 $s<m$. 注意到, $\mathbf{v}_1,\ldots,\mathbf{v}_s,\mathbf{v}_{s+1}$ 线性相关, 存在不全为零的复数 $k_1,\ldots,k_s,k_{s+1}$ 使得

$$k_1\mathbf{v}_1+\cdots+k_s\mathbf{v}_s+k_{s+1}\mathbf{v}_{s+1}=\mathbf{0}.\tag{5.2}$$

设 $A\mathbf{v_i}=\lambda_i\mathbf{v}_i$, 则 $\lambda_1,\ldots,\lambda_m$ 两两不同. 将 A 作用在上式两边, 有

$$k_1\lambda_1\mathbf{v}_1+\cdots+k_s\lambda_s\mathbf{v}_s+k_{s+1}\lambda_{s+1}\mathbf{v}_{s+1}=\mathbf{0}.$$

用加减消元法消去 $\mathbf{v}_{s+1}$, 可得

$$k_1(\lambda_1-\lambda_{s+1})\mathbf{v}_1+\cdots+k_s(\lambda_s-\lambda_{s+1})\mathbf{v}_s=\mathbf{0}.$$

由于 $\lambda_1,\ldots,\lambda_s$ 是线性无关, 故

$$k_1(\lambda_1-\lambda_{s+1})=\cdots=k_s(\lambda_s-\lambda_{s+1})=0,$$

结合 $\lambda_1,\dots,\lambda_m$ 两两不同可知 $k_1=\cdots=k_s=0$. 由于 $k_1,\dots,k_s,k_{s+1}$ 不全为零, 只能 $k_{s+1}\neq 0$, 代入(5.2)得 $\mathbf{v}_{s+1}=\mathbf{0}$, 矛盾!

□

推论 5.3.4. 如果 $A\in M_n(\mathbf{C})$ 有 n 个不同的特征值, 则 A 可对角化.

证明: 对每个特征值, 取出一个特征向量, 共取出 n 个特征向量, 它们属于不同的特征值, 由命题8.4.2知它们线性无关, 再利用命题5.3.2即知 A 可对角化. □

对一般的方阵 $A\in M_n(\mathbf{C})$, 其特征多项式可能有重根. 如果 $p_A(\lambda)$ 分解为

$$p_A(\lambda)=(\lambda-z_1)^{n_1}\cdots(\lambda-z_r)^{n_r},$$

其中 $z_1,\dots,z_r$ 是不同的复数, $n_1,\dots,n_l$ 是正整数, 满足 $n_1+\ldots+n_r=n$, 则称 z_i 是 $p_A(\lambda)$ 的 n_i 重根, 并称 z_i 作为 A 的特征值的代数重数为 n_i, 或称 z_i 是 A 的 n_i 重特征根, 也称 A 的 1 重特征值为单特征值.

对 A 的特征值 λ, 令

$$V_\lambda=\{\mathbf{x}\in\mathbf{C}^n|A\mathbf{x}=\lambda\mathbf{x}\},$$

它是 $\mathbf{C}^n$ 的子空间, 称之为 A 的属于 λ 的特征子空间. 这样, $\mathbf{x}$ 是 A 的属于特征值 λ 的特征向量当且仅当 $\mathbf{x}\in V_\lambda\backslash\{\mathbf{0}\}$. 可以通过在复数域上求解齐次线性方程

$$(\lambda I_n-A)\mathbf{x}=\mathbf{0},$$

来确定出 V_λ.

命题 5.3.5. 不同特征子空间的和是直和. 具体的说, 设 $\lambda_1,\dots,\lambda_r$ 是 A 的不同特征值, 则有

$$V_{\lambda_1}+\cdots+V_{\lambda_r}=V_{\lambda_1}\oplus\cdots\oplus V_{\lambda_r}.$$

证明: 只需证明任何 $\mathbf{x}\in V_{\lambda_1}+\cdots+V_{\lambda_r}$ 只有唯一的表示

$$\mathbf{x}=\mathbf{v}_1+\cdots+\mathbf{v}_r,\quad \mathbf{v}_i\in V_{\lambda_i},\quad \forall 1\le i\le r.$$

这是因为, 如果

$$\mathbf{x}=\mathbf{v}_1+\cdots+\mathbf{v}_r=\mathbf{w}_1+\cdots+\mathbf{w}_r,\quad \mathbf{v}_i,\mathbf{w}_i\in V_{\lambda_i},\quad \forall 1\le i\le r,$$

且 $(\mathbf{v}_1,\dots,\mathbf{v}_r)\neq(\mathbf{w}_1,\dots,\mathbf{w}_r)$, 则有

$$\sum_{i,\mathbf{v}_i\neq\mathbf{w}_i}(\mathbf{v}_i-\mathbf{w}_i)=\mathbf{0},$$

但是 $\{\mathbf{v}_i - \mathbf{w}_i | \mathbf{v}_i \neq \mathbf{w}_i\}$ 是属于不同特征值的特征向量, 由命题8.4.2知此向量组线性无关, 与上式矛盾!

□

定义 5.3.6. 设 λ 是 $A \in M_n(\mathbf{C})$ 的特征值, 称特征子空间 V_λ(在复数域 $\mathbf{C}$ 上) 的维数 $\dim(V_\lambda)$ 为 λ 作为 A 的特征值的几何重数.

命题 5.3.7. *特征值的几何重数不超过其代数重数.*

证明: 设 λ_0 是 A 的特征值, 几何重数为 q. 取 V_{λ_0} 的基 $\mathbf{v}_1, \ldots, \mathbf{v}_q$, 并将它扩充为 $\mathbf{C}^n$ 的基 $\mathbf{v}_1, \ldots, \mathbf{v}_q, \mathbf{v}_{q+1}, \ldots, \mathbf{v}_n$. 设对每个 $1 \leq j \leq n$ 有

$$A\mathbf{v}_j = \sum_{k=1}^{n} c_{kj} \mathbf{v}_k.$$

写成矩阵的形式, 即

$$A[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}.$$

记 $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, $C = (c_{kj})_{1 \leq k,j \leq n}$, 由 $\mathbf{v}_1, \ldots, \ldots, \mathbf{v}_n$ 是 $\mathbf{C}^n$ 的基可知 S 可逆, 结合上式可知 $A = SCS^{-1}$, 由此可得 A 的特征多项式为

$$p_A(\lambda) = \det(\lambda I_n - A) = \det(S(\lambda I_n - C)S^{-1}) = \det(\lambda I_n - C).$$

注意到

$$\lambda I_n - C = \begin{pmatrix} (\lambda - \lambda_0) I_q & * \\ O & * \end{pmatrix}$$

是分块上三角矩阵, 可得 $p_A(\lambda)$ 含有因子 $\det((\lambda - \lambda_0) I_q) = (\lambda - \lambda_0)^q$. 特别的, λ_0 的代数重数至少是 q, 大于等于 λ_0 的几何重数. □

在上述证明中, 我们建立了如下结果.

命题 5.3.8. *设 S 是可逆方阵, $B = S^{-1}AS$, 则 A, B 的特征多项式相同.*

定义 5.3.9. 称 $A \in M_n(\mathbf{C})$ 的特征值 λ_0 为半单特征值, 如果 λ_0 的几何重数等于代数重数, 否则称 λ_0 为亏损特征值.

定理 5.3.10. (1) *复方阵可对角化的充分必要条件是它的所有特征值都是半单的.*

(2) *实方阵 (在 $\mathbf{R}$ 上) 可对角化的充分必要条件是其特征多项式的根都是实数, 且其特征值都半单.*

证明: (1) “$\Rightarrow$” 设 $S^{-1}AS=\mathrm{diag}\{d_1,\ldots,d_n\}$, 则

$$p_A(\lambda)=(\lambda-d_1)\cdots(\lambda-d_n).$$

设 $d_1,\ldots,d_n$ 按值分类共有 n_1 个 λ_1,..., n_r 个 λ_r, 则 A 的全部特征值为 $\lambda_1,\ldots,\lambda_r$, 且 λ_i 的代数重数为 n_i. 对每个 $1\le i\le r$, 令

$$W_i=\mathrm{span}(\{S\mathbf{e}_j|d_j=\lambda_i\}),$$

则有 $\dim(W_i)=n_i$ 且 $W_i\subseteq V_{\lambda_i}$, 由此可得 $n_i\le\dim(V_{\lambda_i})$. 但由命题8.5.2可知 $\dim(V_{\lambda_i})\le n_i$, 因而 $\dim(V_{\lambda_i})=n_i$, 这就证明了 A 的每个特征值都是半单的.

“$\Leftarrow$” 设 A 的所有特征值为 $\lambda_1,\ldots,\lambda_r$. 设 λ_i 的代数重数与几何重数等于 n_i, 则 $\sum\limits_{i=1}^{r}n_i=n$. 对每个 i, 取 V_{λ_i} 的一组基 S_i, 它由 n_i 个属于特征值 λ_i 的特征向量组成. 我们来证明向量组 $\bigcup\limits_{i=1}^{r}S_i$ 线性无关, 由此就找到了 A 的 $\sum\limits_{i=1}^{r}n_i=n$ 个线性无关的特征向量, 由命题5.3.2 可知 A 可以对角化. 为了证明前述断言, 设 $S_i=\{\mathbf{v}_{ij}:1\le j\le n_i\}$, 设 $k_{ij},1\le i\le r,1\le j\le n_i$ 是满足如下等式的任何一组常数:

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}k_{ij}\mathbf{v}_{ij}=\mathbf{0},$$

令 $\mathbf{w}_i=\sum\limits_{j=1}^{n_i}k_{ij}\mathbf{v}_{ij}$, 则 $\mathbf{w}_i\in V_{\lambda_i}$. 由命题5.3.5可知每个 $\mathbf{w}_i$ 都等于零向量. 再由 S_i 是 V_{λ_i} 的基可得所有 $k_{i1},\ldots,k_{in_i}$ 都等于零. 这就验证了向量组 $\bigcup\limits_{i=1}^{r}S_i$ 线性无关.

(2) “$\Rightarrow$” 设 $A\in M_n(\mathbf{R})$ 在 $\mathbf{R}$ 上可对角化, 即存在可逆实矩阵 S 使得 $S^{-1}AS=\mathrm{diag}\{d_1,\ldots,d_n\}$, 其中 d_i 都是实数. 由第 (1) 部分 “$\Rightarrow$” 的证明, A 的全部特征值为 $d_1,\cdots,d_n$ 都是实数, 且都是半单的.

“$\Leftarrow$” 设实方阵 A 的 (不同的) 特征值 $\lambda_1,\ldots,\lambda_r$ 都是实数. 设 λ_i 的代数重数与几何重数等于 n_i, 则 $\sum\limits_{i=1}^{r}n_i=n$. 对每个 $i\le r$, 利用 gauss 消元法, 可用初等行变换把实方阵 λ_iI_n-A 化为阶梯行实矩阵 B, 设 B 的阶梯行数为 t, 则在 $\mathbf{C}$ 上求解方程 $(\lambda_iI_n-A)\mathbf{x}=\mathbf{0}$ 所得解空间 V_{λ_i} 的维数 (作为 $\mathbf{C}$ 上的维数) 等于 $n-t$, 由几何重数的定义可知 $n-t=n_i$. 现在考虑在 $\mathbf{R}$ 上求解方程 $(\lambda_iI_n-A)\mathbf{x}=\mathbf{0}$, 设其解空间为

$$W_{\lambda_i}=\{\mathbf{x}\in\mathbf{R}^n:(\lambda_iI_n-A)\mathbf{x}=\mathbf{0}\},$$

则 W_{λ_i} 作为 $\mathbf{R}$ 上线性空间的维数等于 $n-t=n_i$. 取 W_{λ_i} 的基 T_i, 与第 (1) 部分 “$\Leftarrow$” 的证明一样, 可知向量组 $\bigcup\limits_{i=1}^{r}T_i$ 在 $\mathbf{R}$ 上线性无关. 结合命题5.3.2的证明过程可知 A 在 $\mathbf{R}$ 上可以对角化. $\square$

5.4 相似矩阵

定义 5.4.1. 设 A, B 是 n 阶方阵, 称 A 与 B 相似, 或 A 相似于 B, 如果存在 n 阶可逆矩阵 S 使得 $S^{-1}AS = B$.

命题 5.4.2. *方阵的相似是等价关系, 即满足:*

(1) *自反性, 即 A 与 A 相似;*

(2) *对称性, 即如果 A 与 B 相似, 则 B 与 A 相似;*

(3) *传递性, 即如果 A 与 B 相似且 B 与 C 相似, 则 A 与 C 相似.*

命题 5.4.3. *设方阵 A 与 B 相似, 则有*

(1) *A, B 的特征多项式相同.*

(2) *A 与 B 的特征值一样, 且每个特征值作为 A 特征值的几何 (代数) 重数等于作为 B 特征值的几何 (代数) 重数.*

证明: 设 $B = S^{-1}AS$, 其中 S 是可逆矩阵, 则有

$$p_B(\lambda) = \det(\lambda I_n - B) = \det(S^{-1}(\lambda I_n - A)S) = \det(\lambda I_n - A) = p_A(\lambda).$$

由此可知 A, B 的特征值完全一样, 且每个特征值的代数重数也一样. 注意到, 对每个特征值 λ 有

$$B\mathbf{w} = \lambda\mathbf{w} \iff S^{-1}AS\mathbf{w} = \lambda\mathbf{w} \iff A(S\mathbf{w}) = \lambda S\mathbf{w},$$

则 S 建立了从 B 的 λ 特征子空间到 A 的 λ 特征子空间的线性同构, 由此可知 λ 作为 A 特征值的几何重数等于作为 B 特征值的几何重数.

□

命题 5.4.4. *任何复方阵都相似于上三角矩阵.*

证明: 设 $A \in M_n(\mathbf{C})$, 由代数基本定理可知 A 的特征多项式 $p_A(\lambda)$ 一定有复根, 即 A 一定有特征根. 取 A 的一个特征值 d_1, 设 $\mathbf{v}_1$ 是属于特征值 d_1 的特征向量. 将 $\mathbf{v}_1$ 扩充为 $\mathbf{C}^n$ 的基 $\mathbf{v}_1, \ldots, \mathbf{v}_n$, 设

$$A\mathbf{v}_j = \sum_{i=1}^{n} c_{ij}\mathbf{v}_i, \quad \forall 1 \leq j \leq n.$$

令 $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, 注意到 $A\mathbf{v}_1 = d_1\mathbf{v}_1$, 有

$$AS = S\begin{pmatrix} d_1 & \beta \\ O & C \end{pmatrix},$$

其中 $\beta=(c_{12},\dots,c_{1n})$ 是 $1\times(n-1)$ 矩阵, $C=(c_{ij})_{2\le i,j\le n}$ 是 $(n-1)$ 阶复矩阵. 上式表明 A 相似于 $\begin{pmatrix} d_1 & \beta \\ O & C \end{pmatrix}$. 进一步由归纳假设, C 相似于上三角矩阵, 存在可逆矩阵 T 与上三角矩阵 U 使得 $T^{-1}CT=U$, 由此可得

$$\begin{pmatrix} 1 & O \\ O & T^{-1} \end{pmatrix}\begin{pmatrix} d_1 & \beta \\ O & C \end{pmatrix}\begin{pmatrix} 1 & O \\ O & T \end{pmatrix}=\begin{pmatrix} d_1 & \beta T \\ O & T^{-1}CT \end{pmatrix}=\begin{pmatrix} d_1 & \beta T \\ O & U \end{pmatrix},$$

这表明 $\begin{pmatrix} d_1 & \beta \\ O & C \end{pmatrix}$ 相似于上三角矩阵. 结合起来就证明了 A 相似于三上角矩阵. □

定理 5.4.5 (Cayley-Hamilton)**.** *设方阵 A 的特征多项式为 $p_A(\lambda)$, 则有 $p_A(A)=O$.*

证明: 由命题5.4.4, 存在可逆矩阵 T 与上三角矩阵 U 使得 $T^{-1}AT=U$, 从而有 $p_A(\lambda)=p_U(\lambda)$, 由此可得

$$p_A(A)=p_U(A)=p_U(TUT^{-1})=Tp_U(U)T^{-1}.$$

只需验证对于上三角矩阵 U 有 $p_U(U)=O$. 注意到, $p_U(\lambda)=(\lambda-u_{11})\cdots(\lambda-u_{nn})$, 只需验证

$$(U-u_{11}I_n)\cdots(U-u_{nn}I_n)=O.$$

设 $\{\mathbf{e}_i\}_{1\le i\le n}$ 是 $\mathbf{C}^n$ 的标准基, 则有

$$U\mathbf{e}_k=u_{1k}\mathbf{e}_1+\cdots+u_{kk}\mathbf{e}_k,$$

从而有

$$(U-u_{kk}I_n)\mathbf{e}_k=u_{1k}\mathbf{e}_1+\cdots+u_{(k-1)k}\mathbf{e}_{k-1}. \tag{5.3}$$

我们用归纳法证明对每个 $k\le n$, 有

$$(U-u_{11}I_n)\cdots(U-u_{kk}I_n)\mathbf{e}_k=\mathbf{0}.$$

由(5.3)可知 $k=1$ 版本的断言成立. 对 $k>1$, 由(5.3)可得

$$(U-u_{11}I_n)\cdots(U-u_{kk}I_n)\mathbf{e}_k=(U-u_{11}I_n)\cdots(U-u_{(k-1)(k-1)}I_n)\left(u_{1k}\mathbf{e}_1+\cdots+u_{(k-1)k}\mathbf{e}_{k-1}\right),$$

再用归纳假设即得上式右边等于 $\mathbf{0}$, 从而完成了断言的证明.

注意到, 前述断言表明

$$(U-u_{11}I_n)\cdots(U-u_{nn}I_n)$$

作用在每个 $\mathbf{e}_i(1\le i\le n)$ 上都等于 $\mathbf{0}$, 因此有

$$(U-u_{11}I_n)\cdots(U-u_{nn}I_n)=O.$$

□

Cayley-Hamilton 定理的另一证明. 设 A 给出的线性映射为 $\varphi : \mathbf{R}^n \to \mathbf{R}^n$, 取 $\mathbf{R}^n$ 的标准基 $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, 则有

$$\varphi(\mathbf{e}_j) = a_{1j}\mathbf{e}_1 + \cdots + a_{nj}\mathbf{e}_n,$$

可把上式改写为

$$-a_{1j}\mathbf{e}_1 - \cdots + (\varphi - a_{jj})\mathbf{e}_j - \cdots - a_{nj}\mathbf{e}_n = \mathbf{0}.$$

进一步可写成矩阵乘法的形式:

$$\begin{bmatrix} \varphi - a_{11} & -a_{21} & \cdots & -a_{n1} \\ -a_{12} & \varphi - a_{22} & \cdots & -a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{1n} & -a_{2n} & \cdots & \varphi - a_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{5.4}$$

令 λ 为未知元, 考虑如下矩阵

$$B = \begin{bmatrix} \lambda - a_{11} & -a_{21} & \cdots & -a_{n1} \\ -a_{12} & \lambda - a_{22} & \cdots & -a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{1n} & -a_{2n} & \cdots & \lambda - a_{nn} \end{bmatrix},$$

它的 ij 矩阵元都是 λ 的函数, 记为 $B_{ij}(\lambda)$. 这样, 可把(5.4)改写成

$$\sum_{j=1}^{n} B_{kj}(\varphi)\mathbf{e}_j = \mathbf{0}. \tag{5.5}$$

设 B 的伴随矩阵为 C, 记 C 的 ij 矩阵元为 $C_{ij}(\lambda)$. 注意到 $B_{ij}(\lambda), C_{ij}(\lambda)$ 都是 λ 的多项式, 且 B 的行列式为 $\det(B) = p_A(\lambda)$. 进一步, 由命题8.3.6可得

$$CB = \det(B)I_n = p_A(\lambda)I_n,$$

即对任何指标 i, j 有

$$\sum_{k=1}^{n} C_{ik}(\lambda)B_{kj}(\lambda) = \delta_{ij}p_A(\lambda).$$

这是关于 λ 的多项式之间的恒等式, 把上述等式中的 λ 替换成 φ, 则有等式

$$\sum_{k=1}^{n} C_{ik}(\varphi)B_{kj}(\varphi) = \delta_{ij}p_A(\varphi). \tag{5.6}$$

结合(5.5)与(5.6), 可知对任何指标 i 有

$$\begin{aligned}\mathbf{0} &= \sum_{k=1}^{n} C_{ik}(\varphi)\left(\sum_{j=1}^{n} B_{kj}(\varphi)\mathbf{e}_j\right)\\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} C_{ik}(\varphi)B_{kj}(\varphi)\mathbf{e}_j\\ &= \sum_{j=1}^{n}\left(\sum_{k=1}^{n} C_{ik}(\varphi)B_{kj}(\varphi)\right)\mathbf{e}_j\\ &= \sum_{j=1}^{n}\delta_{ij}p_A(\varphi)\mathbf{e}_j\\ &= p_A(\varphi)\mathbf{e}_i.\end{aligned}$$

这表明 $p_A(\varphi)$ 作用在每个 $\mathbf{e}_i$ 上都等于零, 则它是从 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 的零映射, 其表示矩阵等于零矩阵, 即有 $p_A(A) = O$.

□

定义 5.4.6. 对复数 λ 与正整数 m, 定义 m 阶 Jordan 块矩阵为

$$J_m(d) = \begin{pmatrix} d & 1 & & & \\ & d & 1 & & \\ & & \ddots & \ddots & \\ & & & d & 1 \\ & & & & d \end{pmatrix},$$

也称之为一个 d-Jordan 块.

Jordan 块 $J_m(d)$ 的特征多项式为

$$\det(\lambda I_m - J_m(d)) = (\lambda - d)^m,$$

故 $J_m(d)$ 有唯一的特征值 d, 其代数重数为 m. 另一方面, 特征值 d 的特征子空间为

$$V_d = \{\mathbf{x} \in \mathbf{R}^m | (dI_m - J_m(d))\mathbf{x} = \mathbf{0}\} = \mathrm{span}(\{\mathbf{e}_1\}),$$

可得特征值 d 的几何重数为 1.

定理 5.4.7 (Jordan 标准型)**.** 每个 n 阶复方阵都相似于如下形式的矩阵

$$\begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_r}(\lambda_r) \end{bmatrix},$$

其中 $J_{n_i}(\lambda_i)$ 是 n_i 阶 *Jordan* 块矩阵, $n_1+\cdots+n_r=n$.

设 n 阶复方阵 A 相似于定理8.6.5中所述的标准型矩阵, 则有:

(1) A 的全部特征值为 n_1 个 λ_1,..., n_r 个 λ_r, 其中 $\lambda_1,\ldots,\lambda_r$不必 两两不同;

(2) A 的特征值 λ 的代数重数等于标准型中 λ-Jordan 块的阶数的总和;

(3) A 的特征值 λ 的几何重数等于标准型中 λ-Jordan 块的个数.

(4) 对 A 的特征值 λ, 可通过考察 $\dim\ker\left((A-\lambda I_n)^t\right)(t\in\mathbf{Z}_+)$ 完全确定标准型中 λ-Jordan 块的阶数信息:

$$\dim\ker\left((A-\lambda I_n)^t\right)=\sum_{1\le i\le r,\lambda_i=\lambda}\min\{t,n_i\}.$$

由此可得, A 的标准型中 $J_m(\lambda)$ 块的数目等于

$$2\dim\ker\left((A-\lambda I_n)^m\right)-\dim\ker\left((A-\lambda I_n)^{m-1}\right)-\dim\ker\left((A-\lambda I_n)^{m+1}\right),$$

其中约定 $\dim\ker\left((A-\lambda I_n)^0\right)=0$.

此定理的证明超出课程要求, 只需了解定理的结论.

定义 5.4.8. 称 n 阶复矩阵 A,B 可同时对角化, 如果存在可逆矩阵 S 使得 $S^{-1}AS$ 与 $S^{-1}BS$ 都是对角矩阵.

命题 5.4.9. *设 n 阶复矩阵 A,B 都可以对角化, 则如下三个命题彼此等价:*

(1) A,B 可同时对角化;

(2) 存在 n 个线性无关的向量, 同时是 A 与 B 的特征向量;

(3) A,B 可交换, 即 $AB=BA$.

证明: “(1) ⇒ (2) ⇒ (3)” 是直接的. 我们只需要证明 “(3) ⇒ (1)”, 假设 A,B 可交换, 且都可对角化, 来证明 A,B 可同时对角化.

设存在可逆矩阵 S 使得 $S^{-1}AS=\mathrm{diag}\{a_1,\ldots,a_n\}$, 记 $A'=\mathrm{diag}\{a_1,\ldots,a_n\}$, $B'=S^{-1}BS$, 则 A',B' 可交换. 对 A' 的每个特征值 λ, A' 的 λ 子空间为

$$V_\lambda=\mathrm{span}\{\mathbf{e}_i|a_i=\lambda\}.$$

对任何 $\mathbf{x}\in V_\lambda$, 有

$$A'(B'\mathbf{x})=B'(A'\mathbf{x})=\lambda B'\mathbf{x},$$

可知 $B'\mathbf{x}\in V_\lambda$. 这表明 B' 将 A' 的每个特征子空间 V_λ 映到 V_λ 自身. 对于 B 的每个特征值 μ, 由 B 可对角化可知 μ 的几何重数等于 μ 的代数重数. 注意到 μ 在每个 V_λ 上的几何重数不

超过它在 V_λ 上的代数重数, 且有

$$\mu\text{的代数重数} = \sum_\lambda(\mu\text{在}V_\lambda\text{上的代数重数}),\quad \mu\text{的几何重数} = \sum_\lambda(\mu\text{在}V_\lambda\text{上的几何重数}),$$

故 μ 在每个 V_λ 上的几何重数等于它在 V_λ 上的代数重数, 即 B' 在每个 V_λ 上可对角化. 这样, A', B' 在每个 V_λ 上都可以同时对角化, 由此可知 A, B 可同时对角化. □

5.5 例子与应用

例 5.5.1 (处理线性递推数列)**.** 设 $\{x_n\}_{n=1}^\infty$ 是 k 阶线性递推数列, 即满足

$$x_{n+k} = a_{k-1}x_{n+k-1} + \cdots + a_0x_n.$$

求数列的通项公式.

解. 定义向量

$$\mathbf{v}_n = \begin{pmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+k-1} \end{pmatrix},$$

则 $\{\mathbf{v}_n\}_{n=1}^\infty$ 满足递推关系 $\mathbf{v}_{n+1} = A\mathbf{v}_n$, 其中 A 是如下的 $k \times k$ 矩阵:

$$A = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ a_0 & a_1 & \cdots & a_{k-2} & a_{k-1} \end{bmatrix}.$$

由此可得

$$\mathbf{v}_{n+1} = A^n\mathbf{v}_1,$$

问题转化为计算 A 的幂 A^n. 如果 A 相似于一个更简单的矩阵 B, 即 $A = SBS^{-1}$, 则 $A^n = SB^nS^{-1}$.

考虑 A 的特征多项式

$$p_A(\lambda) = \det\begin{bmatrix} \lambda & -1 & & & \\ & \lambda & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & -1 \\ -a_0 & -a_1 & \cdots & -a_{k-2} & \lambda - a_{k-1} \end{bmatrix}.$$

我们用归纳法证明:

$$p_A(\lambda) = \lambda^k - a_{k-1}\lambda^{k-1} - \cdots - a_2\lambda^2 - a_1\lambda - a_0.$$

对第一行做 Laplace 展开, 结合归纳假设, 可得

$$\begin{aligned}
&p_A(\lambda)\\
=&\lambda \cdot \det \begin{bmatrix} \lambda & -1 & & & \\ & \lambda & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & -1 \\ -a_1 & -a_2 & \cdots & -a_{k-2} & \lambda - a_{k-1} \end{bmatrix} + \det \begin{bmatrix} 0 & -1 & & & \\ & \lambda & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & -1 \\ -a_0 & -a_2 & \cdots & -a_{k-2} & \lambda - a_{k-1} \end{bmatrix}\\
=&\lambda \cdot (\lambda^{k-1} - a_{k-1}\lambda^{k-2} - \cdots - a_2\lambda - a_1) + \mathrm{sign}(23\ldots(k-1)1)\cdot(-1)^{k-2}(-a_0)\\
=&\lambda^k - a_{k-1}\lambda^{k-1} - \cdots - a_2\lambda^2 - a_1\lambda - a_0.
\end{aligned}$$

也称上述多项式为线性递推数列 $\{x_n\}_{n=1}^{\infty}$ 的特征方程, 称其根为特征根.

(1) 如果 A 的特征值都是单特征值, 设 A 的 k 个特征值为 $\lambda_1, \ldots, \lambda_k$. 由推论5.3.4 知 A 可对角化, 即可具体找出可逆矩阵 S 使得 $A = S \cdot \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\} \cdot S^{-1}$, 这种情形下可算出

$$A^n = S \cdot \mathrm{diag}\{\lambda_1^n, \ldots, \lambda_k^n\} \cdot S^{-1}.$$

(2) 如果 $p_A(\lambda)$ 有重根, 设

$$p_A(\lambda) = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_r)^{n_r},$$

其中 $\lambda_1, \ldots, \lambda_r$ 互不相同, n_i 是正整数且 $n_1 + \cdots + n_r = k$. 由定理8.6.5知存在可逆矩阵 S 使得

$$A = S \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \vdots & \\ & & J_{n_r}(\lambda_r) \end{bmatrix} S^{-1},$$

由此可得

$$A^m = S \begin{bmatrix} (J_{n_1}(\lambda_1))^m & & \\ & \vdots & \\ & & (J_{n_r}(\lambda_r))^m \end{bmatrix} S^{-1},$$

从而将计算转化为计算 Jordan 块的幂 $(J_q(\lambda))^m$.

记 $J=J_q(\lambda)$, 则只有当 $l-k$ 等于 0 或 1 时, J_{kl} 才不是零, 结合矩阵乘法的定义式, 可得

$$(J^m)_{ij}=\sum_{i_1}\sum_{i_2}\cdots\sum_{i_{m-1}}J_{ii_1}\cdots J_{i_{m-1}j}=\sum_{i_1-i,i_2-i_1,\ldots,j-i_{m-1}\in\{0,1\}}J_{ii_1}\cdots J_{i_{m-1}j}.$$

由此可知, 当 $i>j$ 时, $(J^m)_{ij}=0$; 当 $i\le j$ 时, 记 $(i_1-i,i_2-i_1,\ldots,j-i_{m-1})=(\delta_1,\cdots,\delta_m)$, 则上述求和式是对满足条件 $\delta_1+\cdots+\delta_m=j-i$ 的所有 $0,1$ 有序组 $(\delta_1,\cdots,\delta_m)$ 求和, 当 $j-i>m$ 时, 不存在满足条件的 $0,1$ 序列, 相应的 $(J^m)_{ij}=0$. 当 $j-i\le m$ 时, 有

$$(J^m)_{ij}=\sum_{(\delta_1,\cdots,\delta_m)}\lambda^{m-(j-i)}=\binom{m}{j-i}\lambda^{m-(j-i)}=C_m^{j-i}\lambda^{m-(j-i)},$$

其中 $\binom{m}{j-i}=C_m^{j-i}$ 是组合数. 如果约定当 $p<0$ 或 $p>m$ 时 $\binom{m}{p}=C_m^p=0$, 则可把前述所有情形统一写成

$$(J^m)_{ij}=\binom{m}{j-i}\lambda^{m-(j-i)}=C_m^{j-i}\lambda^{m-(j-i)},\quad\forall i,j$$

这就完成了整个计算. □

第六章　实对称矩阵

6.1　实对称矩阵的谱分解

定义 6.1.1. 设 A 是 $m\times n$ 的复矩阵, 定义 A 的共轭转置矩阵 $\overline{A}^T$ 为

$$(\overline{A}^T)_{ij}=\overline{A_{ji}},\quad \forall 1\le i\le n, 1\le j\le m.$$

定义 6.1.2. 定义 $\mathbf{C}^n$ 上的厄米内积 $\langle\cdot,\cdot\rangle:\mathbf{C}^n\times\mathbf{C}^n\to\mathbf{C}$ 为: 对 $\mathbf{C}^n$ 中的两个向量 $\mathbf{z}=\begin{pmatrix} z_1\\ \vdots\\ z_n\end{pmatrix},\mathbf{w}=\begin{pmatrix} w_1\\ \vdots\\ w_n\end{pmatrix}$, 有

$$\langle\mathbf{z},\mathbf{w}\rangle=\sum_{i=1}^n\overline{z}_iw_i=\overline{\mathbf{z}}^T\mathbf{w},$$

其中 $\overline{\mathbf{z}}^T$ 表示 $n\times 1$ 矩阵 $\mathbf{z}$ 的共轭转置矩阵.

厄米内积满足:

(1) 关于第一个输入是共轭线性的, 即有

$$\langle k_1\mathbf{z}_1+k_2\mathbf{z}_2,\mathbf{w}\rangle=\overline{k}_1\langle\mathbf{z}_1,\mathbf{w}\rangle+\overline{k}_2\langle\mathbf{z}_2,\mathbf{w}\rangle.$$

(2) 关于第二个输入是复线性的, 即有

$$\langle\mathbf{z},k_1\mathbf{w}_1+k_2\mathbf{w}_2\rangle=k_1\langle\mathbf{z},\mathbf{w}_1\rangle+k_2\langle\mathbf{z},\mathbf{w}_2\rangle.$$

(3) 关于两个输入是共轭对称的, 即

$$\langle\mathbf{z},\mathbf{w}\rangle=\overline{\langle\mathbf{w},\mathbf{z}\rangle}.$$

命题 6.1.3. 设 A 是 $m\times n$ 复矩阵, 则对任何 $\mathbf{x}\in\mathbf{C}^n$ 与 $\mathbf{y}\in\mathbf{C}^m$, 有

$$\langle A\mathbf{x},\mathbf{y}\rangle=\langle\mathbf{x},\overline{A}^T\mathbf{y}\rangle.$$

证明: 利用厄米内积的定义直接验证即可:

$$\langle A\mathbf{x}, \mathbf{y}\rangle = \overline{A\mathbf{x}}^T\mathbf{y} = \overline{\mathbf{x}}^T\overline{A}^T\mathbf{y} = \langle \mathbf{x}, \overline{A}^T\mathbf{y}\rangle.$$

□

命题 6.1.4. 实对称矩阵的特征多项式的根都是实数, 即实对称矩阵在 $\mathbf{C}$ 上的特征值都是实数.

证明: 设 A 是 n 阶实对称矩阵, $\lambda \in \mathbf{C}$ 是它的一个特征值, 则存在非零向量 $\mathbf{x} \in \mathbf{C}^n$ 使得 $A\mathbf{x} = \lambda\mathbf{x}$, 利用命题6.1.3可得

$$\overline{\lambda}\langle \mathbf{x}, \mathbf{x}\rangle = \langle A\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, \overline{A}^T\mathbf{x}\rangle = \langle \mathbf{x}, A\mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle.$$

注意到 $\mathbf{x} \neq 0$, 则 $\langle \mathbf{x}, \mathbf{x}\rangle \neq 0$, 结合上式可知 $\overline{\lambda} = \lambda$, 即 λ 是实数. □

定理 6.1.5 (实对称矩阵的谱分解)**.** 设 A 是 n 阶实对称矩阵, 则存在实正交矩阵 Q, 使得 Q^TAQ 是实对角矩阵. 特别的, $Q^{-1}AQ$ 是实对角矩阵, 即 A 在 $\mathbf{R}$ 上可对角化.

证明: 由代数基本定理, A 的特征多项式 $p_A(\lambda)$ 存在一个复根 λ_1, 再由命题8.7.1知 $\lambda_1 \in \mathbf{R}$. 在 $\mathbf{R}$ 上考虑方程 $(\lambda_1 I_n - A)\mathbf{v} = \mathbf{0}$, 由于其系数矩阵的行列式等于零, 可知此齐次方程有非零解 $\mathbf{v}_1 \in \mathbf{R}^n$, $\mathbf{v}_1$ 是 A 的属于特征值 λ_1 的特征向量. 不妨设 $\mathbf{v}_1$ 是单位长度的, 并将它扩充为 $\mathbf{R}^n$ 的单位正交基 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. 设

$$A\mathbf{v}_j = \sum_{i=1}^n \mathbf{v}_i c_{ij}, \quad \forall 1 \leq j \leq n.$$

其中 c_{ij} 都是实数. 令 $C = (c_{ij})_{1\leq i,j\leq n}$, $Q_0 = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$, 则 Q_0 是实正交矩阵, 且有 $Q_0^TAQ_0 = C$.

注意到, 对任何 i, j, 有

$$c_{ij} = \langle A\mathbf{v}_j, \mathbf{v}_i\rangle = \langle \mathbf{v}_j, A^T\mathbf{v}_i\rangle = \langle \mathbf{v}_j, A\mathbf{v}_i\rangle = c_{ji},$$

其中 $\langle \cdot, \cdot\rangle$ 表示 $\mathbf{R}^n$ 中的内积. 上式表明 C 是实对称矩阵, 结合 $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ 可知 C 形如

$$C = \begin{bmatrix} \lambda_1 & O \\ O & D \end{bmatrix},$$

其中 $D = (c_{ij})_{2\leq i,j\leq n}$ 是 $(n-1)$ 阶实对称矩阵. 对 D 用归纳假设, 存在 $(n-1)$ 阶正交矩阵 P 使得 $P^TDP = \Lambda$ 是实对角矩阵, 由此可得

$$\begin{pmatrix} 1 & O \\ O & P^T \end{pmatrix}\begin{pmatrix} \lambda_1 & O \\ O & D \end{pmatrix}\begin{pmatrix} 1 & O \\ O & P \end{pmatrix} = \begin{pmatrix} \lambda_1 & O \\ O & P^TDP \end{pmatrix} = \begin{pmatrix} \lambda_1 & O \\ O & \Lambda \end{pmatrix}.$$

令 $Q = Q_0 \begin{pmatrix} 1 & O \\ O & P \end{pmatrix}$, 它是两个正交矩阵的乘积, 因而也是正交矩阵, 且有

$$Q^T AQ = \begin{pmatrix} 1 & O \\ O & P^T \end{pmatrix} Q_0^T AQ_0 \begin{pmatrix} 1 & O \\ O & P \end{pmatrix} = \begin{pmatrix} \lambda_1 & O \\ O & \Lambda \end{pmatrix},$$

这就完成了证明. □

由上述定理, 对任何实对称矩阵 A, 存在实正交矩阵 Q 与实对角矩阵 Λ, 使得 $Q^T AQ = \Lambda$, 亦即 $A = Q\Lambda Q^T$, 称之为实对称矩阵 A 的谱分解.

命题 6.1.6. *实对称矩阵不同特征值的 (实) 特征向量相互正交.*

证明: 设 $A\mathbf{v}_i = \lambda_i \mathbf{v}_i (i = 1, 2)$, 且 $\lambda_1 \neq \lambda_2$, 则有

$$\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle A\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, A^T \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, A\mathbf{v}_2 \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle,$$

结合 $\lambda_1 \neq \lambda_2$ 即得 $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. □

定义 6.1.7. 称实方阵 A 与 B 正交相似, 如果存在实正交矩阵 Q 使得 $Q^T AQ = B$.

这样, 可把定理8.7.2叙述为: 实对称矩阵一定正交相似于对角矩阵.

注 6.1.8. 由定理8.7.2, 对于实对称矩阵 A, 存在正交矩阵 Q 使得

$$Q^{-1} AQ = Q^T AQ = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}.$$

令 $\{\mathbf{e}_i\}$ 为 $\mathbf{R}^n$ 的标准基, 令 $\mathbf{v}_i = Q\mathbf{e}_i$, 由 Q 正交可知 $\{\mathbf{v}_i\}$ 为 $\mathbf{R}^n$ 的单位正交基, 且有 $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. 这样, 就找到了 $\mathbf{R}^n$ 的单位正交基, 其成员都是 A 的特征向量.

6.2 二次型

可以把 n 阶实对称矩阵等同成 $\mathbf{R}^n$ 上的二次函数. 对于 n 阶实对称矩阵 A, 它唯一确定如下二次函数

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \langle \mathbf{x}, A\mathbf{x} \rangle = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i<j} a_{ij} x_i x_j,$$

称之为由 A 确定的二次型. 人们经常关心 $Q(x)$ 的符号.

定义 6.2.1. 设 A 是 n 阶实对称矩阵, 令 $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ 为由 A 给出的二次型.

(1) 称 A 是正定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) > 0$;

(2) 称 A 是半正定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) \geq 0$;

(3) 称 A 是负定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) < 0$;

(4) 称 A 是半负定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) \leq 0$;

(5) 称 A 是不定的, 如果 $Q(\mathbf{x})$ 既能取正值又能取负值.

定义 6.2.2. 称方阵 A 第 i 个顺序主子阵的行列式为 A 的第 i 个顺序主子式.

定理 6.2.3 (正定矩阵的刻画)**.** *设 A 是实对称矩阵, 则以下命题彼此等价:*

(1) A 是正定矩阵;

(2) A 的特征值都是正数;

(3) 存在可逆矩阵 B 使得 $A = B^T B$;

(4) A 的顺序主子式都大于零.

证明: 利用定理8.7.2, 存在正交矩阵 P 及对角矩阵 Λ, 使得 $A = P^T \Lambda P = P^{-1}\Lambda P$, A 的所有特征值恰为 Λ 的所有对角元.

<u>“(1) ⇒ (2)”</u> 设 A 正定, 令 $\mathbf{y} = P\mathbf{x}$, 则对任何非零的 $\mathbf{x}$, 有

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T P^T \Lambda P\mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^{n} \Lambda_{ii} y_i^2 > 0.$$

当 $\mathbf{x}$ 取遍所有非零向量时, $\mathbf{y} = P\mathbf{x}$ 也取遍所有非零向量. 特别的, 取 $\mathbf{y} = \mathbf{e}_i$ 代入上式可得每个 Λ_{ii} 都是正数.

<u>“(2) ⇒ (3)”</u> 假设 A 的特征值都是正数, 即 Λ 的对角元都是正数, 定义

$$B = \mathrm{diag}\{\sqrt{\Lambda_{11}}, \ldots, \sqrt{\Lambda_{nn}}\} \cdot P,$$

则 B 是可逆矩阵, 且有

$$B^T B = \left(P^T \mathrm{diag}\{\sqrt{\Lambda_{11}}, \ldots, \sqrt{\Lambda_{nn}}\}\right) \cdot \left(\mathrm{diag}\{\sqrt{\Lambda_{11}}, \ldots, \sqrt{\Lambda_{nn}}\} P\right) = P^T \Lambda P = A.$$

<u>“(3) ⇒ (1)”</u> 设存在可逆矩阵 B 使得 $A = B^T B$, 则对任何非零向量 $\mathbf{x}$, $B\mathbf{x} \neq \mathbf{0}$. 由此可得

$$Q(x) = \langle \mathbf{x}, A\mathbf{x} \rangle = \langle \mathbf{x}, B^T B\mathbf{x} \rangle = \langle B\mathbf{x}, B\mathbf{x} \rangle > 0,$$

这表明 A 正定.

<u>“(1) ⇒ (4)”</u> 假设 A 正定, 记 A 的第 k 个顺序主子阵为 A_k, 来证 $\det(A_k) > 0$. 由于已证明命题 (1), (2) 等价, 可知 A 的所有特征值 $\lambda_1, \ldots, \lambda_n$ 都是正数, 故 $\det(A) = \lambda_1 \cdots \lambda_n > 0$. 这

就证明了正定矩阵的行列式是正的. 进一步, 对不全为零的 $x_1,\ldots,x_k$, 由 A 正定的定义可知

$$Q(x_1,\ldots,x_k,0,\ldots,0)=\sum_{1\le i\le k}\sum_{1\le j\le k}a_{ij}x_ix_j=(x_1,\ldots,x_k)A_k\begin{pmatrix}x_1\\ \vdots\\ x_k\end{pmatrix}>0,$$

这表明 A_k 是正定的, 由前述有 $\det(A_k)>0$.

"(4) ⇒ (1)" 设 A 的所有顺序主子式都大于零, 来证明 A 正定. 由假设 $a_{11}=\det(A_1)>0$, 可将 A 的二次型配方:

$$Q(\mathbf{x})=a_{11}(x_1+\frac{a_{12}}{a_{11}}x_2+\cdots+\frac{a_{1n}}{a_{11}}x_n)^2+R(x_2,\ldots,x_n),\tag{6.1}$$

其中 $R(x_2,\ldots,x_n)$ 是有关 $(n-1)$ 个变元 $x_2,\ldots,x_n$ 的二次型

$$R(x_2,\ldots,x_n)=\sum_{2\le i\le n}\sum_{2\le j\le n}(a_{ij}-\frac{a_{1i}a_{1j}}{a_{11}})x_ix_j,$$

其系数矩阵为如下的 $(n-1)$ 阶方阵

$$B=\left(a_{ij}-\frac{a_{1i}a_{1j}}{a_{11}}\right)_{2\le i,j\le n}.$$

记 B 的第 l 个顺序主子式为 B_l. 考察对 A 的如下初等行变换: 对每个 $i\ge 2$, 将 A 的第 i 行减去第 1 行的 $\frac{a_{i1}}{a_{11}}$ 倍, 则 A 变为如下分块矩阵

$$C=\begin{bmatrix}a_{11} & *\\ O & B\end{bmatrix}.$$

对每个 $k\ge 2$, C 的第 k 个顺序主子阵 C_k 是由 A_k 做若干次行倍加变换得到, 因而有

$$\det(A_k)=\det(C_k)=a_{11}\cdot\det(B_{k-1}).$$

这样, B 的所有顺序主子式都大于零, 利用归纳假设可得 B 正定, 即对不全为零的 $x_2,\ldots,x_n$ 有 $R(x_2,\ldots,x_n)>0$. 再结合(6.1)式可得 A 也正定. □

定理 6.2.4 (半正定矩阵的刻画). *设 A 是实对称矩阵, 则以下命题彼此等价:*

(1) *A 是半正定矩阵;*

(2) *A 的特征值都是非负数;*

(3) *存在方阵 B 使得 $A=B^TB$;*

证明: 仿照定理8.8.2的证明即可. □

命题 6.2.5. 设 n 阶实对称矩阵 A 是不定的, 则存在非零向量 $\mathbf{x} \in \mathbf{R}^n$ 使得 $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ 取值为零.

证明*:* 利用定理8.7.2, 存在正交矩阵 P 及对角矩阵 Λ, 使得 $A = P^T \Lambda P = P^{-1}\Lambda P$, A 的所有特征值恰为 Λ 的所有对角元. 记 $\mathbf{y} = P\mathbf{x}$, 则有

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T P^T \Lambda P\mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^{n} \Lambda_{ii} y_i^2.$$

由假设 A 不定, 则 λL 的对角元有正有负. 不妨设 $\Lambda_{11} < 0 < \Lambda_{22}$, 取 $\mathbf{y}$ 的分量为:

$$y_1 = 1, \quad y_2 = \sqrt{\frac{-\Lambda_{11}}{\Lambda_{22}}}, \quad y_3 = \cdots = y_n = 0,$$

令 $\mathbf{x} = P^{-1}\mathbf{y}$ 即可.

□

定义 6.2.6. 设 A 是 n 阶实对称矩阵, $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ 是 A 给出的二次型. 对非零向量 $\mathbf{x}$, 称

$$Q(\frac{\mathbf{x}}{||\mathbf{x}||}) = \frac{\mathbf{x}^T A\mathbf{x}}{||\mathbf{x}||^2} = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

为 $\mathbf{x}$ 关于 A 的 Rayleigh 商.

命题 6.2.7. 设 A 是 n 阶实对称矩阵, 则 $\frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ 在 $\mathbf{R}^n\backslash\{\mathbf{0}\}$ 上的最大与最小值分别等于 A 的最大与最小特征值.

证明*:* 利用定理8.7.2, 存在正交矩阵 P 及对角矩阵 Λ, 使得 $A = P^T \Lambda P = P^{-1}\Lambda P$, A 的所有特征值恰为 Λ 的所有对角元. 记 $\mathbf{y} = P\mathbf{x}$, 由 P 是正交矩阵可得 $||\mathbf{x}||^2 = ||\mathbf{y}||^2 = \sum\limits_{i=1}^{n} y_i^2$, 且有

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T P^T \Lambda P\mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \sum_{i=1}^{n} \Lambda_{ii} y_i^2.$$

注意到

$$\min\{L_{11}, \ldots, L_{nn}\} \cdot \sum_{i=1}^{n} y_i^2 \leq \sum_{i=1}^{n} \Lambda_{ii} y_i^2 \leq \max\{L_{11}, \ldots, L_{nn}\} \cdot \sum_{i=1}^{n} y_i^2,$$

即可知对非零向量 $\mathbf{x}$, Rayleigh 商 $\frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ 的值介于 A 的最小与最大特征值之间.

□

6.3 奇异值分解

回忆命题2.4.10(或者定理2.2.2的证明), 对任何线性映射 $f:\mathbf{R}^n\to\mathbf{R}^m$, 可找到 $\mathbf{R}^n$ 的基 $\{\mathbf{a}_1,\ldots,\mathbf{a}_n\}$ 与 $\mathbf{R}^m$ 的基 $\{\mathbf{b}_1,\ldots,\mathbf{b}_m\}$, 使得

$$f(\mathbf{a}_i)=\begin{cases}\mathbf{b}_i, & \text{如果} i\le r;\\ \mathbf{0}, & \text{如果} i>r;\end{cases}$$

其中 r 是 $\mathrm{Im}(f)$ 的维数. 这样, 就对 f 的定义域与值域给出了很清楚的描述: $\{\mathbf{a}_{r+1},\ldots,\mathbf{a}_n\}$ 是 $\ker(f)$ 的基; $\{\mathbf{b}_1,\ldots,\mathbf{b}_r\}$ 是 $\mathrm{Im}(f)$ 的基; 如果令

$$X=\mathrm{span}(\{\mathbf{a}_1,\ldots,\mathbf{a}_r\}),\quad Y=\mathrm{span}(\{\mathbf{b}_{r+1},\ldots,\mathbf{b}_m\}),$$

则给出直和分解

$$\mathbf{R}^n=X\oplus\ker(f),\quad \mathbf{R}^m=\mathrm{Im}(f)\oplus Y,$$

在此分解下, f 将 X 线性同构成 $\mathrm{Im}(f)$, 在 $\ker(f)$ 上恒为零. 也可用矩阵语言叙述上述结果. 设 f 的表示矩阵为 A, 令 $Q=[\mathbf{a}_1,\ldots,\mathbf{a}_n]$, $P=[\mathbf{b}_1,\ldots,\mathbf{b}_m]$, 则有

$$P^{-1}AQ=\begin{bmatrix} I_r & O\\ O & O\end{bmatrix},$$

即 A 可相抵于最简形式的矩阵 $D_r=\begin{bmatrix} I_r & O\\ O & O\end{bmatrix}$.

现在对 f 的定义域与值域赋予标准的内积结构, 人们关心能否为 $\mathbf{R}^n$ 与 $\mathbf{R}^m$ 找到合适的单位正交基, 使得在这两个基下 f 的表示形式是最简的. 这就是如下的定理.

定理 6.3.1 (奇异值分解)**.** *设 A 是 $m\times n$ 的实矩阵, 则存在 m 阶正交矩阵 U 和 n 阶正交矩阵 V, 使得 $A=U\Sigma V^T$, 其中 Σ 是如下形式的分块矩阵:*

$$\Sigma=\begin{bmatrix} diag\{\sigma_1,\ldots,\sigma_r\} & O\\ O & O\end{bmatrix},\quad \sigma_1\ge\cdots\ge\sigma_r>0.$$

证明: A^TA 是 n 阶实对称矩阵, 由定理8.7.2(或注6.1.8), 存在 A^TA 的一组特征向量 $\{\mathbf{v}_i\}_{1\le i\le n}$ 构成 $\mathbf{R}^n$ 的单位正交基. 设 $\mathbf{v}_i$ 是 A^TA 的属于特征值 λ_i 的特征向量, 适当排序后不妨设 $\lambda_1\ge\cdots\ge\lambda_n$. 对任何 $i,j\le n$, 有

$$\langle A\mathbf{v}_i,A\mathbf{v}_j\rangle=\langle\mathbf{v}_i,A^TA\mathbf{v}_j\rangle=\langle\mathbf{v}_i,\lambda_j\mathbf{v}_j\rangle=\lambda_j\delta_{ij},$$

其中 δ_{ij} 是 Kronecker 记号, 当 $i\neq j$ 时 $\delta_{ij}=0$, 当 $i=j$ 时 $\delta_{ij}=1$. 在上式中取 $i=j$ 可知 $\lambda_i\geq 0$, 且 $\lambda_i=0 \iff A\mathbf{v}_i=\mathbf{0}$. 另外, 当 $i\neq j$ 时, 有 $\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle=0$, 说明不同的 $A\mathbf{v}_i$ 彼此正交.

设 r 是满足 $\lambda_r>0$ 的最大的指标, 由前述有 $A\mathbf{v}_{r+1}=\cdots=A\mathbf{v}_n=\mathbf{0}$, 且 $\{\frac{A\mathbf{v}_1}{\sqrt{\lambda_1}},\cdots,\frac{A\mathbf{v}_r}{\sqrt{\lambda_r}}\}$ 是 $\mathbf{R}^m$ 的单位正交向量组. 令 $\mathbf{u}_i=\frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}(1\leq i\leq r)$, 并将其扩充为 $\mathbf{R}^m$ 的单位正交基 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$. 这样就有

$$A[\mathbf{v}_1,\ldots,\mathbf{v}_n]=[\mathbf{u}_1,\ldots,\mathbf{u}_m]\begin{bmatrix}\operatorname{diag}\{\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r}\} & O\\ O & O\end{bmatrix}.$$

令 $V=[\mathbf{v}_1,\ldots,\mathbf{v}_n]$, $U=[\mathbf{u}_1,\ldots,\mathbf{u}_m]$, 则 V,U 是正交矩阵, 且有

$$A=U\begin{bmatrix}\operatorname{diag}\{\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r}\} & O\\ O & O\end{bmatrix}V^T.$$

□

定义 6.3.2. 设 A 是 $m\times n$ 实矩阵, 称半正定方阵 A^TA 的特征值的非负平方根为 A 的奇异值.

推论 6.3.3. *设 A 是 $m\times n$ 实矩阵, 则 A^TA 的非零特征值与 AA^T 的非零特征值在计重数的意义下完全一样.*

证明: 由定理8.9.1, A 有奇异值分解 $A=U\Sigma V^T$, 其中 U,V 是正交矩阵. 由此可得

$$A^TA=V(\Sigma^T\Sigma)V^T,\quad AA^T=U(\Sigma\Sigma^T)U^T,$$

即 A^TA, AA^T 分别正交相似于 $\Sigma^T\Sigma,\Sigma\Sigma^T$, 从而可知 A^TA 与 AA^T 的非零特征值都是 $\sigma_1^2,\ldots,\sigma_r^2$. □

推论 6.3.4. *设 A 是 $m\times n$ 实矩阵, 则有*

(1) A^TA 是可逆矩阵当且仅当 A 是列满秩矩阵;

(2) AA^T 是可逆矩阵当且仅当 A 行满秩.

证明: 由定理8.9.1的证明可知 A^TA 的全部非零特征值为 $\sigma_1^2,\ldots,\sigma_r^2$. 由此可得 A^TA 可逆当且仅当 $r=n$, 当且仅当 $\operatorname{rank}(A)=n$, 即 A 是列满秩矩阵.

类似的, AA^T 可逆当且仅当 $r=m$, 当且仅当 $\operatorname{rank}(A)=m$, 即 A 是行满秩矩阵. □

设 A 给出的线性映射为 $f:\mathbf{R}^n\to\mathbf{R}^m$. 在定理8.9.1的证明中, 我们找到了 $\mathbf{R}^n$ 的单位正交基 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 与 $\mathbf{R}^m$ 的单位正交基 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$, 使得

$$f(\mathbf{v}_i)=\begin{cases}\sqrt{\lambda_i}\cdot\mathbf{u}_i, & \text{如果} i\leq r;\\ \mathbf{0}, & \text{如果} i>r.\end{cases}$$

这样, 就对 f 的定义域与值域给出了很清楚的描述: $\{\mathbf{v}_{r+1},\ldots,\mathbf{v}_n\}$ 是 $\ker(f)$ 的单位正交基; $\{\mathbf{u}_1,\ldots,\mathbf{u}_r\}$ 是 $\mathrm{Im}(f)$ 的单位正交基; 如果令

$$X=\mathrm{span}(\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}),\quad Y=\mathrm{span}(\{\mathbf{u}_{r+1},\ldots,\mathbf{u}_m\}),$$

则给出正交直和分解

$$\mathbf{R}^n=X\oplus\ker(f),\quad \mathbf{R}^m=\mathrm{Im}(f)\oplus Y,$$

在此分解下, f 将 X 线性同构成 $\mathrm{Im}(f)$, 在 $\ker(f)$ 上恒为零. 定理8.9.1就是用矩阵的语言来叙述这些信息.

由 $\{\mathbf{v}_i\}_{1\le i\le n}$ 是 $\mathbf{R}^n$ 的单位正交基, 任何向量 $\mathbf{x}$ 可表示成

$$\mathbf{x}=\sum_{i=1}^n\mathbf{v}_i\langle\mathbf{v}_i,\mathbf{x}\rangle,$$

由此可得

$$f(\mathbf{x})=\sum_{i=1}^r\sqrt{\lambda_i}\cdot\mathbf{u}_i\langle\mathbf{v}_i,\mathbf{x}\rangle=\sum_{i=1}^r\sqrt{\lambda_i}\cdot\mathbf{u}_i(\mathbf{v}_i^T\mathbf{x})=\left(\sum_{i=1}^r\sqrt{\lambda_i}\cdot\mathbf{u}_i\mathbf{v}_i^T\right)\mathbf{x}.$$

上式说明 $\sum\limits_{i=1}^r\sqrt{\lambda_i}\cdot\mathbf{u}_i\mathbf{v}_i^T$ 是 f 的表示矩阵, 所以有

$$A=\sum_{i=1}^r\sqrt{\lambda_i}\cdot\mathbf{u}_i\mathbf{v}_i^T=U_r\Sigma_rV_r^T,$$

其中 $U_r=[\mathbf{u}_1,\ldots,\mathbf{u}_r]$, $V_r=[\mathbf{v}_1,\ldots,\mathbf{v}_r]$, $\Sigma_r=\mathrm{diag}\{\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r}\}$. 称上式为 A 的简化奇异值分解.

如果 A 是可逆矩阵, 则 A 的奇异值分解 (SDV) 中 $r=m=n$, 且 A 的逆矩阵为

$$A^{-1}=(U\Sigma V^T)^{-1}=V\Sigma^{-1}U^T.$$

这就启发人们对一般的 (不一定可逆, 甚至不一定是方阵的) 矩阵 A 定义如下 $n\times m$ 矩阵:

$$A^+=\sum_{i=1}^r\frac{1}{\sqrt{\lambda_i}}\cdot\mathbf{v}_i\mathbf{u}_i^T=V_r\Sigma_r^{-1}U_r^T,$$

称之为 A 的 Moore-Penrose 广义逆, 简称广义逆. A^+ 满足

$$A^+A=V_rV_r^T,\quad AA^+=U_rU_r^T.$$

为了进一步解释上述等式的意义, 我们需要回顾如下两个例子.

例 6.3.5 (正交投影的表示矩阵)**.** 设 V 是 $\mathbf{R}^n$ 的子空间. 利用定理3.5.5, 可定义 $\mathbf{R}^n$ 向 V 的正交投影 $P_V:\mathbf{R}^n\to\mathbf{R}^n$. 如果为 V 选定了一组单位正交基 $\{\mathbf{q}_1,\ldots,\mathbf{q}_r\}$, 则对任何 $\mathbf{x}\in\mathbf{R}^n$ 有

$$P_V(\mathbf{x})=\sum_{i=1}^r\mathbf{q}_i\langle\mathbf{q}_i,\mathbf{x}\rangle=\sum_{i=1}^r\mathbf{q}_i(\mathbf{q}_i^T\mathbf{x})=\left(\sum_{i=1}^r\mathbf{q}_i\mathbf{q}_i^T\right)\cdot\mathbf{x},$$

由此可得正交投影映射 P_V 的表示矩阵为 $\sum\limits_{i=1}^r\mathbf{q}_i\mathbf{q}_i^T$. 记 $Q=[\mathbf{q}_1,\ldots,\mathbf{q}_r]$, 则 P_V 的表示矩阵可以表示成 QQ^T.

若只是为 V 找到一组 (不一定单位正交的) 基 $\mathbf{v}_1,\ldots,\mathbf{v}_r$, 则前述计算就会变得复杂. 比如, 为了确定 $\mathbf{a}$ 向 V 的正交投影, 设 $\mathbf{a}_V=\sum\limits_{i=1}^r x_i\mathbf{v}_i$, 则由 $\mathbf{a}-\mathbf{a}_V\perp\mathbf{v}_k$ 可得

$$\sum_{i=1}^r\langle\mathbf{v}_k,\mathbf{v}_i\rangle x_i=\langle\mathbf{v}_k,\mathbf{a}\rangle.$$

这样就等价于求解线性方程组

$$\begin{pmatrix}\langle\mathbf{v}_1,\mathbf{v}_1\rangle & \cdots & \langle\mathbf{v}_1,\mathbf{v}_r\rangle\\ \vdots & & \vdots\\ \langle\mathbf{v}_r,\mathbf{v}_1\rangle & \cdots & \langle\mathbf{v}_r,\mathbf{v}_r\rangle\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_r\end{pmatrix}=\begin{pmatrix}\langle\mathbf{v}_1,\mathbf{a}\rangle\\ \vdots\\ \langle\mathbf{v}_r,\mathbf{a}\rangle\end{pmatrix}.$$

记 $C=[\mathbf{v}_1,\ldots,\mathbf{v}_r]$, 则上述方程组可写成

$$C^TC\begin{pmatrix}x_1\\ \vdots\\ x_r\end{pmatrix}=C^T\mathbf{a}.$$

由推论6.3.4可知 C^TC 可逆, 从而解得

$$P_V(\mathbf{a})=C\begin{bmatrix}x_1\\ \vdots\\ x_r\end{bmatrix}=C(C^TC)^{-1}C^T\mathbf{a}.$$

由此可得正交投影映射 P_V 的表示矩阵为

$$P_V=C(C^TC)^{-1}C^T.$$

例 6.3.6 (最小二乘解)**.** 设 A 是 $m\times n$ 矩阵, 它的列空间 $\mathcal{R}(A)$ 是 $\mathbf{R}^m$ 的子空间. 考虑向 $\mathcal{R}(A)$ 的正交投影, 称其表示矩阵为关于 A 的正交投影矩阵, 记为 P_A. 这是求解线性最小二乘问题的关键.

在 §3.6 中我们将最佳线性拟合问题叙述成求解如下最小值问题

$$\min_{\mathbf{x}\in\mathbf{R}^n} ||A\mathbf{x}-\mathbf{b}||,$$

称之为最小二乘问题, 并将其最小值点 $\mathbf{x}_0$ 称为方程组 $A\mathbf{x}=\mathbf{b}$ 的最小二乘解. 注意到, 在实际问题中, $\mathbf{b}$ 一般不属于 $\mathcal{R}(A)$, 因此方程 $A\mathbf{x}=\mathbf{b}$ 是无解的. 人们称 $\mathbf{x}_0$ 称为方程组 $A\mathbf{x}=\mathbf{b}$ 的最小二乘解, 并不是指 $\mathbf{x}_0$ 是该方程的实际解, 而只是表明 $\mathbf{x}_0$ 是函数 $||A\mathbf{x}-\mathbf{b}||$ 的最小值点. 利用正交投影的概念, 可知最小二乘解 $\mathbf{x}_0$ 满足 $A\mathbf{x}_0=P_A\mathbf{b}$.

(1) 如果 A 是列满秩的, 则 A 的列向量构成 $\mathcal{R}(A)$ 的基, 利用例6.3.5的结论, 有

$$P_A = A(A^TA)^{-1}A^T.$$

由此可得最小二乘解满足

$$A\mathbf{x}_0 = A(A^TA)^{-1}A^T\mathbf{b}.$$

在此等式两边左乘 A^T 后可得 $\mathbf{x}_0=(A^TA)^{-1}A^T\mathbf{b}$, 即当 A 列满秩时有唯一的最小二乘解.

(2) 如果 A 不是列满秩的, 可以找 $\mathcal{R}(A)$ 的一组基, 将它们从左到右排列成矩阵 B, 则 $\mathcal{R}(B)=\mathcal{R}(A)$. 由此可得

$$P_A = P_B = B(B^TB)^{-1}B^T.$$

特别的, 可用 Gram-Schmidt 正交化为 $\mathcal{R}(A)$ 选择一组单位正交基 $\mathbf{q}_1,\dots,\mathbf{q}_r$. 记 $Q=[\mathbf{q}_1,\dots,\mathbf{q}_r]$, 则给出了 A 的 QR 分解 $A=QR$, 其中

$$R = (\langle \mathbf{q}_i, \mathbf{a}_j\rangle)_{1\le i\le r, 1\le j\le n}.$$

由前述有

$$P_A = P_Q = Q(Q^TQ)^{-1}Q^T = QQ^T,$$

由此可得最小二乘解满足 $A\mathbf{x}_0=P_A\mathbf{b}=QQ^T\mathbf{b}$, 即有 $QR\mathbf{x_0}=QQ^T\mathbf{b}$. 在此式两边左乘 Q^T 就得到 $R\mathbf{x}_0=Q^T\mathbf{b}$.

回到 Moore-Penrose 广义逆的定义, 可知 A^+A 是从 $\mathbf{R}^n$ 向 $\mathcal{N}(A)^\perp=\mathcal{R}(A^T)$ 的正交投影, AA^+ 是从 $\mathbf{R}^m$ 向 $\mathcal{R}(A)$ 的正交投影. 由此可用广义逆求解最小二乘问题.

定理 6.3.7. *$\mathbf{x}_0=A^+\mathbf{b}$ 是方程组 $A\mathbf{x}=\mathbf{b}$ 的一个最小二乘解, 且是模长最小的最小二乘解.*

证明: 由例6.3.6, $\mathbf{x}$ 是最小二乘解当且仅当 $A\mathbf{x}=P_A\mathbf{b}$. 由广义逆的定义, AA^+ 是从 $\mathbf{R}^m$ 向 $\mathcal{R}(A)$ 的正交投影, 即有 $P_A=AA^+$. 结合这两点, 可知对于 $\mathbf{x}_0=A^+\mathbf{b}$, 有

$$A\mathbf{x}_0 = A(A^+\mathbf{b}) = (AA^+)\mathbf{b} = P_A\mathbf{b},$$

表明 $\mathbf{x}_0 = A^+\mathbf{b}$ 是最小二乘解.

另一方面, 对任何别的最小二乘解 $\mathbf{x}$, 有 $A\mathbf{x} = A\mathbf{x}_0$, 可知 $\mathbf{x} - \mathbf{x}_0 \in \mathcal{N}(A)$. 注意到,

$$\mathbf{x}_0 = A^+\mathbf{b} = V_r\Sigma_r^{-1}U_r^T\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = \mathcal{N}(A)^\perp.$$

特别的有 $\mathbf{x}_0 \perp (\mathbf{x} - \mathbf{x}_0)$, 由此可得

$$||\mathbf{x}||^2 = ||\mathbf{x}_0||^2 + ||\mathbf{x} - \mathbf{x}_0||^2 \geq ||\mathbf{x}_0||^2,$$

这就证明了 $\mathbf{x}_0 = A^+\mathbf{b}$ 是模长最小的最小二乘解. □

6.3.1 谱范数与低秩近似

设 A 是 $m \times n$ 实矩阵, 令

$$||A|| = \max_{\mathbf{x} \neq \mathbf{0} \in \mathbf{R}^n} \frac{||A\mathbf{x}||}{||\mathbf{x}||},$$

称为 A 的谱范数. 设 A 的奇异值分解为 $A = U\Sigma V^T$, 非零奇异值从大到小排列为 $\sigma_1 \geq \ldots \geq \sigma_r$, 则对非零向量 $\mathbf{x} = x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n$, 有 $A\mathbf{x} = x_1\sigma_1\mathbf{u}_1 + \ldots + x_r\sigma_r\mathbf{u}_r$, 由此可得

$$\frac{||A\mathbf{x}||}{||\mathbf{x}||} = \frac{\sqrt{x_1^2\sigma_1^2 + \cdots + x_r^2\sigma_r^2}}{\sqrt{x_1^2 + \cdots + x_n^2}} \leq \sigma_1.$$

结合 $\frac{||A\mathbf{v}_1||}{||\mathbf{v}_1||} = \sigma_1$, 可得 $||A|| = \sigma_1$, 即矩阵的谱范数等于最大的奇异值.

定理 6.3.8. *设 $m \times n$ 矩阵 A 的简化奇异值分解为*

$$A = \sum_{i=1}^{r} \sigma_i \cdot \mathbf{u}_i\mathbf{v}_i^T,$$

其中 $\sigma_1 \geq \ldots \geq \sigma_r$ 是 A 的所有非零奇异值. 对任何正整数 $k \leq r$, 令

$$A_k = \sum_{i=1}^{k} \sigma_i \cdot \mathbf{u}_i\mathbf{v}_i^T,$$

则对任何秩不超过 k 的 $m \times n$ 矩阵 B, 有 $||A - A_k|| \leq ||A - B||$, 即 A_k 是 A 的最佳秩 k 逼近.

证明: 不妨设 $k < r$, 令 $V_{k+1} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$. 由 $\operatorname{rank}(B) \leq k$ 可知 $\dim(\mathcal{N}(B)) = n - \operatorname{rank}(B) \geq n - k$, 由此可得

$$\begin{aligned}\dim(\mathcal{N}(B) \cap V_{k+1}) &= \dim(\mathcal{N}(B)) + \dim(V_{k+1}) - \dim\left((\mathcal{N}(B) + V_{k+1})\right) \\ &\geq (n-k) + (k+1) - n = 1,\end{aligned}$$

从而存在非零向量 $\mathbf{x}\in\mathcal{N}(B)\cap V_{k+1}$. 设 $\mathbf{x}=x_1\mathbf{v}_1+\ldots+x_{k+1}\mathbf{v}_{k+1}$, 则

$$\frac{||(A-B)\mathbf{x}||}{||\mathbf{x}||}=\frac{||A\mathbf{x}||}{||\mathbf{x}||}=\frac{\sqrt{x_1^2\sigma_1^2+\cdots+x_{k+1}^2\sigma_{k+1}^2}}{\sqrt{x_1^2+\cdots+x_{k+1}^2}}\geq\sigma_{k+1},$$

由此可得 $||A-B||\geq\sigma_{k+1}=||A-A_k||$. □

6.3.2 主成分分析

定理 6.3.9 (Ky Fan 樊畿). 设 n 阶实对称矩阵 A 的特征值为 $\lambda_1\geq\cdots\geq\lambda_n$, 对应的特征向量为 $\mathbf{u}_1,\cdots,\mathbf{u}_n$, 则有

$$\max_{n\times m\text{矩阵满足}Q^TQ=I} trace(Q^TAQ)=\sum_{i=1}^m\lambda_i,$$

且 $Q=[\mathbf{u}_1,\cdots,\mathbf{u}_m]$ 时取得最大值.

证明: 实对称矩阵 A 可正交相似于对角矩阵, 即存在正交矩阵 P 使得 $A=P^T\Lambda P$, 其中 $\Lambda=\mathrm{diag}\{\lambda_1,\cdots,\lambda_n\}$. 这样, 可对原来的最值问题换元处理: 令 $\widetilde{Q}=PQ$, 则 $Q,\widetilde{Q}$ 相互确定 (因为 $P^{-1}\widetilde{Q}=Q$), 且有

$$\widetilde{Q}^T\widetilde{Q}=I_m\iff Q^TP^TPQ=I_m\iff Q^TQ=I_m.$$

由此可得

$$\max_{n\times m\text{矩阵满足}Q^TQ=I}\mathrm{trace}(Q^TAQ)=\max_{n\times m\text{矩阵满足}\widetilde{Q}^T\widetilde{Q}=I}\mathrm{trace}(\widetilde{Q}^T\Lambda\widetilde{Q}),$$

转化为对 Λ 证明 Ky Fan 不等式.

以下假设 $A=\Lambda$. 将 Q 写成列向量分块形式 $Q=[\mathbf{q}_1,\cdots,\mathbf{q}_m]$, 令

$$f(Q)=\mathrm{trace}(Q^TAQ)=\mathbf{q}_1^TA\mathbf{q}_1+\cdots+\mathbf{q}_m^TA\mathbf{q}_m,$$

视之为 m^2 个变元 $q_{ij}(1\leq i,j\leq m)$ 的函数, 来求 f 在如下 $C_m^2+m=\frac{m(m+1)}{2}$ 个约束条件

$$\langle\mathbf{q}_k,\mathbf{q}_l\rangle-\delta_{kl}=0,\quad 1\leq k\leq l\leq m$$

下的最大值. 由最值定理, f 有最大值点 V, 这里 V 是 $n\times m$ 矩阵且满足 $V^TV=I_m$. 注意到, V 是 f 在前述 $\frac{m(m+1)}{2}$ 个约束条件下的条件极值点, 利用 Lagrange 乘子法, 存在实数 $\lambda_{kl}(1\leq k\leq l\leq m)$ 使得 V 的 nm 个矩阵元与这些 λ_{kl} 满足 Lagrange 辅助函数

$$F=f-\sum_{1\leq k\leq l\leq m}\lambda_{kl}\left(\langle\mathbf{q}_k,\mathbf{q}_l\rangle-\delta_{kl}\right)$$

的临界点方程:

$$\frac{\partial F}{\partial q_{ij}}=0,\quad \forall 1\le i\le n,\forall 1\le j\le m.$$

将上述方程化简, 即为

$$\lambda_i q_{ij}=\lambda_{jj}q_{ij}+\sum_{k<j}\lambda_{kj}q_{ik}+\sum_{k>j}\lambda_{jk}q_{ik}. \tag{6.2}$$

将 $\{\lambda_{kl}\}_{k\le l}$ 扩充为一个 m 阶对称矩阵 K, 即对任何 $k\le l$ 定义 $K_{kl}=K_{lk}=\lambda_{kl}$. 由此可将(9.2)改写成矩阵形式:

$$AQ=QK.$$

这样, Lagrange 乘子法断言 f 在约束条件下的最大值点 V 满足: 存在 m 阶对称矩阵 K, 使得 $AV=VK$, 即有 $V^TAV=K$. 由此可得

$$\max_{Q^TQ=I_m} f(Q)=f(V)=\mathrm{trace}(V^TAV)=\mathrm{trace}(K).$$

将 V 的列向量组 $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ 扩充为 $\mathbf{R}^n$ 的单位正交基 $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$. 记 $\widetilde{V}=[\mathbf{v}_1,\cdots,\mathbf{v}_n]$, 令 $B=\widetilde{V}^TA\widetilde{V}$, 则 B 是对称矩阵. 注意到 B 的矩阵元为

$$B_{ij}=\langle \mathbf{v}_i,A\mathbf{v}_j\rangle,$$

再有 $AV=VK$ 可知对 $1\le i\le m$ 有 $A\mathbf{v}_i\in\mathrm{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$, 结合这两点可得 B 是如下形式的分块矩阵

$$B=\begin{pmatrix} K & O\\ O & *\end{pmatrix}.$$

这样, K 的特征值都是 B 的特征值, 而 B 正交相似于 A, 则 K 的特征值是 $\lambda_1,\cdots,\lambda_n$ 的一部分, 所以有

$$\mathrm{trace}(K)=\sum(K\text{的特征值})\le\lambda_1+\cdots+\lambda_m,$$

从而完成了整个证明.

□

问题 6.3.10. 给定 $\mathbf{R}^m$ 中 n 个点 $\mathbf{a}_1,\ldots,\mathbf{a}_n$. 求 $\mathbf{R}^n$ 的 k 维子空间 $\mathcal{M}$, 使得 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 到 $\mathcal{M}$ 的距离的平方和

$$E=\sum_{i=1}^n||\mathbf{a}_i-P_{\mathcal{M}}(\mathbf{a}_i)||^2$$

取得最小值, 其中 $P_{\mathcal{M}}$ 表示向 $\mathcal{M}$ 的正交投影.

解. 设 $\mathcal{M}$ 的一组单位正交基为 $\mathbf{q}_1, \ldots, \mathbf{q}_k$, 记 $Q = [\mathbf{q}_1, \cdots, \mathbf{q}_k]$, 则 $P_{\mathcal{M}} = QQ^T$, 由此可得

$$\begin{aligned} E &= \sum_{i=1}^{n} \langle (\mathbf{a}_i - QQ^T\mathbf{a}_i, \mathbf{a}_i - QQ^T\mathbf{a}_i \rangle \\ &= \sum_{i=1}^{n} \left(\langle \mathbf{a}_i, \mathbf{a}_i \rangle - 2\langle \mathbf{a}_i, QQ^T\mathbf{a}_i \rangle + \langle QQ^T\mathbf{a}_i, QQ^T\mathbf{a}_i \rangle \right) \\ &= \sum_{i=1}^{n} \left(\langle \mathbf{a}_i, \mathbf{a}_i \rangle - 2\langle Q^T\mathbf{a}_i, Q^T\mathbf{a}_i \rangle + \langle Q^T\mathbf{a}_i, Q^TQQ^T\mathbf{a}_i \rangle \right) \\ &= \sum_{i=1}^{n} \left(\langle \mathbf{a}_i, \mathbf{a}_i \rangle - \langle Q^T\mathbf{a}_i, Q^T\mathbf{a}_i \rangle \right). \end{aligned}$$

注意到, 对矩阵 $B = [\mathbf{b}_1, \cdots, \mathbf{b}_p]$, 有如下写法:

$$\sum_{i=1}^{p} \langle \mathbf{b}_i, \mathbf{b}_i \rangle = \sum_{i=1}^{p} \mathbf{b}_i^T\mathbf{b}_i = \text{trace}(B^TB) = \text{trace}(BB^T).$$

记 $A = [\mathbf{a}_1, \cdots, \mathbf{a}_k]$, 利用上述用迹的记法, 可把 E 表示成

$$E = \text{trace}(A^TA) - \text{trace}\left((Q^TA)(Q^TA)^T\right) = \text{trace}(A^TA) - \text{trace}(Q^TAA^TQ).$$

这样, 求解 E 的最小值等于求解如下最大值问题:

$$\max_{Q\text{是}m\times k\text{矩阵, 满足}Q^TQ=I_k} \text{trace}(Q^TAA^TQ),$$

此问题是定理6.3.9特例. 设 A 的简化奇异值分解为

$$A = \sum_{i=1}^{r} \sigma_i \cdot \mathbf{u}_i\mathbf{v}_i^T,$$

其中 $\sigma_1 \geq \ldots \geq \sigma_r$ 是 A 的所有非零奇异值. 对任何正整数 $k \leq r$, 利用定理6.3.9的结论可得

$$\max_{Q\text{是}m\times k\text{矩阵, 满足}Q^TQ=I_k} \text{trace}(Q^TAA^TQ) = \sigma_1^2 + \cdots + \sigma_k^2.$$

进一步, 可得原问题中 E 的最小值为

$$\min E = \text{trace}(A^TA) - \left(\sigma_1^2 + \cdots + \sigma_k^2\right) = \sum_{i=k+1}^{r} \sigma_i^2.$$

□

设 A 是 $m\times n$ 实矩阵, 令

$$||A||_F=\sqrt{\mathrm{trace}(A^TA)}=\sqrt{\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2},$$

称为 A 的 Frobenius 范数. 设 A 的奇异值分解为 $A=U\Sigma V^T$, 则

$$||A||_F^2=\mathrm{trace}(V\Sigma^TU^TU\Sigma V^T)=\mathrm{trace}(V(\Sigma^T\Sigma)V^{-1})=\sum_{i=1}^r\sigma_i^2,$$

即矩阵的 Frobenius 范数的平方等于所有奇异值的平方和.

定理 6.3.11. 设 $m\times n$ 矩阵 A 的简化奇异值分解为

$$A=\sum_{i=1}^r\sigma_i\cdot\mathbf{u}_i\mathbf{v}_i^T,$$

其中 $\sigma_1\geq\ldots\geq\sigma_r$ 是 A 的所有非零奇异值. 对任何正整数 $k\leq r$, 令

$$A_k=\sum_{i=1}^k\sigma_i\cdot\mathbf{u}_i\mathbf{v}_i^T,$$

则对任何秩不超过 k 的 $m\times n$ 矩阵 B, 有 $||A-A_k||_F\leq||A-B||_F$, 即 A_k 是 A 在 *Frobenius* 范数意义下的最佳秩 k 逼近.

证明*:* 设 $B=[\mathbf{b}_1,\cdots,\mathbf{b}_n]$, 它的列向量空间为 $\mathcal{R}(B)=\mathcal{M}$, 其维数等于 p, 则 $p\leq k$. 不妨设 $p\geq 1$, 由 Frobenius 范数的定义可得

$$||A-B||_F^2=\sum_{i=1}^n||\mathbf{a}_i-\mathbf{b}_i||^2\geq\sum_{i=1}^n||\mathbf{a}_i-P_{\mathcal{M}}\mathbf{a}_i||^2.$$

上式左边恰为问题6.3.10中的 E, 由问题6.3.10 的结论有

$$\min E=\sum_{i=p+1}^r\sigma_i^2\geq\sum_{i=k+1}^r\sigma_i^2=||A-A_k||_F^2.$$

结合这两步, 就完成了定理的证明. □

第七章　线性空间与线性映射

本章考虑的数域 $\mathbf{F}$ 为 $\mathbf{R}$ 或 $\mathbf{C}$.

7.1　线性空间

定义 7.1.1 (线性空间)**.** 称非空集合 V 为 $\mathbf{F}$ 上的线性空间, 如果 V 上有加法运算 $V \times V \xrightarrow{+} V$, 并有 $\mathbf{F}$ 的元素与 V 中成员的数乘运算 $\mathbf{F} \times V \xrightarrow{\cdot} V$, 要求这两种运算满足如下运算法则:

(1) 加法是结合的: $(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})$;

(2) 加法是交换的: $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$;

(3) 加法有零元, 即存在一个元素 $\mathbf{0} \in V$ 满足对任何 $\mathbf{a} \in V$ 有 $\mathbf{a}+\mathbf{0} = \mathbf{a} = \mathbf{0}+\mathbf{a}$;

(4) 加法有逆, 对每个 $\mathbf{a} \in V$, 存在 $-\mathbf{a} \in V$, 使得 $\mathbf{a}+(-\mathbf{a}) = \mathbf{0}$;

(5) 数乘作用有单位, $1\mathbf{a} = \mathbf{a}$;

(6) 数乘作用是结合的, $k(l\mathbf{a}) = (kl)\mathbf{a}$;

(7) 数乘作用保持向量加法, $k(\mathbf{a}+\mathbf{b}) = k\mathbf{a}+k\mathbf{b}$;

(8) 数乘关于数的分配律: $(k+l)\mathbf{a} = k\mathbf{a}+l\mathbf{a}$.

定义 7.1.2 (子空间)**.** 设 V 是 $\mathbf{F}$ 上的线性空间, W 是 V 的非空子集, 称 W 是 V 的线性子空间, 如果对任何 $\mathbf{x},\mathbf{y} \in W$ 及任何 $k \in \mathbf{F}$, 有 $\mathbf{x}+\mathbf{y} \in W$, 且 $k\mathbf{x} \in W$, 即要求 W 关于加法与数乘运算都是封闭的.

定义 7.1.3 (子空间的和与直和)**.** 设 W_1, W_2 是 V 的子空间, 令

$$W_1 + W_2 = \{\mathbf{w}_1 + \mathbf{w}_2 | \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2\},$$

它是 V 的子空间, 称之为 W_1 与 W_2 的和. 进一步, 如果 $W_1 \cap W_2 = \{\mathbf{0}\}$, 则称 $W_1 + W_2$ 为 W_1 与 W_2 的直和, 记作 $W_1 \oplus W_2$.

定义 7.1.4. 设 V 是 $\mathbf{F}$ 上的线性空间, $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 是 V 中的一组向量, 令

$$\mathrm{span}(S) = \mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_m\} = \{k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m | k_1, \cdots, k_m \in \mathbf{F}\},$$

易验证 span(S) 是线性子空间, 称之为由 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 生成的子空间.

定义 7.1.5. 设 V 是 $\mathbf{F}$ 上的线性空间. 称 V 中的一组向量 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 是线性相关的, 如果存在不全为零的 $k_1, \cdots, k_m \in \mathbf{F}$, 使得 $k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$; 等价的说, 即 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 中存在一个向量可以表示成其余向量的线性组合.

如果 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 不是线性相关的, 则称它们线性无关; 等价的说, 即 $k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$ 当且仅当 $k_1, \cdots, k_m$ 全为零. 约定空向量组为线性无关的.

定义 7.1.6. 设 V 是 $\mathbf{F}$ 上的线性空间, 称向量组 $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 为 V 的一个基底 (basis), 如果它们线性无关且生成 V; 等价的说, 即 V 中的每个向量都可以唯一的表示成 $\mathbf{a}_1, \cdots, \mathbf{a}_m$ 的线性组合.

约定零子空间的基底为空集.

定义 7.1.7. 设 V 是 $\mathbf{R}^n$ 的线性子空间, 定义 V 的维数为 V 的任意一组基底中向量的个数, 记为 $\dim(V)$.

定理 7.1.8. *设 V 是 $\mathbf{F}$ 上的线性空间, W_1, W_2 是 V 的有限维子空间, 则有*

$$dim(W_1 + W_2) = dim(W_1) + dim(W_2) - dim(W_1 \cap W_2).$$

7.2 向量的坐标表示

设 V 是 $\mathbf{F}$ 上的有限维线性空间, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 是它的一组基. 任何 $\mathbf{v} \in V$ 可唯一的表示成这组基的线性组合

$$\mathbf{v} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, \quad x_1, \ldots, x_n \in \mathbf{F}. \tag{7.1}$$

称向量 $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{F}^n$ 为 $\mathbf{v}$ 在基 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 下的坐标, 并把(8.1)写作

$$\mathbf{v} = [\mathbf{e}_1, \ldots, \mathbf{e}_n] \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

可从映射的角度理解基. 上述基 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 确定了线性映射 $\Phi : \mathbf{F}^n \to V$,

$$\Phi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

显然 Φ 是线性同构, 其逆 $\Phi^{-1}(\mathbf{v})$ 就是给出 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标. 这样, 给定 V 的一个基, 等价于给出 V 到 $\mathbf{F}^n$ 的一个线性同构.

设 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的另一组基, 则 $\mathbf{v}$ 在 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下有坐标表示

$$\mathbf{v}=[\mathbf{v}_1,\ldots,\mathbf{v}_n]\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix}.$$

我们来确定同一个 $\mathbf{v}$ 在不同基下的坐标表示之间的关系. 为此, 对每个 $j\le n$, 设

$$\mathbf{v}_j=T_{1j}\mathbf{e}_1+\cdots+T_{nj}\mathbf{e}_n=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} T_{1j}\\ \vdots\\ T_{nj}\end{pmatrix},$$

则可形式化的写成

$$[\mathbf{v}_1,\ldots,\mathbf{v}_n]=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} T_{11}&\cdots&T_{1n}\\ \vdots&\ddots&\vdots\\ T_{n1}&\cdots&T_{nn}\end{pmatrix}.$$

记 $T=(T_{ij})$, 称之为从基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 到基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的过渡矩阵, 它是可逆矩阵, 其逆 T^{-1} 是从基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 到基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 的过渡矩阵.

设 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标为 $\begin{pmatrix} x_1\\ \vdots\\ x_n\end{pmatrix}\in\mathbf{F}^n$, 基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下的坐标为 $\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix}\in\mathbf{F}^n$, 则有

$$[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} x_1\\ \vdots\\ x_n\end{pmatrix}=[\mathbf{v}_1,\ldots,\mathbf{v}_n]\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix}=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} T_{11}&\cdots&T_{1n}\\ \vdots&\ddots&\vdots\\ T_{n1}&\cdots&T_{nn}\end{pmatrix}\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix},$$

由此可得

$$\begin{pmatrix} x_1\\ \vdots\\ x_n\end{pmatrix}=T\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix},$$

即 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标等于 T 乘以它在基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下的坐标.

7.3 线性映射

设 V, W 是 $\mathbf{F}$ 上的有限维线性空间, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 与 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 分别是 V 与 W 的基. 线性映射 $f : V \to W$ 由它在 V 的基上的值 $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)$ 唯一确定, 设

$$f(\mathbf{v}_j) = \sum_{i=1}^{m} F_{ij}\mathbf{w}_i, \quad j = 1, \ldots, n. \tag{7.2}$$

令 $F = (F_{ij})_{1\le i\le m, 1\le j\le n}$, 称之为 f 在前述两组基下的表示矩阵. 由此可将(8.2)形式化的记为

$$[f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)] = [\mathbf{w}_1, \ldots, \mathbf{w}_m]F.$$

利用表示矩阵, 可确定 $\mathbf{v}$ 的坐标 $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 与 $f(\mathbf{v})$ 的坐标 $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ 之间的关系:

$$\begin{aligned}[\mathbf{w}_1, \ldots, \mathbf{w}_m]\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = f(\mathbf{v}) = f\left([\mathbf{v}_1, \ldots, \mathbf{v}_n]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) \\ =[f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]F\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.\end{aligned}$$

由此可得

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = F\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

即像点的坐标等于表示矩阵乘以原像点的坐标.

命题 7.3.1 (复合映射的表示矩阵等于表示矩阵的乘积)**.** *设 U, V, W 都是有限维线性空间, 为它们各选定一组基. 设 $f : U \to V$ 与 $g : V \to W$ 是线性映射, 则在前述选定的基下, 复合映射 $g \circ f$ 的表示矩阵等于 g 的表示矩阵 G 乘以 f 的表示矩阵 F, 即 $g \circ f$ 的表示矩阵为 GF.*

命题 7.3.2 (换基前后的表示矩阵相抵)**.** *设从 V 的基 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 到另一个基 $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ 的过渡矩阵为 T; 从 W 的基 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 到另一个基 $\mathbf{w}'_1, \ldots, \mathbf{w}'_m$ 的过渡矩阵为 S. 如果线性映射 $f : V \to W$ 在基 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 与 $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 下的表示矩阵为 F, 则 $f : V \to W$ 在基 $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ 与 $\mathbf{w}'_1, \ldots, \mathbf{w}'_m$ 下的表示矩阵为 $S^{-1}FT$.*

证明: 将 $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ 简记为 $[\mathbf{v}]$, 对其他基也采用同样的记法, 则有

$$[\mathbf{v}'] = [\mathbf{v}]T, \quad [\mathbf{w}'] = [\mathbf{w}]S, \quad [f(\mathbf{v})] = [\mathbf{w}]F,$$

由此可得

$$[f(\mathbf{v}')] = f([\mathbf{v}]T) = [f(\mathbf{v})]T = [\mathbf{w}]FT = [\mathbf{w}']S^{-1}FT,$$

这表明 f 在基 $[\mathbf{v}']$ 与 $[\mathbf{w}']$ 下的表示矩阵为 $S^{-1}FT$. □

对于线性空间 V 到自身的线性映射 $f: V \to V$(称为 V 上的线性变换), 我们为定义域与值域的 V 选取同样的基 $[\mathbf{v}]$, 由此可谈论 f 的表示矩阵 F. 设 $[\mathbf{v}']$ 是另一组基, 且从 $[\mathbf{v}]$ 到 $[\mathbf{v}']$ 的过渡矩阵为 S, 则由命题8.11.2 可知, f 在基 $[\mathbf{v}']$ 下的表示矩阵为 $S^{-1}FS$. 也就是说, 线性变换在换基前后的表示矩阵相似.

例 7.3.3. 已知在 $\mathbf{R}^3$ 中, 线性变换 T 在基

$$\mathbf{v}_1 = \begin{bmatrix} 8 \\ -6 \\ 7 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 7 \\ -13 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 9 \\ -3 \\ 7 \end{bmatrix}$$

下的表示矩阵为

$$A = \begin{bmatrix} 1 & -18 & 15 \\ -1 & -22 & 20 \\ 1 & -25 & 22 \end{bmatrix}.$$

求 T 在基

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

下的表示矩阵.

解. 记 $\mathbf{R}^3$ 的标准基为 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, 则从基 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 到基 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 的过渡矩阵为

$$[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]V,$$

其中 V 是将三个列向量 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 从左至右排列成的 3×3 矩阵. 类似的,

$$[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]W.$$

结合这两者, 可知从基 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 到基 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ 的过渡矩阵为

$$[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]W = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]V^{-1}W.$$

已知 T 在 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 下的表示矩阵为 A, 则 T 在 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ 下的表示矩阵为

$$(V^{-1}W)^{-1}A(V^{-1}W) = W^{-1}VAV^{-1}W.$$

$\square$

第八章　附录一

8.1　Gram-Schmidt 正交化与 QR 分解

定理 8.1.1 (Gram-Schmidt 正交化)**.** *设* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *是内积空间* V *的一个线性无关向量组, 则存在单位正交向量组* $\{\mathbf{q}_1, \cdots, \mathbf{q}_r\}$ *使得* $span(\{\mathbf{q}_1, \cdots, \mathbf{q}_r\}) = span(\{\mathbf{v}_1, \cdots, \mathbf{v}_r\})$*.*

证明:

- 取 $\mathbf{w}_1 = \mathbf{v}_1$, 令 $\mathbf{q}_1 = \frac{\mathbf{w}_1}{||\mathbf{w}_1||}$;
- 取 $\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{q}_1\langle\mathbf{q}_1, \mathbf{v}_2\rangle$, 令 $\mathbf{q}_2 = \frac{\mathbf{w}_2}{||\mathbf{w}_2||}$;
- 取 $\mathbf{w}_3 = \mathbf{v}_3 - \mathbf{q}_1\langle\mathbf{q}_1, \mathbf{v}_3\rangle - \mathbf{q}_2\langle\mathbf{q}_2, \mathbf{v}_3\rangle$, 令 $\mathbf{q}_3 = \frac{\mathbf{w}_3}{||\mathbf{w}_3||}$;
- ......;
- 取 $\mathbf{w}_r = \mathbf{v}_r - \mathbf{q}_1\langle\mathbf{q}_1, \mathbf{v}_r\rangle - \cdots - \mathbf{q}_{r-1}\langle\mathbf{q}_{r-1}, \mathbf{v}_r\rangle$, 令 $\mathbf{q}_r = \frac{\mathbf{w}_r}{||\mathbf{w}_r||}$.

□

定义 8.1.2. 称方阵 Q 为正交矩阵, 如果它满足如下四个彼此等价的断言:

(1) $Q^TQ = I_n$;

(2) Q 的列向量构成 $\mathbf{R}^n$ 的单位正交基;

(3) $QQ^T = I_n$;

(4) Q 的行向量构成 $\mathbf{R}^n$ 的单位正交基.

将 Gram-Schmidt 正交化用于矩阵 A 的列向量组, 得到矩阵 A 的 QR 分解.

定理 8.1.3 (QR 分解)**.** *设* A *是* n *阶可逆矩阵, 则存在一个正交矩阵* Q *与一个对角元为正数的上三角矩阵* R*, 使得* $A = QR$*. 进一步, 满足上述条件的* Q, R *都是唯一的.*

证明: 将 A 表示成列向量分块矩阵 $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$, 对 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 做 Gram-Schmidt 正交化,

得到单位正交基 $\mathbf{q}_1, \cdots, \mathbf{q}_n$. 令 $Q = [\mathbf{q}_1, \cdots, \mathbf{q}_n]$,

$$R = \begin{pmatrix} \langle \mathbf{q}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{q}_n, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{q}_n, \mathbf{a}_n \rangle \end{pmatrix}.$$

易知 Q 是正交矩阵, R 是上三角矩阵, 且有 $A = QR$, 得到所需要的分解. □

例 8.1.4 (QR 分解可用于计算行列式)**.** 设 n 阶方阵 A 的 QR 分解为 $A = QR$, 则

$$\det(A) = \det(QR) = \det(Q)\det(R) = \pm\det(R) = \pm\prod_{i=1}^{n} R_{ii} = \pm\prod_{i=1}^{n}\langle \mathbf{q}_i, \mathbf{a}_i \rangle.$$

结合 Cauchy-Schwartz 不等式, 可得 Hadamard 不等式

$$|\det(A)| \leq \prod_{i=1}^{n} ||\mathbf{a}_i||,$$

其几何意义为: 平行体的 Volume 不超过它所有棱长的乘积.

定理 8.1.5 (一般 QR 分解)**.** *设 A 是 $m \times n$ 矩阵, 秩为 r, 则存在 $m \times r$ 矩阵 Q 与 $r \times n$ 矩阵 R, 满足如下条件:*

(1) $A = QR$;

(2) $Q^T Q = I_r$, 即 Q 的列向量是单位正交向量组;

(3) R 是上三角矩阵且对角元非负.

证明: 仿照 Gram-Schmidt 正交化, 先构造 $\mathbf{R}^m$ 的单位正交向量组 $Q = \{\mathbf{q}_1, \cdots, \mathbf{q}_r\}$. 初始时, 令 $Q_0 = \emptyset$. 依次对 A 对列向量 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 做如下考察: 如果 $\mathbf{a}_k \in \mathrm{span}(Q_{k-1})$, 则令 $Q_k = Q_{k-1}$, 转而考虑下一个 $\mathbf{a}_{k+1}$; 如果 $\mathbf{a}_k \notin \mathrm{span}(Q_{k-1})$, 则令 $\mathbf{u}$ 为 $\mathbf{a}_k$ 在 $\mathrm{span}(Q_{k-1})$ 垂直方向上的分量, 即

$$\mathbf{u} = \mathbf{a}_k - \sum_{\mathbf{q} \in Q_{k-1}} \mathbf{q}\langle \mathbf{q}, \mathbf{a}_k \rangle.$$

令 $Q_k = Q_{k-1} \cup \{\frac{\mathbf{u}}{||\mathbf{u}||}\}$, 并转到考虑下一个 $\mathbf{a}_{k+1}$. 当所有 n 步考察结束时, 把所得的 Q_n 记为 Q, 则 Q 是由 r 个单位正交向量构成, 且对每个 k 都有 $|Q_k| \leq k$. 按照加入 Q 的先后顺序, 将 Q 的成员依次编号为 $\mathbf{q}_1, \cdots, \mathbf{q}_r$, 则有

$$A = [\mathbf{q}_1, \cdots, \mathbf{q}_r] \begin{pmatrix} \langle \mathbf{q}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{q}_r, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{q}_r, \mathbf{a}_n \rangle \end{pmatrix},$$

这就给出了 A 的 QR 分解. □

例 8.1.6. 求如下矩阵 A 的 QR 分解.

$$A=\begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}.$$

解. 对 A 的列向量做 Gram-Schmidt 正交化, 可得

$$Q=\begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix},$$

$$R=\begin{pmatrix} \langle \mathbf{q}_1,\mathbf{a}_1\rangle & \langle \mathbf{q}_1,\mathbf{a}_2\rangle & \langle \mathbf{q}_1,\mathbf{a}_3\rangle \\ \langle \mathbf{q}_2,\mathbf{a}_1\rangle & \langle \mathbf{q}_2,\mathbf{a}_2\rangle & \langle \mathbf{q}_2,\mathbf{a}_3\rangle \\ \langle \mathbf{q}_3,\mathbf{a}_1\rangle & \langle \mathbf{q}_3,\mathbf{a}_2\rangle & \langle \mathbf{q}_3,\mathbf{a}_3\rangle \end{pmatrix}=\begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}.$$

□

8.2 正交投影与最小二乘法

设 V 是 $\mathbf{R}^n$ 的子空间, 考虑 $\mathbf{R}^n$ 向 V 的正交投影 $P_V:\mathbf{R}^n\to\mathbf{R}^n$. 如果为 V 选定了一组单位正交基 $\{\mathbf{q}_1,\ldots,\mathbf{q}_r\}$, 则对任何 $\mathbf{x}\in\mathbf{R}^n$ 有

$$P_V(\mathbf{x})=\sum_{i=1}^r \mathbf{q}_i\langle \mathbf{q}_i,\mathbf{x}\rangle=\sum_{i=1}^r \mathbf{q}_i(\mathbf{q}_i^T\mathbf{x})=\left(\sum_{i=1}^r \mathbf{q}_i\mathbf{q}_i^T\right)\cdot\mathbf{x},$$

由此可得正交投影映射 P_V 的表示矩阵为 $\sum\limits_{i=1}^r \mathbf{q}_i\mathbf{q}_i^T$. 记 $Q=[\mathbf{q}_1,\ldots,\mathbf{q}_r]$, 则 P_V 的表示矩阵可以表示成 QQ^T.

若只是为 V 找到一组 (不一定单位正交的) 基 $\mathbf{v}_1,\ldots,\mathbf{v}_r$, 则前述计算就会变得复杂. 比如, 为了确定 $\mathbf{a}$ 向 V 的正交投影, 设 $\mathbf{a}_V=\sum\limits_{i=1}^r x_i\mathbf{v}_i$, 则由 $\mathbf{a}-\mathbf{a}_V\perp\mathbf{v}_k$ 可得

$$\sum_{i=1}^r \langle \mathbf{v}_k,\mathbf{v}_i\rangle x_i=\langle \mathbf{v}_k,\mathbf{a}\rangle.$$

这样就等价于求解线性方程组

$$\begin{pmatrix} \langle \mathbf{v}_1,\mathbf{v}_1\rangle & \cdots & \langle \mathbf{v}_1,\mathbf{v}_r\rangle \\ \vdots & & \vdots \\ \langle \mathbf{v}_r,\mathbf{v}_1\rangle & \cdots & \langle \mathbf{v}_r,\mathbf{v}_r\rangle \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}=\begin{pmatrix} \langle \mathbf{v}_1,\mathbf{a}\rangle \\ \vdots \\ \langle \mathbf{v}_r,\mathbf{a}\rangle \end{pmatrix}.$$

记 $C=[\mathbf{v}_1,\ldots,\mathbf{v}_r]$, 则上述方程组可写成

$$C^TC\begin{pmatrix}x_1\\ \vdots\\ x_r\end{pmatrix}=C^T\mathbf{a}.$$

由于 C 是列满秩的, 可知 C^TC 可逆, 从而解得

$$P_V(\mathbf{a})=C\begin{bmatrix}x_1\\ \vdots\\ x_r\end{bmatrix}=C(C^TC)^{-1}C^T\mathbf{a}.$$

由此可得正交投影映射 P_V 的表示矩阵为

$$P_V=C(C^TC)^{-1}C^T.$$

设 A 是 $m\times n$ 矩阵, 它的列空间 $\mathcal{R}(A)$ 是 $\mathbf{R}^m$ 的子空间. 考虑向 $\mathcal{R}(A)$ 的正交投影, 称其表示矩阵为关于 A 的正交投影矩阵, 记为 P_A. 这是求解线性最小二乘问题的关键.

我们将最佳线性拟合问题叙述成求解如下最小值问题

$$\min_{\mathbf{x}\in\mathbf{R}^n}||A\mathbf{x}-\mathbf{b}||,$$

称之为最小二乘问题, 并将其最小值点 $\mathbf{x}_0$ 称为方程组 $A\mathbf{x}=\mathbf{b}$ 的最小二乘解. 注意到, 在实际问题中, $\mathbf{b}$ 一般不属于 $\mathcal{R}(A)$, 因此方程 $A\mathbf{x}=\mathbf{b}$ 是无解的. 人们称 $\mathbf{x}_0$ 称为方程组 $A\mathbf{x}=\mathbf{b}$ 的最小二乘解, 并不是指 $\mathbf{x}_0$ 是该方程的实际解, 而只是表明 $\mathbf{x}_0$ 是函数 $||A\mathbf{x}-\mathbf{b}||$ 的最小值点. 利用正交投影的概念, 可知最小二乘解 $\mathbf{x}_0$ 满足 $A\mathbf{x}_0=P_A\mathbf{b}$.

(1) 如果 A 是列满秩的, 则 A 的列向量构成 $\mathcal{R}(A)$ 的基, 由前述正交投影映射的表示矩阵公式有

$$P_A=A(A^TA)^{-1}A^T.$$

由此可得最小二乘解满足

$$A\mathbf{x}_0=A(A^TA)^{-1}A^T\mathbf{b}.$$

在此等式两边左乘 A^T 后可得 $\mathbf{x}_0=(A^TA)^{-1}A^T\mathbf{b}$, 即当 A 列满秩时有唯一的最小二乘解.

(2) 如果 A 不是列满秩的, 可以找 $\mathcal{R}(A)$ 的一组基, 将它们从左到右排列成矩阵 B, 则 $\mathcal{R}(B)=\mathcal{R}(A)$. 由此可得

$$P_A=P_B=B(B^TB)^{-1}B^T.$$

特别的, 可用 Gram-Schmidt 正交化为 $\mathcal{R}(A)$ 选择一组单位正交基 $\mathbf{q}_1,\ldots,\mathbf{q}_r$. 记 $Q=[\mathbf{q}_1,\ldots,\mathbf{q}_r]$, 则给出了 A 的 QR 分解 $A=QR$, 其中

$$R=(\langle\mathbf{q}_i,\mathbf{a}_j\rangle)_{1\le i\le r,1\le j\le n}.$$

由前述有

$$P_A = P_Q = Q(Q^TQ)^{-1}Q^T = QQ^T,$$

由此可得最小二乘解满足 $A\mathbf{x}_0 = P_A\mathbf{b} = QQ^T\mathbf{b}$, 即有 $QR\mathbf{x_0} = QQ^T\mathbf{b}$. 在此式两边左乘 Q^T 就得到 $R\mathbf{x}_0 = Q^T\mathbf{b}$.

8.3 行列式

定义 8.3.1. 设 $A = (a_{ij})_{1\le i,j\le n}$ 是 n 阶方阵, 定义 A 的行列式为

$$\begin{aligned}\det A &= \sum_{\sigma\in S_n}\left(\operatorname{sign}(\sigma)a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}\right)\\ &= \sum_{\sigma\in S_n}\left(\operatorname{sign}(\sigma)a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n}\right),\end{aligned}$$

其中 $\operatorname{sign}(\sigma)$ 表示置换 σ 的符号, $\operatorname{sign}(\sigma) = 1$ 当且仅当可通过偶数次对换从 σ 变成恒同置换.

命题 8.3.2 (行列式的性质)**.** (1) *行列式关于每行 (列) 都是线性的; 交换矩阵中某两行 (列) 后行列式变成相反数; 如果矩阵中有两个行 (列) 向量相等则行列式等于零;*

(2) *初等行 (列) 变换对行列式的改变规则是: 对换两行 (列) 之后行列式变成相反数; 倍加变换不改变行列式; 将某行 (列) 倍乘 k 之后行列式也乘以 k.*

(3) $\det(A^T) = \det A$;

(4) *A 是可逆矩阵的充分必要条件是* $\det A \ne 0$;

(5) $\det(AB) = \det(A)\cdot\det(B)$;

(6) *如果 A 可逆, 则有* $\det(A^{-1}) = \frac{1}{\det(A)}$.

定义 8.3.3. 设 $A = (a_{ij})_{1\le i,j\le n}$ 是 n 阶方阵, 令 $M_{i,j}$ 为将 A 的第 i 行与第 j 列删去所得 $(n-1)\times(n-1)$ 矩阵的行列式, 称之为矩阵元 a_{ij} 的余子式; 称 $(-1)^{i+j}M_{ij}$ 为 a_{ij} 的代数余子式.

定理 8.3.4 (Laplace 展开)**.** *设 A 是 n 阶矩阵, 则有*

$$\begin{aligned}\det(A) &= \sum_{j=1}^{n}(-1)^{i+j}a_{ij}M_{i,j}\\ &= \sum_{i=1}^{n}(-1)^{i+j}a_{ij}M_{i,j}.\end{aligned}$$

定义 8.3.5. 设 A 是 n 阶方阵, 定义它的代数余子式矩阵 C 为

$$C_{ij} = (-1)^{i+j} M_{ij},$$

即 C_{ij} 为 a_{ij} 的代数余子式.

称 C^T 为 A 的伴随矩阵, 记作 $\mathrm{adj}(A) = C^T$.

命题 8.3.6. 对任何 n 阶方阵 A, 有

$$adj(A) \cdot A = A \cdot adj(A) = \det(A) I_n.$$

特别的, 如果 A 可逆, 则 A 的逆矩阵为

$$A^{-1} = \frac{1}{\det(A)} adj(A).$$

推论 8.3.7 (Cramer 法则)**.** 设 A 是可逆的 n 阶方阵, 则方程组

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

有唯一解

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad i = 1, \ldots, n,$$

其中 A_i 是由将 A 的第 i 个列向量换成 $\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ 所得的方阵.

8.4　特征值与特征向量

定义 8.4.1. 设 $A \in M_n(\mathbf{C})$, 称 $\lambda \in \mathbf{C}$ 是 A 的 (复) 特征值, 如果存在非零向量 $\mathbf{x} \in \mathbf{C}^n$ 使得 $A\mathbf{x} = \lambda\mathbf{x}$. 称满足此条件的非零向量 $\mathbf{x}$ 为 A 的一个属于特征值 λ 的 (复) 特征向量 (*eigenvector*).

若 $A \in M_n(\mathbf{R})$, 称实数 λ 为 A 的一个特征值 (*eigenvalue*), 如果存在非零向量 $\mathbf{x} \in \mathbf{R}^n$ 使得 $A\mathbf{x} = \lambda\mathbf{x}$. 称满足此条件的非零向量 $\mathbf{x}$ 为 A 的一个属于特征值 λ 的特征向量 (*eigenvector*).

命题 8.4.2. 属于不同特征值的特征向量线性无关. 具体的说, 设 $A \in M_n(\mathbf{C})$, $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 是 A 的特征向量, 且属于不同的特征值, 则 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 线性无关.

命题 8.4.3. *λ 是 A 的 (复) 特征值的充分必要条件是 $\det(\lambda I_n - A) = 0$, 即 λ 是 A 的特征多项式的根.*

定义 8.4.4. 对 n 阶方阵 A, 令

$$p_A(\lambda) = \det(\lambda I_n - A) = \det \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix},$$

称之为矩阵 A 的特征多项式, 它是 λ 的 n 次多项式, 首项系数为 1.

命题 8.4.5. *设 A 是 n 阶方阵, 则它的特征多项式形如*

$$p_A(\lambda) = \lambda^n - trace(A)\lambda^{n-1} + \cdots + (-1)^n \det(A).$$

进一步, A 的所有 n 个特征值 (计重数) 的和等于 $trace(A)$, 所有 n 个特征值的乘积等于 $\det(A)$.

8.5 对角化

定义 8.5.1. 称 $A \in M_n(\mathbf{C})$ 是可对角化的, 如果存在可逆复矩阵 S 使得 $S^{-1}AS$ 是对角矩阵. 进一步, 称满足如上条件的矩阵 S 对角化 A.

对方阵 $A \in M_n(\mathbf{C})$, 其特征多项式可能有重根. 如果 $p_A(\lambda)$ 分解为

$$p_A(\lambda) = (\lambda - z_1)^{n_1} \cdots (\lambda - z_r)^{n_r},$$

其中 $z_1, \ldots, z_r$ 是不同的复数, $n_1, \ldots, n_l$ 是正整数, 满足 $n_1 + \ldots + n_r = n$, 则称 z_i 是 $p_A(\lambda)$ 的 n_i 重根, 并称 z_i 作为 A 的特征值的代数重数为 n_i.

对 A 的特征值 λ, 令

$$V_\lambda = \{\mathbf{x} \in \mathbf{C}^n | A\mathbf{x} = \lambda\mathbf{x}\},$$

它是 $\mathbf{C}^n$ 的子空间, 称之为 A 的属于 λ 的特征子空间. 称特征子空间 V_λ(在复数域 $\mathbf{C}$ 上) 的维数 $\dim(V_\lambda)$ 为 λ 作为 A 的特征值的几何重数.

命题 8.5.2. *特征值的几何重数不超过其代数重数.*

定义 8.5.3. 称 $A \in M_n(\mathbf{C})$ 的特征值 λ_0 为半单特征值, 如果 λ_0 的几何重数等于代数重数, 否则称 λ_0 为亏损特征值.

定理 8.5.4. (1) 复方阵可对角化的充分必要条件是它的所有特征值都是半单的.

(2) 实方阵 (在 $\mathbf{R}$ 上) 可对角化的充分必要条件是其特征多项式的根都是实数, 且其特征值都半单.

8.6 相似

定义 8.6.1. 设 A, B 是 n 阶方阵, 称 A 与 B 相似, 或 A 相似于 B, 如果存在 n 阶可逆矩阵 S 使得 $S^{-1}AS = B$.

命题 8.6.2. 设方阵 A 与 B 相似, 则有

(1) A, B 的特征多项式相同.

(2) A 与 B 的特征值一样, 且每个特征值作为 A 特征值的几何 (代数) 重数等于作为 B 特征值的几何 (代数) 重数.

定理 8.6.3 (Cayley-Hamilton)**.** 设方阵 A 的特征多项式为 $p_A(\lambda)$, 则有 $p_A(A) = O$.

定义 8.6.4. 对复数 λ 与正整数 m, 定义 m 阶 Jordan 块矩阵为

$$J_m(d) = \begin{pmatrix} d & 1 & & & \\ & d & 1 & & \\ & & \ddots & \ddots & \\ & & & d & 1 \\ & & & & d \end{pmatrix},$$

也称之为一个 d-Jordan 块.

Jordan 块 $J_m(d)$ 的特征多项式为

$$\det(\lambda I_m - J_m(d)) = (\lambda - d)^m,$$

故 $J_m(d)$ 有唯一的特征值 d, 其代数重数为 m. 另一方面, 特征值 d 的特征子空间为

$$V_d = \{\mathbf{x} \in \mathbf{R}^m | (dI_m - J_m(d))\mathbf{x} = \mathbf{0}\} = \mathrm{span}(\{\mathbf{e}_1\}),$$

可得特征值 d 的几何重数为 1.

定理 8.6.5 (Jordan 标准型)**.** 每个 n 阶复方阵都相似于如下形式的矩阵

$$\begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_r}(\lambda_r) \end{bmatrix},$$

其中 $J_{n_i}(\lambda_i)$ 是 n_i 阶 *Jordan* 块矩阵, $n_1 + \cdots + n_r = n$.

设 n 阶复方阵 A 相似于定理8.6.5中所述的标准型矩阵, 则有:

(1) A 的全部特征值为 n_1 个 λ_1,..., n_r 个 λ_r, 其中 $\lambda_1, \ldots, \lambda_r$不必 两两不同;

(2) A 的特征值 λ 的代数重数等于标准型中 λ-Jordan 块的阶数的总和;

(3) A 的特征值 λ 的几何重数等于标准型中 λ-Jordan 块的个数.

(4) 对 A 的特征值 λ, 可通过考察 $\dim\ker((A-\lambda I_n)^t)(t \in \mathbf{Z}_+)$ 完全确定标准型中 λ-Jordan 块的阶数信息:

$$\dim\ker\left((A-\lambda I_n)^t\right) = \sum_{1\le i\le r, \lambda_i=\lambda} \min\{t, n_i\}.$$

由此可得, A 的标准型中 $J_m(\lambda)$ 块的数目等于

$$2\dim\ker((A-\lambda I_n)^m) - \dim\ker\left((A-\lambda I_n)^{m-1}\right) - \dim\ker\left((A-\lambda I_n)^{m+1}\right),$$

其中约定 $\dim\ker((A-\lambda I_n)^0) = 0$.

8.7 实对称矩阵的谱分解

命题 8.7.1. *实对称矩阵的特征多项式的根都是实数, 即实对称矩阵在 $\mathbf{C}$ 上的特征值都是实数.*

定理 8.7.2 (实对称矩阵的谱分解)**.** *设 A 是 n 阶实对称矩阵, 则存在实正交矩阵 Q, 使得 Q^TAQ 是实对角矩阵. 特别的, $Q^{-1}AQ$ 是实对角矩阵, 即 A 在 $\mathbf{R}$ 上可对角化.*

命题 8.7.3. *实对称矩阵不同特征值的 (实) 特征向量相互正交.*

定义 8.7.4. 称实方阵 A 与 B 正交相似, 如果存在实正交矩阵 Q 使得 $Q^TAQ = B$.

这样, 可把定理8.7.2叙述为: 实对称矩阵一定正交相似于对角矩阵. 对于实对称矩阵 A, 存在正交矩阵 Q 使得

$$Q^{-1}AQ = Q^TAQ = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}.$$

令 $\{\mathbf{e}_i\}$ 为 $\mathbf{R}^n$ 的标准基, 令 $\mathbf{v}_i = Q\mathbf{e}_i$, 由 Q 正交可知 $\{\mathbf{v}_i\}$ 为 $\mathbf{R}^n$ 的单位正交基, 且有 $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$. 这样, 就找到了 $\mathbf{R}^n$ 的单位正交基, 其成员都是 A 的特征向量.

8.8 二次型

可以把 n 阶实对称矩阵等同成 $\mathbf{R}^n$ 上的二次函数. 对于 n 阶实对称矩阵 A, 它唯一确定如下二次函数

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = \langle \mathbf{x}, A\mathbf{x}\rangle = \sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j = \sum_{i=1}^n a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_ix_j,$$

称之为由 A 确定的二次型. 人们经常关心 $Q(x)$ 的符号.

定义 8.8.1. 设 A 是 n 阶实对称矩阵, 令 $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ 为由 A 给出的二次型.

(1) 称 A 是正定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) > 0$;

(2) 称 A 是半正定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) \geq 0$;

(3) 称 A 是负定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) < 0$;

(4) 称 A 是半负定的, 如果对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 都有 $Q(\mathbf{x}) \leq 0$;

(5) 称 A 是不定的, 如果 $Q(\mathbf{x})$ 既能取正值又能取负值.

定理 8.8.2 (正定矩阵的刻画). *设 A 是实对称矩阵, 则以下命题彼此等价:*

(1) A 是正定矩阵;

(2) A 的特征值都是正数;

(3) 存在可逆矩阵 B 使得 $A = B^T B$;

(4) A 的顺序主子式都大于零.

定义 8.8.3. 设 A 是 n 阶实对称矩阵, $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ 是 A 给出的二次型. 对非零向量 $\mathbf{x}$, 称

$$Q(\frac{\mathbf{x}}{||\mathbf{x}||}) = \frac{\mathbf{x}^T A\mathbf{x}}{||\mathbf{x}||^2} = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

为 $\mathbf{x}$ 关于 A 的 Rayleigh 商.

命题 8.8.4. *设 A 是 n 阶实对称矩阵, 则 $\frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ 在 $\mathbf{R}^n \backslash \{\mathbf{0}\}$ 上的最大与最小值分别等于 A 的最大与最小特征值.*

8.9 奇异值分解

定理 8.9.1 (奇异值分解). *设 A 是 $m \times n$ 的实矩阵, 则存在 m 阶正交矩阵 U 和 n 阶正交矩阵 V, 使得 $A = U\Sigma V^T$, 其中 Σ 是如下形式的分块矩阵:*

$$\Sigma = \begin{bmatrix} diag\{\sigma_1, \ldots, \sigma_r\} & O \\ O & O \end{bmatrix}, \quad \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

证明: $A^T A$ 是 n 阶实对称矩阵, 由定理8.7.2, 存在 $A^T A$ 的一组特征向量 $\{\mathbf{v}_i\}_{1\leq i\leq n}$ 构成 $\mathbf{R}^n$ 的单位正交基. 设 $\mathbf{v}_i$ 是 $A^T A$ 的属于特征值 λ_i 的特征向量, 适当排序后不妨设 $\lambda_1 \geq \cdots \geq \lambda_n$. 对任何 $i, j \leq n$, 有

$$\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = \langle \mathbf{v}_i, A^T A\mathbf{v}_j\rangle = \langle \mathbf{v}_i, \lambda_j\mathbf{v}_j\rangle = \lambda_j\delta_{ij},$$

其中 δ_{ij} 是 Kronecker 记号, 当 $i \neq j$ 时 $\delta_{ij} = 0$, 当 $i = j$ 时 $\delta_{ij} = 1$. 在上式中取 $i = j$ 可知 $\lambda_i \geq 0$, 且 $\lambda_i = 0 \iff A\mathbf{v}_i = \mathbf{0}$. 另外, 当 $i \neq j$ 时, 有 $\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = 0$, 说明不同的 $A\mathbf{v}_i$ 彼此正交.

设 r 是满足 $\lambda_r > 0$ 的最大的指标, 由前述有 $A\mathbf{v}_{r+1} = \cdots = A\mathbf{v}_n = \mathbf{0}$, 且 $\{\frac{A\mathbf{v}_1}{\sqrt{\lambda_1}}, \cdots, \frac{A\mathbf{v}_r}{\sqrt{\lambda_r}}\}$ 是 $\mathbf{R}^m$ 的单位正交向量组. 令 $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}(1 \le i \le r)$, 并将其扩充为 $\mathbf{R}^m$ 的单位正交基 $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$. 这样就有

$$A[\mathbf{v}_1, \dots, \mathbf{v}_n] = [\mathbf{u}_1, \dots, \mathbf{u}_m] \begin{bmatrix} \mathrm{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}\} & O \\ O & O \end{bmatrix}.$$

令 $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, $U = [\mathbf{u}_1, \dots, \mathbf{u}_m]$, 则 V, U 是正交矩阵, 且有

$$A = U \begin{bmatrix} \mathrm{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}\} & O \\ O & O \end{bmatrix} V^T$$

. □

定义 8.9.2. 设 A 是 $m \times n$ 实矩阵, 称半正定方阵 A^TA 的特征值的非负平方根为 A 的奇异值.

设 A 给出的线性映射为 $f : \mathbf{R}^n \to \mathbf{R}^m$. 在定理8.9.1的证明中, 我们找到了 $\mathbf{R}^n$ 的单位正交基 $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 与 $\mathbf{R}^m$ 的单位正交基 $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$, 使得

$$f(\mathbf{v}_i) = \begin{cases} \sqrt{\lambda_i} \cdot \mathbf{u}_i, & \text{如果} i \le r; \\ \mathbf{0}, & \text{如果} i > r. \end{cases}$$

这样, 就对 f 的定义域与值域给出了很清楚的描述: $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ 是 $\ker(f)$ 的单位正交基; $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ 是 $\mathrm{Im}(f)$ 的单位正交基; 如果令

$$X = \mathrm{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_r\}), \quad Y = \mathrm{span}(\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}),$$

则给出正交直和分解

$$\mathbf{R}^n = X \oplus \ker(f), \quad \mathbf{R}^m = \mathrm{Im}(f) \oplus Y,$$

在此分解下, f 将 X 线性同构成 $\mathrm{Im}(f)$, 在 $\ker(f)$ 上恒为零. 定理8.9.1就是用矩阵的语言来叙述这些信息.

由 $\{\mathbf{v}_i\}_{1 \le i \le n}$ 是 $\mathbf{R}^n$ 的单位正交基, 任何向量 $\mathbf{x}$ 可表示成

$$\mathbf{x} = \sum_{i=1}^n \mathbf{v}_i \langle \mathbf{v}_i, \mathbf{x} \rangle,$$

由此可得

$$f(\mathbf{x}) = \sum_{i=1}^r \sqrt{\lambda_i} \cdot \mathbf{u}_i \langle \mathbf{v}_i, \mathbf{x} \rangle = \sum_{i=1}^r \sqrt{\lambda_i} \cdot \mathbf{u}_i (\mathbf{v}_i^T \mathbf{x}) = \left(\sum_{i=1}^r \sqrt{\lambda_i} \cdot \mathbf{u}_i \mathbf{v}_i^T \right) \mathbf{x}.$$

上式说明 $\sum_{i=1}^{r}\sqrt{\lambda_i}\cdot\mathbf{u}_i\mathbf{v}_i^T$ 是 f 的表示矩阵, 所以有

$$A=\sum_{i=1}^{r}\sqrt{\lambda_i}\cdot\mathbf{u}_i\mathbf{v}_i^T=U_r\Sigma_rV_r^+,$$

其中 $U_r=[\mathbf{u}_1,\ldots,\mathbf{u}_r]$, $V_r=[\mathbf{v}_1,\ldots,\mathbf{v}_r]$, $\Sigma_r=\mathrm{diag}\{\sqrt{\lambda_1},\ldots,\sqrt{\lambda_r}\}$. 称上式为 A 的简化奇异值分解.

如果 A 是可逆矩阵, 则 A 的奇异值分解 (SDV) 中 $r=m=n$, 且 A 的逆矩阵为

$$A^{-1}=(U\Sigma V^T)^{-1}=V\Sigma^{-1}U^T.$$

这就启发人们对一般的 (不一定可逆, 甚至不一定是方阵的) 矩阵 A 定义如下 $n\times m$ 矩阵:

$$A^+=\sum_{i=1}^{r}\frac{1}{\sqrt{\lambda_i}}\cdot\mathbf{v}_i\mathbf{u}_i^T=V_r\Sigma_r^{-1}U_r^+,$$

称之为 A 的 Moore-Penrose 广义逆, 简称广义逆. A^+ 满足

$$A^+A=V_rV_r^T,\quad AA^+=U_rU_r^T.$$

这样, A^+A 是从 $\mathbf{R}^n$ 向 $\mathcal{N}(A)^\perp=\mathcal{R}(A^T)$ 的正交投影, AA^+ 是从 $\mathbf{R}^m$ 向 $\mathcal{R}(A)$ 的正交投影. 由此可用广义逆求解最小二乘问题.

定理 8.9.3. *$\mathbf{x}_0=A^+\mathbf{b}$ 是方程组 $A\mathbf{x}=\mathbf{b}$ 的一个最小二乘解, 且是模长最小的最小二乘解.*

证明: $\mathbf{x}$ 是最小二乘解当且仅当 $A\mathbf{x}=P_A\mathbf{b}$. 由广义逆的定义, AA^+ 是从 $\mathbf{R}^m$ 向 $\mathcal{R}(A)$ 的正交投影, 即有 $P_A=AA^+$. 结合这两点, 可知对于 $\mathbf{x}_0=A^+\mathbf{b}$, 有

$$A\mathbf{x}_0=A(A^+\mathbf{b})=(AA^+)\mathbf{b}=P_A\mathbf{b},$$

表明 $\mathbf{x}_0=A^+\mathbf{b}$ 是最小二乘解.

另一方面, 对任何别的最小二乘解 $\mathbf{x}$, 有 $A\mathbf{x}=A\mathbf{x}_0$, 可知 $\mathbf{x}-\mathbf{x}_0\in\mathcal{N}(A)$. 注意到,

$$\mathbf{x}_0=A^+\mathbf{b}=V_r\Sigma_r^{-1}U_r^+\mathbf{b}\in\mathrm{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}=\mathcal{N}(A)^\perp.$$

特别的有 $\mathbf{x}_0\perp(\mathbf{x}-\mathbf{x}_0)$, 由此可得

$$||\mathbf{x}||^2=||\mathbf{x}_0||^2+||\mathbf{x}-\mathbf{x}_0||^2\geq||\mathbf{x}_0||^2,$$

这就证明了 $\mathbf{x}_0=A^+\mathbf{b}$ 是模长最小的最小二乘解. □

8.10 向量的坐标表示

设 V 是 $\mathbf{F}$ 上的有限维线性空间, $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 是它的一组基. 任何 $\mathbf{v}\in V$ 可唯一的表示成这组基的线性组合

$$\mathbf{v}=x_1\mathbf{e}_1+\cdots+x_n\mathbf{e}_n,\quad x_1,\ldots,x_n\in\mathbf{F}. \tag{8.1}$$

称向量 $\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\in\mathbf{F}^n$ 为 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标, 并把(8.1)写作

$$\mathbf{v}=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}.$$

可从映射的角度理解基. 上述基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 确定了线性映射 $\Phi:\mathbf{F}^n\to V$,

$$\Phi\left(\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\right)=x_1\mathbf{e}_1+\cdots+x_n\mathbf{e}_n.$$

显然 Φ 是线性同构, 其逆 $\Phi^{-1}(\mathbf{v})$ 就是给出 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标. 这样, 给定 V 的一个基, 等价于给出 V 到 $\mathbf{F}^n$ 的一个线性同构.

设 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的另一组基, 则 $\mathbf{v}$ 在 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下有坐标表示

$$\mathbf{v}=[\mathbf{v}_1,\ldots,\mathbf{v}_n]\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}.$$

我们来确定同一个 $\mathbf{v}$ 在不同基下的坐标表示之间的关系. 为此, 对每个 $j\le n$, 设

$$\mathbf{v}_j=T_{1j}\mathbf{e}_1+\cdots+T_{nj}\mathbf{e}_n=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix}T_{1j}\\ \vdots\\ T_{nj}\end{pmatrix},$$

则可形式化的写成

$$[\mathbf{v}_1,\ldots,\mathbf{v}_n]=[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix}T_{11}&\cdots&T_{1n}\\ \vdots&\ddots&\vdots\\ T_{n1}&\cdots&T_{nn}\end{pmatrix}.$$

记 $T=(T_{ij})$, 称之为从基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 到基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的过渡矩阵, 它是可逆矩阵, 其逆 T^{-1} 是从基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 到基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 的过渡矩阵.

设 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标为 $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{F}^n$, 基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下的坐标为 $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbf{F}^n$, 则有

$$[\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = [\mathbf{v}_1,\ldots,\mathbf{v}_n]\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = [\mathbf{e}_1,\ldots,\mathbf{e}_n]\begin{pmatrix} T_{11} & \cdots & T_{1n} \\ \vdots & \ddots & \vdots \\ T_{n1} & \cdots & T_{nn} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

由此可得

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = T\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

即 $\mathbf{v}$ 在基 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 下的坐标等于 T 乘以它在基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 下的坐标.

8.11 线性映射

设 V,W 是 $\mathbf{F}$ 上的有限维线性空间, $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 与 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 分别是 V 与 W 的基. 线性映射 $f:V\to W$ 由它在 V 的基上的值 $f(\mathbf{v}_1),\ldots,f(\mathbf{v}_n)$ 唯一确定, 设

$$f(\mathbf{v}_j)=\sum_{i=1}^{m}F_{ij}\mathbf{w}_i,\quad j=1,\ldots,n. \tag{8.2}$$

令 $F=(F_{ij})_{1\le i\le m,1\le j\le n}$, 称之为 f 在前述两组基下的表示矩阵. 由此可将(8.2)形式化的记为

$$[f(\mathbf{v}_1),\ldots,f(\mathbf{v}_n)]=[\mathbf{w}_1,\ldots,\mathbf{w}_m]F.$$

利用表示矩阵, 可确定 $\mathbf{v}$ 的坐标 $\mathbf{x}=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 与 $f(\mathbf{v})$ 的坐标 $\mathbf{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ 之间的关系:

$$\begin{aligned}[\mathbf{w}_1,\ldots,\mathbf{w}_m]\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = f(\mathbf{v}) = f\left([\mathbf{v}_1,\ldots,\mathbf{v}_n]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) \\ =[f(\mathbf{v}_1),\ldots,f(\mathbf{v}_n)]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = [\mathbf{w}_1,\ldots,\mathbf{w}_m]F\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.\end{aligned}$$

由此可得

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = F\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

即像点的坐标等于表示矩阵乘以原像点的坐标.

命题 8.11.1 (复合映射的表示矩阵等于表示矩阵的乘积)**.** *设 U,V,W 都是有限维线性空间, 为它们各选定一组基. 设 $f:U\to V$ 与 $g:V\to W$ 是线性映射, 则在前述选定的基下, 复合映射 $g\circ f$ 的表示矩阵等于 g 的表示矩阵 G 乘以 f 的表示矩阵 F, 即 $g\circ f$ 的表示矩阵为 GF.*

命题 8.11.2 (换基前后的表示矩阵相抵)**.** *设从 V 的基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 到另一个基 $\mathbf{v}'_1,\ldots,\mathbf{v}'_n$ 的过渡矩阵为 T; 从 W 的基 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 到另一个基 $\mathbf{w}'_1,\ldots,\mathbf{w}'_m$ 的过渡矩阵为 S. 如果线性映射 $f:V\to W$ 在基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 与 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 下的表示矩阵为 F, 则 $f:V\to W$ 在基 $\mathbf{v}'_1,\ldots,\mathbf{v}'_n$ 与 $\mathbf{w}'_1,\ldots,\mathbf{w}'_m$ 下的表示矩阵为 $S^{-1}FT$.*

证明: 将 $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ 简记为 $[\mathbf{v}]$, 对其他基也采用同样的记法, 则有

$$[\mathbf{v}']=[\mathbf{v}]T,\quad [\mathbf{w}']=[\mathbf{w}]S,\quad [f(\mathbf{v})]=[\mathbf{w}]F,$$

由此可得

$$[f(\mathbf{v}')]=f([\mathbf{v}]T)=[f(\mathbf{v})]T=[\mathbf{w}]FT=[\mathbf{w}']S^{-1}FT,$$

这表明 f 在基 $[\mathbf{v}']$ 与 $[\mathbf{w}']$ 下的表示矩阵为 $S^{-1}FT$. □

对于线性空间 V 到自身的线性映射 $f:V\to V$(称为 V 上的线性变换), 我们为定义域与值域的 V 选取同样的基 $[\mathbf{v}]$, 由此可谈论 f 的表示矩阵 F. 设 $[\mathbf{v}']$ 是另一组基, 且从 $[\mathbf{v}]$ 到 $[\mathbf{v}']$ 的过渡矩阵为 S, 则由命题8.11.2 可知, f 在基 $[\mathbf{v}']$ 下的表示矩阵为 $S^{-1}FS$. 也就是说, 线性变换在换基前后的表示矩阵相似.

第九章　附录二: 习题参考答案

9.1　线性空间与映射

例 9.1.1. 设 A, B 是 n 阶方阵, 满足 $AB = BA = 0$, $\mathrm{rank}(A^2) = \mathrm{rank}(A)$. 证明: $\mathrm{rank}(A+B) = \mathrm{rank}(A) + \mathrm{rank}(B)$.

证明: 设 A, B 对应的线性映射分别为 f, g, 则条件可以叙述为 $g \circ f = f \circ g = 0$, 且 $\mathrm{im}(f \circ f) = \mathrm{im}(f)$.

由 $\mathrm{im}(f \circ f) = \mathrm{im}(f)$ 可知 $f(\mathrm{im}(f)) = \mathrm{im}(f)$, 同维数线性空间之间的满射一定是线性同构, 故 $f : \mathrm{im}(f) \to \mathrm{im}(f)$ 是线性同构. 特别的, $\mathrm{im}(f) \cap \ker(f) = \{0\}$. 由此可得 $\dim(\mathrm{im}(f) + \ker(f)) = \dim(\mathrm{im}(f)) + \dim(\ker(f)) = n$, 所以有 $\mathrm{im}(f) \oplus \ker(f) = \mathbf{R}^n$.

在此分解下, 利用条件 $g \circ f = f \circ g = 0$ 可得 g 在 $\mathrm{im}(f)$ 上恒为零, g 把 $\ker(f)$ 映入 $\ker(f)$(这是因为 $\mathrm{Im}(g) \subseteq \ker(f)$). 结合 f 把 $\mathrm{im}(f)$ 映入 $\mathrm{im}(f)$, f 在 $\ker(f)$ 上恒为零, 可得 $\mathrm{im}(f+g) = \mathrm{im}(f) \oplus \mathrm{im}(g)$, 即有 $\mathrm{rank}(A+B) = \mathrm{rank}(A) + \mathrm{rank}(B)$. □

例 9.1.2. 设 A, B 是 n 阶方阵, 满足 $(A+B)^2 = A+B$, 且 $\mathrm{rank}(A+B) = \mathrm{rank}(A)+\mathrm{rank}(B)$. 证明: $A^2 = A, B^2 = B, AB = BA = 0$.

证明: 把矩阵 A, B 给出的线性映射也记为 A, B. 注意到 $\mathrm{Im}(A+B) \subseteq \mathrm{Im}(A) + \mathrm{Im}(B)$, 可得

$$\begin{aligned} rk(A+B) &= dim Im(A+B) \\ &\le dim\,(Im(A) + Im(B)) \\ &= dim Im(A) + dim Im(B) - dim\,(Im(A) \cap Im(B)) \\ &\le rk(A) + rk(B). \end{aligned}$$

由题目条件可知上述不等式都取等号, 从而有

$$Im(A+B) = Im(A) + Im(B), \quad Im(A) \cap Im(B) = \{0\}.$$

利用条件 $(A+B)^2=A+B$, 对任何 x, 有

$$(A^2+AB-A)x=(B-B^2-BA)x\in Im(A)\cap Im(B)=\{0\},$$

即有 $A^2+AB=A$, $B^2+BA=B$.

对任何 x, 由于 $Bx\in Im(A)+Im(B)=Im(A+B)$, 存在 u 使得 $Bx=(A+B)u$, 由此可得

$$Au=B(x-u)\in Im(A)\cap Im(B)=\{0\},$$

即有 $Au=0$, $Bu=Bx$. 这样, 结合 $B^2+BA=B$ 可知

$$BBx-Bx=BBu-Bu=BAu=0,$$

这就证明了 $B^2=B$. 对称的, 有 $A^2=A$, 从而完成了证明.

□

例 9.1.3. 求如下矩阵的逆矩阵:

$$A_n=\begin{pmatrix} 1 & 2 & \cdots & n \\ n & 1 & \cdots & n-1 \\ \vdots & & & \vdots \\ 2 & 3 & \cdots & 1 \end{pmatrix}.$$

解答. 考虑映射 $\tau:\mathbf{R}^n\to\mathbf{R}^n$ 为 $\tau(x_1,x_2,\cdots,x_n)=(x_n,x_1,\cdots,x_{n-1})$, 它的效果是把最后一个分量前置. 令 $\alpha=(1,2,\cdots,n)$ 为 A_n 的第一个行向量, 则 A_n 的行向量依次为 $\alpha,\tau\alpha,\cdots\tau^{n-1}\alpha$. 我们只需找向量 β, 满足

$$\alpha\cdot\beta=1,\quad (\tau\alpha)\cdot\beta=0,\quad \cdots,\quad (\tau^{n-1}\alpha)\cdot\beta=0$$

即可知道 $(A_n)^{-1}$ 的列向量依次为 $\beta^T,(\tau\beta)^T,\cdots(\tau^{n-1}\beta)^T$. 求解 β 等价于求解方程 $A_n\beta^T=\mathbf{e}_1$, 即

$$\begin{cases} 1x_1+2x_2+\cdots+nx_n=1, \\ nx_1+1x_2+\cdots+(n-1)x_n=0 \\ (n-1)x_1+nx_2+\cdots+(n-2)x_n=0, \\ \cdots \\ 2x_1+3x_2+\cdots+1x_n=0, \end{cases}$$

对每个 $2\le i\le n-1$, 考虑上述第 i 个方程与第 $i+1$ 个方程的差, 可得

$$x_2=x_3=\cdots=x_{n-1}=\frac{x_1+x_2+\cdots+x_n}{n},$$

由此可以解出 β.

□

例 9.1.4. 设 A, B 分别为 $m\times n, n\times k$ 矩阵. 证明: $\mathrm{rank}(AB)\geq \mathrm{rank}(A)+\mathrm{rank}(B)-n$.

证明: 设 A, B 给出的线性映射分别为 $f:\mathbf{R}^n\to\mathbf{R}^m$ 与 $g:\mathbf{R}^k\to\mathbf{R}^n$. 令 $V=\mathrm{Im}(g)$, 对线性映射 $f:V\to\mathbf{R}^m$ 用维数公式可得

$$\begin{aligned}\dim f(V)&=\dim V-\dim(\ker(f)\cap V)\\&\geq \dim V-\dim(\ker(f))\\&=\dim V-(n-\dim(\mathrm{Im}(f)))\\&=\dim\mathrm{Im}(g)+\dim\mathrm{Im}(f)-n,\end{aligned}$$

此即 $\mathrm{rank}(AB)\geq \mathrm{rank}(A)+\mathrm{rank}(B)-n$. □

例 9.1.5. 设 A, B, C 分别为 $m\times n, n\times k, k\times s$ 矩阵. 证明:

$$\mathrm{rank}(AB)+\mathrm{rank}(BC)\leq \mathrm{rank}(ABC)+\mathrm{rank}(B).$$

证明: 也用 A, B, C 表示它们给出的线性映射. 记 $V=\mathrm{Im}(B), W=\mathrm{Im}(B\circ C)$, 对线性映射 $A:V\to\mathbf{R}^m$ 与 $A:W\to\mathbf{R}^m$ 用维数公式, 可得

$$\begin{aligned}&\dim A(V)=\dim(V)-\dim(V\cap\ker(A)),\\&\dim A(W)=\dim(W)-\dim(W\cap\ker(A)).\end{aligned}$$

注意到 $W\subseteq V$, 有 $\dim(W\cap\ker(A))\leq\dim(V\cap\ker(A))$, 结合以上两个等式可得

$$\begin{aligned}&\mathrm{rank}(AB)-\mathrm{rank}(ABC)\\=&\dim A(V)-\dim A(W)\\=&\dim(V)-\dim(V\cap\ker(A))-\dim(W)+\dim(W\cap\ker(A))\\\leq&\dim(V)-\dim(W)\\=&\mathrm{rank}(B)-\mathrm{rank}(BC).\end{aligned}$$

□

例 9.1.6. 设 A 是 n 阶方阵. 证明: 存在正整数 $k\leq n$, 使得 $\mathrm{rank}(A^k)=\mathrm{rank}(A^{k+1})=\cdots$.

证明: 把 A 给出的线性映射也记作 A, 并令 $A^k=A\circ\cdots\circ A$ 为 A 的 k 次迭代映射, 约定 $A^0=\mathrm{Id}_{\mathbf{R}^n}$. 令 $V_k=\mathrm{Im}(A^k)$ 为 A^k 的像, 则有 $V_{k+1}=A(V_k)$. 显然 $V_{k+1}\subseteq V_k$, 有

$\dim V_{k+1} \leq \dim V_k$. 注意到

$$\sum_{i=0}^{n} (\dim V_k - \dim V_{k+1}) = n - \dim(V_{n+1}) \leq n,$$

上式左边是 $n+1$ 个非负整数的和, 由于总和不超过 n, 其中必有一个等于零, 即存在 $0 \leq k \leq n$ 使得 $\dim V_k = \dim V_{k+1}$. 结合 $V_{k+1} \subseteq V_k$ 可知 $V_{k+1} = V_k$, 再由 $V_{k+1} = A(V_k)$ 可得 $A(V_k) = V_k$. 反复利用 $V_{i+1} = A(V_i)$ 可得

$$V_k = V_{k+1} = V_{k+2} = \cdots,$$

从而完成了证明. □

例 9.1.7. 证明: 任何秩为 $r > 0$ 的矩阵可以分解成 r 个秩为 1 的矩阵的和.

证明: 利用《线性代数入门》命题 2.3.16, A 与 D_r 相抵, 即存在可逆方阵 P, Q 使得 $PAQ = D_r$. 将 D_r 中的 I_r 分解为 r 个秩为 1 的方阵之和即可, 这是显然的, 只需每次取出 I_r 的单独一列. □

例 9.1.8 (Steinitz 替换定理)**.** 设 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_r\}$ 是线性无关的向量组, 可被 $T = \{\mathbf{b}_1, \cdots, \mathbf{b}_t\}$ 线性表示. 证明: 可从 T 中删去 r 个向量, 再并上 S, 使得得到的向量组 T' 与 T 线性等价.

证明: 对 r 用归纳法. 当 $r = 1$ 时, $\mathbf{a}_1$ 非零, 且可表示为

$$\mathbf{a}_1 = \sum_{i=1}^{t} k_i \mathbf{b}_i,$$

则 $k_i (1 \leq i \leq t)$ 不全为零. 不妨设 $k_1 \neq 0$, 则有

$$\mathbf{b}_1 = \frac{\mathbf{a}_1 - \sum\limits_{i=2}^{t} k_i \mathbf{b}_i}{k_1}.$$

由此可知 $T' = \{\mathbf{a}_1, \mathbf{b}_2, \cdots, \mathbf{b}_t\}$ 与 T 线性等价.

设 $r < m$ 时题述命题成立, 来考虑 $r = m \geq 2$ 的情形. 由于 $\{\mathbf{a}_2, \cdots, \mathbf{a}_m\}$ 线性无关且可被 T 表示, 利用归纳假设可把 T 中某 $m-1$ 个向量替换成 $\{\mathbf{a}_2, \cdots, \mathbf{a}_m\}$, 使得所得的向量组 T' 与 T 线性等价. 记 $p = t - (m-1)$, 不妨设

$$T' = \{\mathbf{a}_2, \cdots, \mathbf{a}_m\} \cup \{\mathbf{b}_1, \cdots, \mathbf{b}_p\}.$$

由条件有 $\mathbf{a}_1 \in \operatorname{span}(T) = \operatorname{span}(T')$, 设

$$\mathbf{a}_1 = \sum_{i=2}^{m} x_i \mathbf{a}_i + \sum_{j=1}^{p} y_j \mathbf{b}_j,$$

则 $y_j(1 \le j \le p)$ 不能全为零, 否则的话 $\mathbf{a}_1$ 可用 $\{\mathbf{a}_2, \cdots, \mathbf{a}_m\}$ 线性表示, 与条件 $S = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 线性无关矛盾! 不妨设 $y_1 \ne 0$, 则有

$$\mathbf{b}_1 = \frac{1}{y_1}\left(\mathbf{a}_1 - \sum_{i=2}^{m} x_i \mathbf{a}_i - \sum_{j=2}^{p} y_j \mathbf{b}_j\right),$$

由此可知 $T'' = \{\mathbf{a}_1, \cdots, \mathbf{a}_m\} \cup \{\mathbf{b}_2, \cdots, \mathbf{b}_p\}$ 与 T' 线性等价. 这样, $\mathrm{span}(T) = \mathrm{span}(T') = \mathrm{span}(T'')$, 且 T'' 是将 T 中 m 个向量替换成 $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ 所得的向量组. □

例 9.1.9. 设 A, B 是 n 阶方阵, 满足 $A^2 = A$, $B^2 = B$, 且 $I_n - A - B$ 是可逆矩阵. 证明: $\mathrm{rank}(A) = \mathrm{rank}(B)$.

证明: 注意到

$$A(I_n - A - B) = A - A^2 - AB = -AB, \quad (I_n - A - B)B = B - AB - B^2 = -AB,$$

结合 $I_n - A - B$ 可逆, 有

$$B = (I_n - A - B)^{-1} A (I_n - A - B).$$

特别的, A 与 B 相抵, 因而有相同的秩.

□

例 9.1.10. 设 A 是 n 阶方阵.

(1) 证明: $A^2 = I_n$ 的充分必要条件是 $\mathrm{rk}(A - I_n) + \mathrm{rk}(A + I_n) = n$.

(2) 证明: $A^2 = A$ 的充分必要条件是 $\mathrm{rk}(A - I_n) + \mathrm{rk}(A) = n$.

证明: (1) 令 $V_+ = \ker(A - I)$, $V_- = \ker(A + I)$, 则有

$$V_+ = \{\mathbf{x} | A\mathbf{x} = \mathbf{x}\}, \quad V_- = \{\mathbf{x} | A\mathbf{x} = -\mathbf{x}\}.$$

显然 $V_+ \cap V_- = \{\mathbf{0}\}$, 由此利用维数公式可得

$$\mathrm{rk}(A - I_n) + \mathrm{rk}(A + I_n) = n \iff \dim V_+ + \dim V_- = n \iff \dim(V_+ + V_-) = n \iff V_+ + V_- = \mathbf{R}^n.$$

(a) 当 $V_+ + V_- = \mathbf{R}^n$ 时, 任何 $\mathbf{x}\mathbf{R}^n$ 可以表示成 $\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_-$, 其中 $\mathbf{x}_\pm \in V_\pm$, 可得

$$A^2\mathbf{x} = A^2\mathbf{x}_+ + A^2\mathbf{x}_- = \mathbf{x}_+ + \mathbf{x}_- = \mathbf{x},$$

说明 $A^2 = I$.

(b) 当 $A^2 = I$ 时, 对每个 $\mathbf{x}$, 令

$$\mathbf{x}_+ = \frac{1}{2}(A\mathbf{x} + \mathbf{x}), \quad \mathbf{x}_- = \frac{1}{2}(A\mathbf{x} - \mathbf{x}),$$

显然 $\mathbf{x}_\pm \in V_\pm$, 且 $\mathbf{x} = \mathbf{x}_+ + \mathbf{x}_-$, 得到所需要的分解.

(2) 类似, 考虑 $V_0 = \ker(A)$, $V_1 = \ker(A - I)$. □

例 9.1.11. 设 A, B 分别是 $l \times n, m \times n$ 矩阵. 证明: $\mathcal{N}(A) \subseteq \mathcal{N}(B)$ 当且仅当存在 $m \times l$ 矩阵 C, 使得 $B = CA$.

证明: "$\Leftarrow$" 是直接的, 若 $B = CA$, 则由 $A\mathbf{x} = \mathbf{0}$ 可知 $CA\mathbf{x} = \mathbf{0}$, 从而有 $\mathcal{N}(A) \subseteq \mathcal{N}(B)$.

"$\Rightarrow$" 设 A, B 给出的线性映射分别为 $A : \mathbf{R}^n \to \mathbf{R}^l$ 与 $B : \mathbf{R}^n \to \mathbf{R}^m$. 取 $\ker(A)$ 的基 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\}$, 把它扩充为 $\mathbf{R}^n$ 的基 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\} \cup \{\mathbf{w}_1, \cdots, \mathbf{w}_t\}$, 则 $\{A(\mathbf{w}_1), \cdots, A(\mathbf{w}_t)\}$ 为 $\mathrm{Im}(A)$ 的基, 将后者扩充为 $\mathbf{R}^l$ 的基底 $\{A(\mathbf{w}_1), \cdots, A(\mathbf{w}_t)\} \cup \{\mathbf{z}_1, \cdots, \mathbf{z}_p\}$

定义线性映射 $C : \mathbf{R}^l \to \mathbf{R}^m$ 为

$$C(A(\mathbf{w}_i)) = B(\mathbf{w}_i), \quad \forall 1 \le i \le t; \quad C(\mathbf{z}_j) \text{ 任意取值}, \quad \forall j \le p.$$

结合 $\ker(A) \subseteq \ker(B)$, 可验证 $B = C \circ A$, 在表示矩阵层面即有 $B = CA$. □

例 9.1.12. 设 A, B 都是 $m \times n$ 矩阵. 证明: $\mathcal{N}(A) = \mathcal{N}(B)$ 当且仅当存在 m 阶可逆矩阵 T, 使得 $B = TA$.

证明: "$\Rightarrow$" 设 $\mathcal{N}(A) = \mathcal{N}(B)$, 利用上例证明中的步骤: 取 $\ker(A)$ 的基 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\}$, 把它扩充为 $\mathbf{R}^n$ 的基 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\} \cup \{\mathbf{w}_1, \cdots, \mathbf{w}_t\}$, 则 $\{A(\mathbf{w}_1), \cdots, A(\mathbf{w}_t)\}$ 为 $\mathrm{Im}(A)$ 的基, 将后者扩充为 $\mathbf{R}^m$ 的基底 $\{A(\mathbf{w}_1), \cdots, A(\mathbf{w}_t)\} \cup \{\mathbf{z}_1, \cdots, \mathbf{z}_p\}$. 由于 $\ker(A) = \ker(B)$, 则 $\{B(\mathbf{w}_1), \cdots, B(\mathbf{w}_t)\}$ 为 $\mathrm{Im}(B)$ 的基, 将后者扩充为 $\mathbf{R}^m$ 的基底 $\{B(\mathbf{w}_1), \cdots, B(\mathbf{w}_t)\} \cup \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

定义线性映射 $T : \mathbf{R}^m \to \mathbf{R}^m$ 为

$$T(A(\mathbf{w}_i)) = B(\mathbf{w}_i), \quad \forall 1 \le i \le t; \quad T(\mathbf{z}_j) = \mathbf{u}_j, \quad \forall j \le p.$$

则 T 为线性同构, 且有 $B = T \circ A$.

□

例 9.1.13. 对分块上三角矩阵 $C = \begin{pmatrix} A & X \\ O & B \end{pmatrix}$, 证明: $\mathrm{rank}(C) \ge \mathrm{rank}(A) + \mathrm{rank}(B)$.

证明: 设 $rk(A) = r$, $rk(B) = s$. 取 A 的行向量极大无关组 $\alpha_{i_1}, \cdots, \alpha_{i_r}$, 取 B 的行向量极大无关组 $\beta_{j_1}, \cdots, \beta_{j_s}$, 它们给出 C 的 $r + s$ 个行向量, 分别为

$$(\alpha_{i_1}|X_{i_1}), \cdots, (\alpha_{i_r}|X_{i_r}), \quad (0|\beta_{j_1}), \cdots, (0|\beta_{j_s}),$$

它们显然线性无关, 从而有 $rk(C) \ge r + s = rk(A) + rk(B)$.

□

例 9.1.14. 设 $A_{m\times n}$ 是秩为 $r>0$ 的矩阵. 令 C 为 A 的主列按顺序组成的矩阵, 令 R 为 A 的行简化阶梯形矩阵的非零行按顺序组成的矩阵. 证明: $A=CR$.

证明: 设 A 经过初等行变换 T 化为简化阶梯形矩阵 $A_0=TA$, 令 C_0 为 A_0 的主列按顺序组成的矩阵. 注意到, 对于 A_0 的某一主列, 不妨设为第 j 列, 有

$$(A_0)_{ij}=\sum_{k=1}^{m}T_{ik}A_{kj},$$

即 A_0 的第 j 列等于 T 乘以 A 的第 j 列 (这其实是显然的, 当对 A 做行变换 T 时, 对 A 的每一列都做一样的行变换 T). 由此可知 $C_0=TC$. 这样, 只要对 A_0 证明 $A_0=C_0R$ 即可, 因为由此可得

$$A=T^{-1}A_0=T^{-1}(C_0R)=(T^{-1}C_0)R=CR.$$

为了方便起见, 假设 A 是行简化阶梯形矩阵, 设其主列的列指标分别为 $j(1)<\cdots<j(r)$, 则 C 是 $m\times r$ 矩阵, R 是 $r\times n$ 矩阵, 且有

$$C_{it}=A_{ij(t)},\quad R_{tj}=A_{tj},\quad \forall i\le m,t\le r,j\le n.$$

由此可得

$$(CR)_{ij}=\sum_{t=1}^{r}C_{it}R_{tj}=\sum_{t=1}^{r}A_{ij(t)}A_{tj}.$$

注意到 $A_{ij(t)}$ 是 A 的 $j(t)$ 主列的矩阵元, 只有当 $i=t$ 时 $A_{ij(t)}$ 非零, 值为 $A_{tj(t)}=1$, 结合上式可得当 $i\le r$ 时有

$$(CR)_{ij}=\sum_{t=1}^{r}A_{ij(t)}A_{tj}=A_{ij(i)}A_{ij}=A_{ij};$$

而当 $i>r$ 时, i 不等于任何 $t\le r$, 可得此时

$$(CR)_{ij}=\sum_{t=1}^{r}A_{ij(t)}A_{tj}=0=A_{ij}.$$

这就验证了 $CR=A$. □

9.2 行列式

例 9.2.1. 计算行列式

$$\begin{vmatrix}1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n\\ 1+x_2y_1 & 1+x_2y_2 & \cdots & 1+x_2y_n\\ \vdots & \vdots & \ddots & \vdots\\ 1+x_ny_1 & 1+x_ny_2 & \cdots & 1+x_ny_n\end{vmatrix}.$$

证明: 只考虑 $n \geq 3$ 的情形. 记所求的行列式为 $\det(A)$, 做行倍加变换, 将第 $i \geq 2$ 行减去第 1 行, 可得

$$\begin{aligned}\det(A) &= \begin{vmatrix} 1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n \\ (x_2-x_1)y_1 & (x_2-x_1)y_2 & \cdots & (x_2-x_1)y_n \\ \vdots & \vdots & \ddots & \vdots \\ (x_n-x_1)y_1 & (x_n-x_1)y_2 & \cdots & (x_n-x_1)y_n \end{vmatrix} \\ &= (x_2-x_1)\cdots(x_n-x_1)\begin{vmatrix} 1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n \\ y_1 & y_2 & \cdots & y_n \\ \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & \cdots & y_n \end{vmatrix} \\ &= 0.\end{aligned}$$

□

例 9.2.2. 已知整数 $1653, 2581, 3451, 4582$ 可以被 29 整除. 证明下面的四阶行列式的值也能被 29 整除.

$$A = \begin{bmatrix} 1 & 6 & 5 & 3 \\ 2 & 5 & 8 & 1 \\ 3 & 4 & 5 & 1 \\ 4 & 5 & 8 & 2 \end{bmatrix},$$

证明: 对矩阵做三次列倍加变换, 对每个 $1 \leq i \leq 3$, 分别将第 4 列加上第 i 列的 10^{4-i} 倍, 矩阵的行列式不变, 可得

$$\det(A) = \begin{vmatrix} 1 & 6 & 5 & 1653 \\ 2 & 5 & 8 & 2581 \\ 3 & 4 & 5 & 3451 \\ 4 & 5 & 8 & 4582 \end{vmatrix}$$

再对第四列做 Laplace 展开, 结合条件即得 $\det(A)$ 能被 29 整除.

□

例 9.2.3 (Hadamard 不等式)**.** 对任意 n 阶矩阵 $T = [\mathbf{t}_1, \cdots, \mathbf{t}_n]$, 有

$$|\det(T)| \leq \prod_{i=1}^{n} ||\mathbf{t}_i||.$$

证明: 不妨设 T 可逆, 则它有 QR 分解 $T = QR$, 其中 Q 是正交矩阵, R 是上三角矩阵, 且 $R_{ij} = \langle \mathbf{q}_i, \mathbf{t}_j \rangle$. 由此可得

$$\det(T) = \det(Q)\det(R) = \pm\det(R) = \pm\prod_{i=1}^{n} R_{ii} = \pm\prod_{i=1}^{n}\langle \mathbf{q}_i, \mathbf{t}_i \rangle.$$

结合 Cauchy-Schwartz 不等式

$$\big|\langle \mathbf{q}_i, \mathbf{t}_i \rangle\big| \le ||\mathbf{q}_i|| \cdot ||\mathbf{t}_i|| = ||\mathbf{t}_i||,$$

即得证 Hadamard 不等式. □

例 9.2.4. 设 n 阶方阵 A 不可逆, 则其伴随矩阵的秩为 0 或 1.

证明: 由 A 不可逆知 $\det(A) = 0$, 从而有 $A \cdot \mathrm{adj}(A) = \det(A) I_n = O$, 即 $\mathcal{R}(\mathrm{adj}(A)) \subseteq \mathcal{N}(A)$. 利用维数公式可得

$$\mathrm{rank}(\mathrm{adj}(A)) = \dim(\mathcal{R}(\mathrm{adj}(A))) \le \dim(\mathcal{N}(A)) = n - \mathrm{rank}(A).$$

当 $\mathrm{rank}(A) = n-1$ 时, 上式表明 $\mathrm{rank}(\mathrm{adj}(A)) \le 1$; 当 $\mathrm{rank}(A) \le n-2$ 时, A 中任何 $(n-1)$ 个行向量线性相关. 由此可知对任何指标 i, j, 删去 A 的第 i 行第 j 列所得矩阵的 $(n-1)$ 个行向量线性相关, 因而 A 余子式 M_{ij} 等于零. 故此时 $\mathrm{adj}(A) = O$, 其秩为零. □

例 9.2.5. 用 A^* 表示 $n(n \ge 3)$ 阶方阵 A 的伴随矩阵. 证明:

$$(A^*)^* = (\det(A))^{n-2} A.$$

证明: 如果 A 可逆, 则由等式 $AA^* = \det(A) I_n$ 可知 $A^* = \det(A) \cdot A^{-1}$, 进而可得 $\det(A^*) = (\det(A))^n \det(A^{-1}) = (\det(A))^{n-1}$. 再结合 $A^*(A^*)^* = \det(A^*) I_n$, 可得

$$(A^*)^* = \det(A^*)(A^*)^{-1} = \det(A^*) \cdot \left(\frac{1}{\det(A)} A\right) = (\det(A))^{n-2} A.$$

如果 A 不可逆, 则由上例的结论, A^* 的秩为 0 或 1. 由 $n \ge 3$, 则 A^* 中任何 $(n-1)$ 行都是线性相关的, 由此可知 A^* 的任何余子式都等于零, 从而可知 $(A^*)^*$ 是零矩阵, 题述结论显然成立. □

例 9.2.6. 设 n 阶方阵 A 满足:

$$a_{ii} > \sum_{j \ne i} |a_{ij}|, \quad \forall 1 \le i \le n.$$

证明: $\det(A) > 0$.

证明: 对 n 进行归纳. 当 $n=1$ 时, 条件为 $a_{11}>0$, 显然 $\det(A)=a_{11}$ 大于零. 设 $(n-1)$ 版本命题成立, 考虑 $n\geq 2$ 版本的命题. 对每个 $1\leq j\leq n-1$, 将 A 的第 j 列减去 A 的第 n 列的 $\frac{a_{nj}}{a_{nn}}$ 倍, 则经过这 $(n-1)$ 次列倍加变换后 A 变成如下分块矩阵

$$\begin{pmatrix} B & * \\ O & a_{nn} \end{pmatrix}$$

其中 $B=\left(a_{ij}-a_{in}\frac{a_{nj}}{a_{nn}}\right)_{1\leq i,j\leq n-1}$. 由于倍加变换不改变行列式, 结合 Laplace 展开可得 $\det(A)=\det(B)\cdot a_{nn}$. 注意到, 对每个 $i\leq n-1$, 有

$$\begin{aligned}
\sum_{j\leq n-1,j\neq i}|B_{ij}| &= \sum_{j\leq n-1,j\neq i}|a_{ij}-a_{in}\frac{a_{nj}}{a_{nn}}| \\
&\leq \sum_{j\leq n-1,j\neq i}\left(|a_{ij}|+|a_{in}|\frac{|a_{nj}|}{a_{nn}}\right) \\
&= \left(\sum_{j\neq n,j\neq i}|a_{ij}|\right)+\frac{|a_{in}|}{a_{nn}}\left(\sum_{j\neq n,j\neq i}|a_{nj}|\right) \\
&< (a_{ii}-|a_{in}|)+\frac{|a_{in}|}{a_{nn}}(a_{nn}-|a_{ni}|) \\
&= a_{ii}-\frac{|a_{in}|}{a_{nn}}|a_{ni}| \\
&\leq a_{ii}-\frac{a_{in}}{a_{nn}}a_{ni} \\
&= B_{ii},
\end{aligned}$$

利用归纳假设有 $\det(B)>0$. 这就证明了 $\det(A)>0$, 从而完成归纳证明.

事实上, 我们证明了题中所述的方阵是正定的, 因为其顺序主子式都大于零. □

例 9.2.7. 设 A,B 分别是 $m\times n$ 与 $n\times m$ 矩阵. 证明:

$$\lambda^n\cdot\det(\lambda I_m-AB)=\lambda^m\cdot\det(\lambda I_n-BA).$$

证明: 考虑如下分块矩阵

$$M=\begin{pmatrix} \lambda I_m & A \\ \lambda B & \lambda I_n \end{pmatrix}$$

的行列式.

注意到,

$$\begin{pmatrix} I_m & O \\ -B & I_n \end{pmatrix}\begin{pmatrix} \lambda I_m & A \\ \lambda B & \lambda I_n \end{pmatrix} = \begin{pmatrix} \lambda I_m & A \\ O & \lambda I_n - BA \end{pmatrix},$$
$$\begin{pmatrix} \lambda I_m & A \\ \lambda B & \lambda I_n \end{pmatrix}\begin{pmatrix} I_m & O \\ -B & I_n \end{pmatrix} = \begin{pmatrix} \lambda I_m - AB & A \\ O & \lambda I_n \end{pmatrix},$$

由前一个式子可得 $\det(M) = \lambda^m \cdot \det(\lambda I_n - BA)$; 由后一个式子可得 $\det(M) = \lambda^n \cdot \det(\lambda I_m - AB)$. 结合起来即得所要证明的等式. □

由此例的结论可得如下两个推论.

(1) 考虑 $\lambda = 1$ 处的值, 可得

$$I_m - AB\text{可逆} \iff \left(\lambda^n \cdot \det(\lambda I_m - AB)\right)\Big|_{\lambda=1} \neq 0$$
$$\iff \left(\lambda^m \cdot \det(\lambda I_n - BA)\right)\Big|_{\lambda=1} \neq 0 \iff I_n - BA\text{可逆}.$$

(2) 对于 n 阶方阵 A, B, 有

$$p_{AB}(\lambda) = p_{BA}(\lambda).$$

例 9.2.8. 设 n 阶方阵 A, B 满足: $A, B, A+B$ 都可逆. 证明: $A^{-1} + B^{-1}$ 可逆.

证明: 由于 A 是可逆矩阵, 则

$$A^{-1} + B^{-1}\text{可逆} \iff A(A^{-1} + B^{-1})\text{可逆} \iff I_n + AB^{-1}\text{可逆}.$$

由于 B 是可逆矩阵, 则

$$A + B\text{可逆} \iff B^{-1}(A + B)\text{可逆} \iff I_n + B^{-1}A\text{可逆}.$$

而由前一例题结论的推论, 有

$$I_n + AB^{-1}\text{可逆} \iff I_n + B^{-1}A\text{可逆},$$

这就完成了本题的证明. □

9.3 二次型

例 9.3.1. 设 A 是 $m \times n$ 实矩阵, I 是单位矩阵. 证明: 当 $\lambda > 0$ 时, 矩阵 $B = \lambda I + A^T A$ 正定.

证明: 对任何非零向量 $\mathbf{x} \in \mathbf{R}^n$, 有

$$\langle \mathbf{x}, B\mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, A^TA\mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle + \langle A\mathbf{x}, A\mathbf{x}\rangle = \lambda||\mathbf{x}||^2 + ||A\mathbf{x}||^2 > 0.$$

这表明 B 是正定矩阵. □

例 9.3.2. 判断如下实对称矩阵是否正定?

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

解. 设 $A = C^TBC$, 其中 C 是可逆矩阵, 则 A 正定当且仅当 B 正定. 这是由于, A 给出的二次函数为

$$Q(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x}\rangle = \langle \mathbf{x}, C^TBC\mathbf{x}\rangle = \langle C\mathbf{x}, BC\mathbf{x}\rangle,$$

可换元 $\mathbf{y} = C\mathbf{x}$, 则有 $Q(\mathbf{x}) = \langle \mathbf{y}, B\mathbf{y}\rangle$, 可表示为由 B 给出的二次型. 注意到, 由于 C 可逆, 当 $\mathbf{x}$ 取遍非零向量时 $\mathbf{y}$ 也取遍非零向量, 反之亦然. 结合正定矩阵的定义, 即可知 A 正定当且仅当 B 正定. □

例 9.3.3. 设 A 是 m 阶正定矩阵, B 是 $m \times n$ 的实矩阵. 证明: B^TAB 为正定矩阵的充分必要条件是 $\mathrm{rank}(B) = n$.

证明: 由正定矩阵的定义,

$$\begin{aligned} B^TAB\text{正定} &\iff \forall \mathbf{x} \neq \mathbf{0} \in \mathbf{R}^n \ \text{有}\langle \mathbf{x}, B^TAB\mathbf{x}\rangle > 0 \\ &\iff \forall \mathbf{x} \neq \mathbf{0} \in \mathbf{R}^n \ \text{有}\langle B\mathbf{x}, AB\mathbf{x}\rangle > 0 \\ &\iff \forall \mathbf{x} \neq \mathbf{0} \in \mathbf{R}^n \ \text{有}B\mathbf{x} \neq \mathbf{0}(\text{这一断言推上一断言用到}A\text{正定}) \\ &\iff B\text{确定的线性映射是单射} \\ &\iff B\text{是列满秩} \\ &\iff \mathrm{rank}(B) = n. \end{aligned}$$

□

例 9.3.4. 设 n 阶矩阵 A 的矩阵元为 $a_{ij} = \frac{1}{i+j}$. 证明: A 是正定矩阵.

证明: 将二次型 $Q = Q(x_1, \cdots, x_n) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}x_ix_j$ 中含有 x_n 的项完全配方, 可得

$$Q = a_{nn}\left(\frac{a_{1n}}{a_{nn}}x_1 + \cdots + \frac{a_{(n-1)n}}{a_{nn}}x_{n-1} + x_n\right)^2 + \sum_{i\leq n-1}\sum_{j\leq n-1}\left(a_{ij} - \frac{a_{in}a_{jn}}{a_{nn}}\right)x_ix_j.$$

将上式中后一部分记为

$$R(x_1,\cdots,x_{n-1})=\sum_{i\le n-1}\sum_{j\le n-1}\left(a_{ij}-\frac{a_{in}a_{jn}}{a_{nn}}\right)x_ix_j,$$

它是关于 $x_1,\cdots,x_{n-1}$ 的二次型.

将上述计算用于题中的矩阵, 记

$$Q_n(x_1,\cdots,x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{i+j}x_ix_j,$$

把含有 x_n 的项完全配方, 可得剩余的关于 $x_1,\cdots,x_{n-1}$ 的二次型为:

$$\begin{aligned}R(x_1,\cdots,x_{n-1})&=\sum_{i\le n-1}\sum_{j\le n-1}\left(\frac{1}{i+j}-\frac{\frac{1}{i+n}\cdot\frac{1}{j+n}}{1/2n}\right)x_ix_j\\&=\sum_{i\le n-1}\sum_{j\le n-1}\frac{1}{i+j}\left(\frac{n-i}{n+i}x_i\right)\left(\frac{n-j}{n+j}x_j\right)\\&=Q_{n-1}\left(\frac{n-1}{n+1}x_1,\frac{n-2}{n+2}x_2,\cdots,\frac{n-(n-1)}{n+(n-1)}x_{n-1}\right).\end{aligned}$$

由归纳假设, $R(x_1,\cdots,x_{n-1})\ge 0$, 且等号成立当且仅当 $x_1=\cdots=x_{n-1}=0$. 由此可得

$$Q_n(x_1,\cdots,x_n)=a_{nn}\left(\frac{a_{1n}}{a_{nn}}x_1+\cdots+\frac{a_{(n-1)n}}{a_{nn}}x_{n-1}+x_n\right)^2+R(x_1,\cdots,x_{n-1})\ge 0,$$

且等号成立当且仅当

$$\frac{a_{1n}}{a_{nn}}x_1+\cdots+\frac{a_{(n-1)n}}{a_{nn}}x_{n-1}+x_n=0\ \text{且}x_1=\cdots=x_{n-1}=0,$$

这就证明了 Q_n 是正定的二次型. □

例 9.3.5. 设 $S=\begin{pmatrix}A&B\\B^T&C\end{pmatrix}$ 是实对称矩阵. 证明: S 正定当且仅当 A 及其 Schur 补都正定.

证明: 设 A 是 n 阶方阵, C 是 m 阶方阵. 若 S 正定, 则 S 的第 n 个顺序主子式大于零, 即 $\det(A)>0$. 这样, 题目断言的条件与结论中都假设了 A 可逆, 因而可讨论 A 的 Schur 补 $D=C-B^TA^{-1}B$. 注意到

$$\begin{pmatrix}I&O\\-B^TA^{-1}&I\end{pmatrix}\begin{pmatrix}A&B\\B^T&C\end{pmatrix}=\begin{pmatrix}A&B\\O&D\end{pmatrix}.$$

上式左边的矩阵乘法表示对 S 作行倍加变换, 由于行倍加变换不改变行列式, 可知 S 的顺序主子式为

$$\det(S_k)=\begin{cases}\det(A_k), & \text{如果} k\le n;\\ \det(A)\cdot\det(D_{k-n}), & \text{如果} k>n;\end{cases}$$

结合方阵正定的充分必要条件是其顺序主子式都大于零, 可得 S 正定当且仅当 A,D 都正定. □

例 9.3.6. 判断平面上的二次曲线 $C: 3x_1^2+4x_1x_2+6x_2^2=1$ 是哪一种二次曲线.

解. 曲线 C 的定义方程可以写成

$$1=\begin{bmatrix} x_1 & x_2\end{bmatrix}\begin{bmatrix}3&2\\2&6\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

. 记上式右边二次型对应的实对称矩阵为 A, 其特征多项式为

$$p_A(\lambda)=(\lambda-3)(\lambda-6)-(-2)(-2)=(\lambda-2)(\lambda-7),$$

特征值为 $\lambda=2$ 或 $\lambda=7$. 通过求解方程组 $(A-\lambda I_2)\mathbf{x}=\mathbf{0}$ 可找出 A 的特征向量: $\frac{1}{\sqrt5}(-2,1)$ 是 $\lambda=2$ 的特征向量; $\frac{1}{\sqrt5}(1,2)$ 是 $\lambda=7$ 的特征向量. 令 $Q=\frac{1}{\sqrt5}\begin{pmatrix}-2&1\\1&2\end{pmatrix}$, 则有 $Q^TAQ=\operatorname{diag}\{2,7\}$. 换元

$$\begin{pmatrix}y_1\\y_2\end{pmatrix}=Q^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix}=Q^T\begin{pmatrix}x_1\\x_2\end{pmatrix},$$

则 C 的定义方程为

$$\begin{aligned}1&=\begin{bmatrix}x_1&x_2\end{bmatrix}A\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}x_1&x_2\end{bmatrix}Q\begin{bmatrix}2&\\&7\end{bmatrix}Q^T\begin{bmatrix}x_1\\x_2\end{bmatrix}\\&=\begin{bmatrix}y_1&y_2\end{bmatrix}\begin{bmatrix}2&\\&7\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}.\end{aligned}$$

注意到, Q 是正交矩阵, 从 (x_1,x_2) 平面到 (y_1,y_2) 平面的变换是等距同构, 保持一切几何量. 在 (y_1,y_2) 平面中, C 的定义方程是 $2y_1^2+7y_2^2=1$, 它是椭圆, 并可进一步确定 C 的所有几何量, 比如长短轴长度. 另外, 还可以求出 C 的焦点, 长短轴的直线方程, 再用 $\begin{pmatrix}x_1\\x_2\end{pmatrix}=Q\begin{pmatrix}y_1\\y_2\end{pmatrix}$ 变换到 (x_1,x_2) 平面中去. □

例 9.3.7. 设 A,B 是 n 阶实对称矩阵, 且 A 正定. 证明: 存在可逆矩阵 P, 使得 P^TAP 和 P^TBP 同时为对角矩阵.

证明: 取正交矩阵 Q 使得 $Q^TAQ=D$ 是对角矩阵. 设 $D=\mathrm{diag}\{d_1,\cdots,d_n\}$, 由 A 正定可知每个 d_i 都是正数. 定义

$$\sqrt{D}=\mathrm{diag}\{\sqrt{d_1},\cdots,\sqrt{d_n}\},\quad \sqrt{\frac{1}{D}}=\mathrm{diag}\{\sqrt{\frac{1}{d_1}},\cdots,\sqrt{\frac{1}{d_n}}\}.$$

考虑实对称矩阵

$$\widetilde{B}=\sqrt{\frac{1}{D}}Q^TBQ\sqrt{\frac{1}{D}},$$

它可相似于对角矩阵. 设正交矩阵 $\widetilde{R}$ 使得 $\widetilde{R}^T\widetilde{B}\widetilde{R}$ 是对角矩阵, 令

$$R=\sqrt{\frac{1}{D}}\widetilde{R}.$$

令 $P=QR$, 则有

$$P^TAP=R^TQ^TAQR=R^TDR=\widetilde{R}^T\sqrt{\frac{1}{D}}D\sqrt{\frac{1}{D}}\widetilde{R}=\widetilde{R}^T\widetilde{R}=I_n,$$

且有

$$P^TBP=R^TQ^TBQR=\widetilde{R}^T\sqrt{\frac{1}{D}}Q^TBQ\sqrt{\frac{1}{D}}\widetilde{R}=\widetilde{R}^T\widetilde{B}\widetilde{R}.$$

这就完成了整个证明. □

例 9.3.8. 设 A,B 是 n 阶实对称矩阵, 且 $A+B$ 正定. 证明: 存在可逆矩阵 P, 使得 P^TAP 和 P^TBP 同时为对角矩阵.

证明: 由条件 $A+B$ 正定, 可对 $A+B$ 与 $-B$ 用前例的结论, 存在可逆矩阵 P 使得 $P^T(A+B)P$ 与 $P^T(-B)P$ 同时对角化, 此时

$$P^TAP=P^T(A+B)P+P^T(-B)P$$

也是对角矩阵. □

9.4 谱理论

例 9.4.1. 对方阵 A, 若多项式 $f(x)$ 满足 $f(A)=O$, 则称 $f(x)$ 是 A 的化零多项式. 证明: A 的特征值都是 A 的化零多项式的根.

证明: 设 $f(x)=a_mx^m+\cdots+a_1x+a_0$ 是 A 的化零多项式, λ 是 A 的特征值, $\mathbf{v}\neq\mathbf{0}$ 是对应的特征向量, 即 $A\mathbf{v}=\lambda\mathbf{v}$. 结合条件 $f(A)=O$, 可得

$$\mathbf{0}=f(A)\mathbf{v}=\left(a_mA^m+\cdots+a_1A+a_0I_n\right)\mathbf{v}=\left(a_m\lambda^m+\cdots+a_1\lambda+a_0\right)\mathbf{v},$$

故有 $f(\lambda)=0$. □

由此问题的结论, 可从 A 满足的等式推出 A 的特征值的信息. 例如:

(1) 如果存在正整数 k 使得 $A^k=O$, 则 A 的特征值只能是 0;

(2) 如果 $A^2=I_n$, 则 A 的特征值只能是 ± 1;

(3) 如果 $A^2=A$, 则 A 的特征值只能是 0 或 1.

例 9.4.2. 设 A 是 n 阶实方阵, 满足 $A^2=I_n$. 证明: A 在 $\mathbf{R}$ 上可对角化.

证明: 由前一题的结论, A 的特征值只能是 ± 1, 设对应的特征子空间分别为

$$V_+=\{\mathbf{v}|A\mathbf{v}=\mathbf{v}\},\quad V_-=\{\mathbf{v}|A\mathbf{v}=-\mathbf{v}\}.$$

对任何 $\mathbf{x}\in\mathbf{R}^n$, 令

$$\mathbf{x}_+=\frac{\mathbf{x}+A\mathbf{x}}{2},\quad \mathbf{x}_-=\frac{\mathbf{x}-A\mathbf{x}}{2},$$

则 $\mathbf{x}_\pm\in V_\pm$, 且有 $\mathbf{x}=\mathbf{x}_++\mathbf{x}_-$. 这就证明了 $\mathbf{R}^n=V_+\oplus V_-$, 由此可得 A 在 $\mathbf{R}$ 上可对角化. □

例 9.4.3. 任何迹为零的方阵相似于一个对角元素全为零的方阵.

证明: 设 $\mathrm{trace}(A)=0$. 如果 A 是数量矩阵, 则 $A=O$, 命题显然成立.

以下假设 A 不是数量矩阵. 我们断言在此假设下, 存在非零向量 $\mathbf{v}$ 使得 $\mathbf{v},A\mathbf{v}$ 线性无关. 否则的话, 对每个非零向量 $\mathbf{v}$, 存在实数 $c(\mathbf{v})$ 使得 $A\mathbf{v}=c(\mathbf{v})\mathbf{v}$. 由此可得

$$c(\mathbf{e}_1+\cdots+\mathbf{e}_n)(\mathbf{e}_1+\cdots+\mathbf{e}_n)=A\sum_{i=1}^n\mathbf{e}_i=c(\mathbf{e}_1)\mathbf{e}_1+\cdots+c(\mathbf{e}_n)\mathbf{e}_n,$$

则有 $c(\mathbf{e}_1)=\cdots=c(\mathbf{e}_n)$, 从而 A 是数量矩阵, 矛盾!

利用上述断言, 取非零向量 $\mathbf{v}$, 使得 $\mathbf{v},A\mathbf{v}$ 线性无关. 令 $\mathbf{v}_1=\mathbf{v}$, $\mathbf{v}_2=A\mathbf{v}$, 则 $\mathbf{v}_1,\mathbf{v}_2$ 线性无关, 可将它们扩充为 $\mathbf{R}^n$ 的一个基 $\mathbf{v}_1,\ldots,\mathbf{v}_n$. 利用这个基, 可把 A 相似于如下形式

$$B=\begin{pmatrix}0 & *\\ * & C\end{pmatrix}.$$

由于相似变换保持迹, 可知 $\mathrm{trace}(C)=0$. 利用归纳假设, 存在 $(n-1)$ 阶矩阵 T 使得 $T^{-1}CT$ 的每个对角元都是零. 注意到, 可进一步把 B 做相似变换

$$\begin{pmatrix}1 & O\\ O & T^{-1}\end{pmatrix}\begin{pmatrix}0 & *\\ * & C\end{pmatrix}\begin{pmatrix}1 & O\\ O & T\end{pmatrix}=\begin{pmatrix}0 & *\\ * & T^{-1}CT\end{pmatrix},$$

最后所得的矩阵的对角元都是零. 这就完成了整个归纳证明. □

例 9.4.4. 设矩阵 $A=\begin{pmatrix}1&x&1\\x&1&y\\1&y&1\end{pmatrix}$, $B=\operatorname{diag}\{0,1,2\}$. 当 x,y 满足什么条件时, A 与 B 相似?

解. A 的特征多项式为

$$p_A(\lambda)=(\lambda-1)^3-2xy-(\lambda-1)(1+x^2+y^2).$$

若 A 相似于对角矩阵 B, 则 $p_A(\lambda)$ 的三根为 $0,1,2$, 由此可得

$$xy=0,\quad x^2+y^2=0,$$

即有 $x=y=0$.

当 $x=y=0$ 时, $A=\begin{pmatrix}1&0&1\\0&1&0\\1&0&1\end{pmatrix}$. 直接计算可得 A 的所有特征值为 $0,1,2$, 且相应的特征子空间为

$$V_0=\operatorname{span}\{(1,0,-1)\},\quad V_1=\operatorname{span}\{(0,1,0)\},\quad V_2=\operatorname{span}\{(1,0,1)\},$$

所以此时 A 相似于 B.

综合这两方面, A,B 相似的充分必要条件是 $x=y=0$. □

例 9.4.5. 证明如下的方阵 A 可以相似对角化.

$$A=\begin{bmatrix}0&0&0&\cdots&1\\1&0&0&\cdots&0\\0&1&0&\cdots&0\\\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&\cdots&1&0\end{bmatrix}.$$

证明: 直接计算 A 的特征多项式

$$\begin{aligned}p_A(\lambda)=\det(\lambda I_n-A)&=\begin{vmatrix}\lambda&0&0&\cdots&-1\\-1&\lambda&0&\cdots&0\\0&-1&\lambda&\cdots&0\\\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&\cdots&-1&\lambda\end{vmatrix}\\&=\lambda^n+\operatorname{sign}(n12\cdots(n-1))(-1)^n\\&=\lambda^n-1,\end{aligned}$$

其特征值都是单重根, 因而都半单. 由此可得 A 在 $\mathbf{C}$ 上可对角化. □

例 9.4.6. 确定分块矩阵 $A=\begin{pmatrix} I_m & B \\ O & -I_n \end{pmatrix}$ 的特征值的几何重数与代数重数.

解. A 的特征多项式为

$$p_A(\lambda)=(\lambda-1)^m(\lambda+1)^n,$$

共有两个特征值 ±1, 代数重数分别为 m 与 n.

特征值 1 的特征子空间为如下方程组的解空间:

$$(A-I_{m+n})\,\mathbf{v}=\mathbf{0}.$$

记 $\mathbf{v}=\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$, 则上述方程组可直接求解

$$\begin{pmatrix} O & B \\ O & -2I_n \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}=\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \iff \mathbf{y}=\mathbf{0}.$$

其解空间是 m 维的, 故特征值 1 的几何重数为 m.

特征值 -1 的特征子空间为如下方程组的解空间:

$$(A+I_{m+n})\,\mathbf{v}=\mathbf{0}.$$

记 $\mathbf{v}=\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$, 则上述方程组可直接求解

$$\begin{pmatrix} 2I_m & B \\ O & O \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}=\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \iff \mathbf{x}=-\frac{1}{2}B\mathbf{y}.$$

其解空间是 n 维的, 故特征值 -1 的几何重数为 n. □

例 9.4.7. 设 σ 为实线性空间 V 上的线性变换, 满足 $\sigma^2=-\mathrm{Id}_V$. 证明: V 的维数为偶数, 且存在 V 的一组基, 使得 σ 在这组基下的表示矩阵为 $\begin{pmatrix} O & I_m \\ -I_m & O \end{pmatrix}$.

证明: 设 σ 在标准基下的表示矩阵为 J, 考虑 J 的复特征值. 由例9.4.1 的结论, J 的特征值只能是 $\pm i$. 设 $\pm i$ 作为 J 的特征值的代数重数分别为 s,t, 则有 $p_J(\lambda)=(\lambda-i)^s(\lambda+i)^t$. 注意到, J 是实方阵, 则其特征多项式是实系数多项式, 因而有

$$(\lambda-i)^s(\lambda+i)^t=p_J(\lambda)=\overline{p_J(\lambda)}=(\lambda+i)^s(\lambda-i)^t,$$

由此可知 $s=t$, 从而 $n=s+t$ 是偶数, 且 $\pm i$ 作为 J 的特征值的代数重数都是 $\frac{n}{2}$.

记 $n=2m$, 令 J 的 $\pm i$ 特征子空间分别为

$$V_i=\{\mathbf{z}\in\mathbf{C}^n|J\mathbf{z}=i\mathbf{z}\},\quad V_{-i}=\{\mathbf{z}\in\mathbf{C}^n|J\mathbf{z}=-i\mathbf{z}\},$$

由于特征值的几何重数不超过代数重数, 可得 $\dim_{\mathbf{C}}(V_{\pm i})\le\frac{n}{2}$. 注意到, 对任何 $\mathbf{z}\in\mathbf{C}^n$, 它可分解为

$$\mathbf{z}=\mathbf{z}_++\mathbf{z}_-,\quad \mathbf{z}_+\in V_i,\quad \mathbf{z}_-\in V_{-i}.$$

事实上, 只要取 $\mathbf{z}_+=\frac{i\mathbf{z}+J\mathbf{z}}{2i}$, $\mathbf{z}_-=\frac{i\mathbf{z}-J\mathbf{z}}{2i}$ 即可. 这个分解表明 $V_i+V_{-i}=\mathbf{C}^n$, 结合 $\dim_{\mathbf{C}}(V_{\pm i})\le\frac{n}{2}$ 可得 $\dim_{\mathbf{C}}(V_{\pm i})=\frac{n}{2}=m$.

我们来研究 V_i 的具体结构. 将 V_i 中的元素 $\mathbf{z}$ 写成实部虚部相加的形式: $\mathbf{z}=\mathbf{x}+i\mathbf{y}$, 其中 $\mathbf{x},\mathbf{y}\in\mathbf{R}^n$. 这样, 有

$$\mathbf{x}+i\mathbf{y}\in V_i\iff J(\mathbf{x}+i\mathbf{y})=i(\mathbf{x}+i\mathbf{y})\iff J\mathbf{x}=-\mathbf{y},J\mathbf{y}=\mathbf{x}\iff J\mathbf{x}=-\mathbf{y},$$

上述最后一步用到了 $J^2=-I_n$. 这样, V_i 中的元素都形如 $\mathbf{x}-iJ\mathbf{x}$, 其中 $\mathbf{x}\in\mathbf{R}^n$. 取 V_i 作为 $\mathbf{C}$ 上线性空间的一组基 $\mathbf{z}_1,\ldots,\mathbf{z}_m$. 设对每个 $1\le k\le m$, 有 $\mathbf{z}_k=\mathbf{x}_k-iJ\mathbf{x}_k$, 记 $\mathbf{y}_k=J\mathbf{x}_k$. 我们断言: $\{\mathbf{x}_1,\ldots,\mathbf{x}_m,\mathbf{y}_1,\ldots,\mathbf{y}_m\}$ 是 $\mathbf{R}^n$ 的一组基. 为了证明此断言, 考虑 $\mathbf{R}$ 上线性空间之间的同构 $\Psi:\mathbf{R}^n\to V_i$, $\Psi(\mathbf{x})=\mathbf{x}-iJ\mathbf{x}$, 有

$$\Psi(\mathbf{x}_k)=\mathbf{z}_k,\quad \Psi(\mathbf{y}_k)=\mathbf{y}_k-iJ\mathbf{y}_k=J\mathbf{x}_k-iJ(J(\mathbf{x}_k))=i\mathbf{z}_k.$$

如果存在实数 $\alpha_1,\ldots,\alpha_m,\beta_1,\ldots,\beta_m$ 使得 $\sum\limits_{k=1}^m\alpha_k\mathbf{x}_k+\sum\limits_{k=1}^m\beta_k\mathbf{y}_k=\mathbf{0}$, 则有

$$\mathbf{0}=\Psi(\sum_{k=1}^m\alpha_k\mathbf{x}_k+\sum_{k=1}^m\beta_k\mathbf{y}_k)=\sum_{k=1}^m(\alpha_k+i\beta_k)\mathbf{z}_k.$$

由 $\mathbf{z}_1,\ldots,\mathbf{z}_m$ 是 V_i 作为 $\mathbf{C}$ 上线性空间的基, 可得对每个 k 都有 $\alpha_k+i\beta_k=0$, 即 $\alpha_1,\ldots,\alpha_m,\beta_1,\ldots,\beta_m$ 全为零, 这就证明了断言.

利用断言, 考虑 J 在 $\mathbf{R}^n$ 的基 $\{\mathbf{x}_1,\ldots,\mathbf{x}_m,\mathbf{y}_1,\ldots,\mathbf{y}_m\}$ 下的表示矩阵, 注意到

$$J\mathbf{x}_k=\mathbf{y}_k,\quad J\mathbf{y}_k=-\mathbf{x}_k,$$

可知 J 的表示矩阵为 $\begin{pmatrix}O&-I_m\\I_m&O\end{pmatrix}$. 若换成基 $\{\mathbf{x}_1,\ldots,\mathbf{x}_m,-\mathbf{y}_1,\ldots,-\mathbf{y}_m\}$, 则相应的 J 的表示矩阵为基 $\begin{pmatrix}O&I_m\\-I_m&O\end{pmatrix}$.

□

例 9.4.8. 设 n 阶实对称矩阵 A, B 都是正定的. 证明: AB 的所有特征值都是正数.

证明: 取正交矩阵 Q 使得 $Q^TAQ = C = \text{diag}\{c_1, \ldots, c_n\}$ 是对角矩阵, 令 $Q^TBQ = D$, 则 C 的对角元 c_i 都是正数, D 是正定的实对称矩阵. 注意到

$$CD = Q^TAQQ^TBQ = Q^{-1}(AB)Q,$$

这表明 CD 与 AB 相似, 故两者的特征值完全一样. 只要证明 CD 的特征值都是正数.

对 CD 的每个特征值 $\lambda \in \mathbf{C}$, 设 $\mathbf{x} \in \mathbf{C}^n$ 是其对应的特征向量, 即有 $CD\mathbf{x} = \lambda\mathbf{x}$. 由此可得 $D\mathbf{x} = \lambda C^{-1}\mathbf{x}$, 进而有

$$\langle \mathbf{x}, D\mathbf{x} \rangle = \langle \mathbf{x}, \lambda C^{-1}\mathbf{x} \rangle = \lambda \sum_{i=1}^{n} \frac{1}{c_i}|x_i|^2, \tag{9.1}$$

其中 $\langle \cdot, \cdot \rangle$ 表示厄米内积. 由于 D 是实对称矩阵, 有

$$\langle \mathbf{x}, D\mathbf{x} \rangle = \langle \overline{D}^T\mathbf{x}, \mathbf{x} \rangle = \langle D\mathbf{x}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, D\mathbf{x} \rangle},$$

这说明 $\langle \mathbf{x}, D\mathbf{x} \rangle$ 是实数, 代入(9.1)可知 λ 是实数. 这样可选取 CD 的属于 λ 的实特征向量 $\mathbf{x}$, 代入(9.1), 结合 D 是正定矩阵可得

$$0 < \langle \mathbf{x}, D\mathbf{x} \rangle = \langle \mathbf{x}, \lambda C^{-1}\mathbf{x} \rangle = \lambda \sum_{i=1}^{n} \frac{1}{c_i}|x_i|^2,$$

从而证明了 λ 是正数. □

例 9.4.9. 设 Q 是正交矩阵. 证明如下断言.

(1) 若 $\lambda \in \mathbf{C}$ 是 Q 的特征值, 则 $\frac{1}{\lambda}$ 也是 Q 的特征值, 且有 $|\lambda| = 1$.

(2) 如果 $\det(Q) = -1$, 则 -1 是 Q 的一个特征值.

(3) 如果 Q 的阶数是奇数, 且 $\det(Q) = 1$, 则 1 是 Q 的一个特征值.

(4) 如果 Q 还是对称矩阵, 则 U 的特征值为 -1 或 1.

证明: (1) 设 $\mathbf{v}$ 是属于特征值 λ 的特征向量, 有 $Q\mathbf{v} = \lambda\mathbf{v}$, 由此可知

$$|\lambda|^2\langle \mathbf{v}, \mathbf{v} \rangle = \langle Q\mathbf{v}, Q\mathbf{v} \rangle = \langle \mathbf{v}, Q^TQ\mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle,$$

故 $|\lambda| = 1$. 进一步, 对 $Q\mathbf{v} = \lambda\mathbf{v}$ 两边取复共轭, 有

$$Q\overline{\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}} = \frac{1}{\lambda}\overline{\mathbf{v}},$$

这表明 $\frac{1}{\lambda}$ 也是 Q 的特征值.(或者利用实方阵的特征值共轭成对出现这个事实)

(2) 由前一小问的结论, Q 的非实特征值共轭成对出现, 每对的乘积等于 1; Q 的实特征值只能是 ± 1. 注意到 $\det(Q)$ 等于 Q 的所有特征值的乘积, 由此可知若 $\det(Q)=-1$, 则 -1 是 Q 的一个特征值.

(3) Q 的非实特征值共轭成对出现, 每对的乘积等于 1. 这些非实特征值的代数重数的总和为偶数, 如果 Q 是奇数阶的, 且 1 不是特征值, 则 -1 的代数重数等于奇数, 由此得出 $\det(Q)=-1$, 矛盾!

(4) 如果 Q 对称, 则特征根是实数, 结合 (1) 可知只能是 ± 1. □

9.5 实对称矩阵的谱分解

例 9.5.1. 设 $A=\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$. 求 4 阶正交矩阵 Q 和对角矩阵 Λ 使得 $Q^TAQ=\Lambda$.

解. 即要找出实对称矩阵 A 的谱分解. 注意到 A 保持子空间 $\mathrm{span}\{\mathbf{e}_1,\mathbf{e}_3\}$ 与 $\mathrm{span}\{\mathbf{e}_2,\mathbf{e}_4\}$. 只需分别计算 A 在这两个子空间上的谱分解.

在 $\mathrm{span}\{\mathbf{e}_1,\mathbf{e}_3\}$ 上 A 的表示矩阵为 $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, 特征值为 $0,1$, 对应的特征向量分别为 $\mathbf{e}_1-\mathbf{e}_3,\mathbf{e}_1+\mathbf{e}_3$.

在 $\mathrm{span}\{\mathbf{e}_2,\mathbf{e}_4\}$ 上 A 的表示矩阵为 $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$, 特征值为 $-\frac{1}{3},1$, 对应的特征向量分别为 $\mathbf{e}_2-\mathbf{e}_4,\mathbf{e}_2+\mathbf{e}_4$.

结合起来, 如下四个特征向量

$$\frac{1}{\sqrt{2}}(\mathbf{e}_1-\mathbf{e}_3),\quad \frac{1}{\sqrt{2}}(\mathbf{e}_1+\mathbf{e}_3),\quad \frac{1}{\sqrt{2}}(\mathbf{e}_2-\mathbf{e}_4),\quad \frac{1}{\sqrt{2}}(\mathbf{e}_2+\mathbf{e}_4)$$

构成 $\mathbf{R}^4$ 的单位正交基, 可用它们给出 A 的谱分解. 令

$$Q=\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\quad \Lambda=\begin{bmatrix} 0 & & & \\ & 1 & & \\ & & -\frac{1}{3} & \\ & & & 1 \end{bmatrix},$$

即有 $Q^TAQ=\Lambda$. □

例 9.5.2 (极分解)**.** 对 n 阶方阵 A, 存在正交矩阵 Q 和对称半正定矩阵 S, 使得 $A=QS$.

证明: 取 A 的奇异值分解 $A=U\Sigma V^T$, 其中 U,V 是 n 阶正交矩阵, $\Sigma=\mathrm{diag}\{\sigma_1,\cdots,\sigma_n\}$ 是对角元非负的对角矩阵. 由此得到分解

$$A=(UV^T)(V\Sigma V^T).$$

令 $Q=UV^T$, $S=V\Sigma V^T$, 则 Q 是正交矩阵, S 是对称半正定矩阵. 这就得到了 A 的极分解.

还有另一个顺序的极分解. 取 A^T 的极分解 $A^T=QS$, 则有 $A=SQ^T$, 将 A 表示成了对称半正定矩阵左乘正交矩阵的乘积.

□

例 9.5.3. 设 $A\in M_{m\times n}(\mathbf{R})$, 定义 A 的算子范数为

$$||A||=\max_{\mathbf{v}\in\mathbf{R}^n,||\mathbf{v}||=1}||A\mathbf{v}||.$$

证明:

$$||A||=\max_{||\mathbf{u}||=||\mathbf{v}||=1}\mathbf{u}^TA\mathbf{v},$$

上式中 $\mathbf{u}\in\mathbf{R}^m,\mathbf{v}\in\mathbf{R}^n$.

证明: 对 $\mathbf{R}^m$ 中给定的向量 $\mathbf{y}$, 对任何单位向量 $\mathbf{u}\in\mathbf{R}^m$, 由 Cauchy-Schwartz 不等式有

$$\mathbf{u}^T\mathbf{y}\le||\mathbf{u}||\cdot||\mathbf{y}||=||\mathbf{y}||.$$

当 $\mathbf{y}=\mathbf{0}$ 时, 任取 $\mathbf{u}$; 当 $\mathbf{y}\neq\mathbf{0}$ 时, 取 $\mathbf{u}=\frac{\mathbf{y}}{||\mathbf{y}||}$, 可使上述不等式成立等号. 这样就有

$$\max_{||\mathbf{u}||=1}\mathbf{u}^T\mathbf{y}=||\mathbf{y}||.$$

由此可得

$$\max_{||\mathbf{u}||=||\mathbf{v}||=1}\mathbf{u}^TA\mathbf{v}=\max_{||\mathbf{v}||=1}\left(\max_{||\mathbf{u}||=1}\mathbf{u}^TA\mathbf{v}\right)=\max_{||\mathbf{v}||=1}||A\mathbf{v}||=||A||.$$

□

例 9.5.4. 设 A 是 n 阶实方阵. 证明: A 的实特征值的绝对值不超过 A 的最大的奇异值.

证明: 设 λ 是 A 的实特征值, $\mathbf{v}$ 是对应的实特征向量, 则有

$$||A||=\max_{\mathbf{x}\neq\mathbf{0}\in\mathbf{R}^n}\frac{||A\mathbf{x}||}{||\mathbf{x}||}\ge\frac{||A\mathbf{v}||}{||\mathbf{v}||}=\frac{||\lambda\mathbf{v}||}{||\mathbf{v}||}=|\lambda|.$$

我们已经证明过 A 的谱范数 $||A||$ 等于 A 的最大奇异值, 结合上式即有 $|\lambda|$ 不超过 A 的最大奇异值. □

例 9.5.5 (Weyl 不等式)**.** 设 A 是 n 阶实方阵, 它的所有奇异值为 $\sigma_1 \geq \cdots \geq \sigma_n$, 它的所有复特征值为 $\lambda_1, \cdots, \lambda_n$, 且有 $|\lambda_1| \geq \cdots \geq |\lambda_n|$. 证明: 对每个正整数 m, 有

$$|\lambda_1 \cdots \lambda_m| \leq \sigma_1 \cdots \sigma_m.$$

例 9.5.6. 设 $m \times n$ 实矩阵 A 的奇异值分解为 $A = U\Sigma V^T$. 求矩阵 $B = \begin{pmatrix} O & A^T \\ A & O \end{pmatrix}$ 的谱分解.

解. 由 $A = U\Sigma V^T$, 有 $U^TAV = \Sigma$. 注意到

$$\begin{pmatrix} V^T & O \\ O & U^T \end{pmatrix}\begin{pmatrix} O & A^T \\ A & O \end{pmatrix}\begin{pmatrix} V & O \\ O & U \end{pmatrix} = \begin{pmatrix} O & V^TA^TU \\ U^TAV & O \end{pmatrix} = \begin{pmatrix} O & \Sigma^T \\ \Sigma & O \end{pmatrix},$$

即 B 正交相似于 $C = \begin{pmatrix} O & \Sigma^T \\ \Sigma & O \end{pmatrix}$, 只需进一步确定 C 的谱分解. 对于 C, 可选取 $\mathbf{R}^{n+m}$ 的基底 $\mathbf{v}_1, \cdots, \mathbf{v}_n, \mathbf{u}_1, \cdots, \mathbf{u}_m$ 使得

$$C\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i & \forall 1 \leq i \leq r; \\ \mathbf{0} & \forall i > r; \end{cases} \qquad C\mathbf{u}_i = \begin{cases} \sigma_i \mathbf{v}_i & \forall 1 \leq i \leq r; \\ \mathbf{0} & \forall i > r; \end{cases}$$

由此可得, 在 $\mathbf{R}^{n+m}$ 的基底

$$\mathbf{v}_1 + \mathbf{u}_1, \mathbf{v}_1 - \mathbf{u}_1, \cdots, \mathbf{v}_r + \mathbf{u}_r, \mathbf{v}_r - \mathbf{u}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_n, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$$

下, C 的表示矩阵为

$$\mathrm{diag}\{\sigma_1, -\sigma_1, \cdots, \sigma_r, -\sigma_r, 0, \cdots, 0\},$$

这就给出了 C 的谱分解.

由此题的结论, 可得特征多项式之间的等式: 令 $C = A^TA$, 则有

$$p_B(\lambda) = \lambda^{m-n} \cdot p_C(\lambda^2).$$

□

例 9.5.7 (Ky Fan 樊畿)**.** 设 n 阶实对称矩阵 A 的特征值为 $\lambda_1 \geq \cdots \geq \lambda_n$, 对应的特征向量为 $\mathbf{u}_1, \cdots, \mathbf{u}_n$, 则有

$$\max_{n\times m\text{矩阵满足}Q^TQ=I} \mathrm{trace}(Q^TAQ) = \sum_{i=1}^m \lambda_i,$$

且 $Q = [\mathbf{u}_1, \cdots, \mathbf{u}_m]$ 时取得最大值.

证明: 实对称矩阵 A 可正交相似于对角矩阵, 即存在正交矩阵 P 使得 $A = P^T \Lambda P$, 其中 $\Lambda = \mathrm{diag}\{\lambda_1, \cdots, \lambda_n\}$. 这样, 可对原来的最值问题换元处理: 令 $\widetilde{Q} = PQ$, 则 $Q, \widetilde{Q}$ 相互确定 (因为 $P^{-1}\widetilde{Q} = Q$), 且有

$$\widetilde{Q}^T \widetilde{Q} = I_m \iff Q^T P^T P Q = I_m \iff Q^T Q = I_m.$$

由此可得

$$\max_{n\times m\text{矩阵满足}Q^TQ=I} \mathrm{trace}(Q^T A Q) = \max_{n\times m\text{矩阵满足}\widetilde{Q}^T\widetilde{Q}=I} \mathrm{trace}(\widetilde{Q}^T \Lambda \widetilde{Q}),$$

转化为对 Λ 证明 Ky Fan 不等式.

以下假设 $A = \Lambda$. 将 Q 写成列向量分块形式 $Q = [\mathbf{q}_1, \cdots, \mathbf{q}_m]$, 令

$$f(Q) = \mathrm{trace}(Q^T A Q) = \mathbf{q}_1^T A \mathbf{q}_1 + \cdots + \mathbf{q}_m^T A \mathbf{q}_m,$$

视之为 m^2 个变元 $q_{ij}(1 \le i, j \le m)$ 的函数, 来求 f 在如下 $C_m^2 + m = \frac{m(m+1)}{2}$ 个约束条件

$$\langle \mathbf{q}_k, \mathbf{q}_l \rangle - \delta_{kl} = 0, \quad 1 \le k \le l \le m$$

下的最大值. 由最值定理, f 有最大值点 V, 这里 V 是 $n \times m$ 矩阵且满足 $V^T V = I_m$. 注意到, V 是 f 在前述 $\frac{m(m+1)}{2}$ 个约束条件下的条件极值点, 利用 Lagrange 乘子法, 存在实数 $\lambda_{kl}(1 \le k \le l \le m)$ 使得 V 的 nm 个矩阵元与这些 λ_{kl} 满足 Lagrange 辅助函数

$$F = f - \sum_{1\le k\le l\le m} \lambda_{kl} \left(\langle \mathbf{q}_k, \mathbf{q}_l \rangle - \delta_{kl} \right)$$

的临界点方程:

$$\frac{\partial F}{\partial q_{ij}} = 0, \quad \forall 1 \le i \le n, \forall 1 \le j \le m.$$

将上述方程化简, 即为

$$\lambda_i q_{ij} = \lambda_{jj} q_{ij} + \sum_{k<j} \lambda_{kj} q_{ik} + \sum_{k>j} \lambda_{jk} q_{ik}. \tag{9.2}$$

将 $\{\lambda_{kl}\}_{k\le l}$ 扩充为一个 m 阶对称矩阵 K, 即对任何 $k \le l$ 定义 $K_{kl} = K_{lk} = \lambda_{kl}$. 由此可将(9.2)改写成矩阵形式:

$$AQ = QK.$$

这样, Lagrange 乘子法断言 f 在约束条件下的最大值点 V 满足: 存在 m 阶对称矩阵 K, 使得 $AV = VK$, 即有 $V^T AV = K$. 由此可得

$$\max_{Q^TQ=I_m} f(Q) = f(V) = \mathrm{trace}(V^T A V) = \mathrm{trace}(K).$$

将 V 的列向量组 $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ 扩充为 $\mathbf{R}^n$ 的单位正交基 $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$. 记 $\widetilde{V}=[\mathbf{v}_1,\cdots,\mathbf{v}_n]$, 令 $B=\widetilde{V}^TA\widetilde{V}$, 则 B 是对称矩阵. 注意到 B 的矩阵元为

$$B_{ij}=\langle\mathbf{v}_i,A\mathbf{v}_j\rangle,$$

再有 $AV=VK$ 可知对 $1\le i\le m$ 有 $A\mathbf{v}_i\in\mathrm{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$, 结合这两点可得 B 是如下形式的分块矩阵

$$B=\begin{pmatrix} K & O \\ O & * \end{pmatrix}.$$

这样, K 的特征值都是 B 的特征值, 而 B 正交相似于 A, 则 K 的特征值是 $\lambda_1,\cdots,\lambda_n$ 的一部分, 所以有

$$\mathrm{trace}(K)=\sum(K\text{的特征值})\le\lambda_1+\cdots+\lambda_m,$$

从而完成了整个证明.

$\square$

例 9.5.8 (Weyl 不等式)**.** 设 A,B 是 n 阶实对称矩阵, $C=A+B$. 设 A,B,C 的所有特征值分别为 $a_1\le a_2\le\cdots\le a_n$, $b_1\le b_2\le\cdots\le b_n$, 与 $c_1\le c_2\le\cdots\le c_n$. 证明:

(1) 对任何 $k\le l$, 有 $c_k\le a_l+b_{k-l+n}$.

(2) 对任何 $i\ge j$, 有 $c_i\ge a_j+b_{i-j+1}$.

证明: 由实对称矩阵的谱分解定理, 存在 A 的 n 个特征向量 $\{\mathbf{x}_i|1\le i\le n\}$ 构成 $\mathbf{R}^n$ 的一组基, 且 $A\mathbf{x}_i=a_i\mathbf{x_i}$. 类似的, 存在 $\mathbf{R}^n$ 的基 $\{\mathbf{y}_i|1\le i\le n\}$ 与 $\{\mathbf{z}_i|1\le i\le n\}$, 分别满足对每个指标 i 有 $B\mathbf{y}_i=b_i\mathbf{y}_i$, $C\mathbf{z}_i=c_i\mathbf{z}_i$.

(1) 对 $k\le l$, 令 $Z=\mathrm{span}\{\mathbf{z}_k,\cdots,\mathbf{z}_n\}$, $X=\mathrm{span}\{\mathbf{x}_1,\cdots,\mathbf{x}_l\}$, $Y=\mathrm{span}\{\mathbf{y}_1,\cdots,\mathbf{y}_{(n-(l-k))}\}$, 由子空间和的维数公式可得

$$\begin{aligned}\dim(X\cap Y\cap Z)&=\dim(X\cap Y)+\dim(Z)-\dim\left((X\cap Y)+Z\right)\\&\ge\dim(X\cap Y)+\dim(Z)-n\\&\ge(\dim(X)+\dim(Y)-n)+\dim(Z)-n\\&=l+n-(l-k)-n+(n-k+1)-n\\&=1.\end{aligned}$$

取单位向量 $\mathbf{v}\in X\cap Y\cap Z$, 可得

$$c_k\le||C\mathbf{v}||=||A\mathbf{v}+B\mathbf{v}||\le||A\mathbf{v}||+||B\mathbf{v}||\le a_l+b_{n-(l-k)}.$$

(2) 对 $i \geq j$, 令 $\widetilde{X} = \mathrm{span}\{\mathbf{x}_j, \cdots, \mathbf{x}_n\}$, $\widetilde{Y} = \mathrm{span}\{\mathbf{y}_{i-j+1}, \cdots, \mathbf{y}_n\}$, $\widetilde{Z} = \mathrm{span}\{\mathbf{z}_1, \cdots, \mathbf{z}_i\}$, 则

$$\dim(\widetilde{X} \cap \widetilde{Y} \cap \widetilde{Z}) \geq \dim(\widetilde{X}) + \dim(\widetilde{Y}) + \dim(\widetilde{Z}) - 2n = 1.$$

取单位向量 $\mathbf{w} \in \widetilde{X} \cap \widetilde{Y} \cap \widetilde{Z}$, 可得

$$c_i \geq \langle \mathbf{w}, C\mathbf{w} \rangle = \langle \mathbf{w}, A\mathbf{w} \rangle + \langle \mathbf{w}, B\mathbf{w} \rangle \geq a_j + b_{i-j+1}.$$

注: 其实只需要对 $-A, -B, -C$ 使用 (1) 的结论就可以得到 (2) 的结论; 或者对 $-A, -B, -C$ 使用 (2) 的结论就可以得到 (1) 的结论. □

9.6 一般线性空间与线性映射

例 9.6.1. 定义 $M_2(\mathbf{R})$ 上的线性变换 $T : M_2(\mathbf{R}) \to M_2(\mathbf{R})$ 满足

$$T(A) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} A \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

求 T 在基 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 下的矩阵.

证明: 将题述基的成员依次记为 $\mathbf{v}_1, \cdots, \mathbf{v}_4$, 则任何矩阵 $A = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ 在此基下的表示为 $A = x\mathbf{v}_1 + y\mathbf{v}_2 + z\mathbf{v}_3 + w\mathbf{v}_4$. 直接计算矩阵乘法, 有

$$T(\mathbf{v}_1) = T(\mathbf{v}_3) = 2\mathbf{v}_1 + 2\mathbf{v}_3, \quad T(\mathbf{v}_2) = T(\mathbf{v}_4) = \mathbf{v}_2 + \mathbf{v}_4.$$

由此可得 T 在此基底下的表示矩阵为

$$\begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

□

例 9.6.2. 设 σ 是 n 维线性空间 V 上的线性变换, 满足 $\sigma^2 = \sigma$. 证明: (1) $\mathrm{Im}(\sigma) \bigcap \mathrm{Ker}(\sigma) = \{\mathbf{0}\}$. (2) $V = \mathrm{Im}(\sigma) \oplus \mathrm{Ker}(\sigma)$.

证明: (1) 设 $\mathbf{x} \in \mathrm{Im}(\sigma) \bigcap \mathrm{Ker}(\sigma)$, 则存在 $\mathbf{y}$ 使得 $\mathbf{x} = \sigma(\mathbf{y})$ 且 $\sigma(\mathbf{x}) = \mathbf{0}$. 由此可得

$$\mathbf{0} = \sigma(\mathbf{x}) = \sigma^2(\mathbf{y}) = \sigma(\mathbf{y}) = \mathbf{x}.$$

这就证明了 $\mathrm{Im}(\sigma) \bigcap \mathrm{Ker}(\sigma) = \{\mathbf{0}\}$.

(2) 由 (1) 的结论, 可得

$$\begin{aligned}\dim(\mathrm{Im}(\sigma) + \mathrm{Ker}(\sigma)) &= \dim(\mathrm{Im}(\sigma)) + \dim(\mathrm{Ker}(\sigma)) - \dim(\mathrm{Im}(\sigma) \cap \mathrm{Ker}(\sigma)) \\ &= \dim(\mathrm{Im}(\sigma)) + \dim(\mathrm{Ker}(\sigma)) \\ &= \dim(V),\end{aligned}$$

即子空间 $\mathrm{Im}(\sigma) + \mathrm{Ker}(\sigma)$ 的维数等于全空间 V 的维数, 所以有 $\mathrm{Im}(\sigma) + \mathrm{Ker}(\sigma) = V$. 再结合 (1) 的结论就得到 $V = \mathrm{Im}(\sigma) \oplus \mathrm{Ker}(\sigma)$. □

例 9.6.3. 设 W 为数域 F 上的 n 维向量空间 F^n 的子空间, 且 $W \neq F^n$. 证明: W 为 F 上某个 n 元齐次线性方程组的解空间.

证明: 取 W 的基 $\mathbf{w}_1, \ldots, \mathbf{w}_m$, 并将它扩充为 F^n 的基 $\mathbf{w}_1, \ldots, \mathbf{w}_n$. 设从标准基到上述基的过度矩阵为 T, 即有

$$[\mathbf{w}_1, \ldots, \mathbf{w}_n] = [\mathbf{e}_1, \ldots, \mathbf{e}_n]T.$$

对 F^n 中的每个向量 $\mathbf{x}$, 设它在基 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 下的坐标为 $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, 则有

$$[\mathbf{e}_1, \ldots, \mathbf{e}_n] \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = [\mathbf{w}_1, \ldots, \mathbf{w}_n] \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = [\mathbf{e}_1, \ldots, \mathbf{e}_n]T \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

由此可得

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = T^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

这样, 就把 W 实现成解空间:

$$\mathbf{x} \in W \iff y_{m+1} = \cdots = y_n = 0 \iff \sum_{j=1}^{n} (T^{-1})_{ij} x_j = 0, \quad \forall i \geq m+1.$$

□

例 9.6.4. 已知在 $\mathbf{R}^3$ 中, 线性变换 T 在基

$$\mathbf{v}_1 = \begin{bmatrix} 8 \\ -6 \\ 7 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 7 \\ -13 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 9 \\ -3 \\ 7 \end{bmatrix}$$

下的表示矩阵为

$$A = \begin{bmatrix} 1 & -18 & 15 \\ -1 & -22 & 20 \\ 1 & -25 & 22 \end{bmatrix}.$$

求 T 在基

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

下的表示矩阵.

解. 记 $\mathbf{R}^3$ 的标准基为 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, 则从基 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 到基 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 的过渡矩阵为

$$[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]V,$$

其中 V 是将三个列向量 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 从左至右排列成的 3×3 矩阵. 类似的,

$$[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]W.$$

结合这两者, 可知从基 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 到基 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ 的过渡矩阵为

$$[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]W = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]V^{-1}W.$$

已知 T 在 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 下的表示矩阵为 A, 则 T 在 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ 下的表示矩阵为

$$(V^{-1}W)^{-1}A(V^{-1}W) = W^{-1}VAV^{-1}W.$$

□

9.7 样卷参考解答

例 9.7.1 (8 分)**.** 判断以下矩阵是否可以相似对角化，并简单说明理由。

$$(a) \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} \qquad (b) \begin{bmatrix} 100 & 200 \\ 0 & 100 \end{bmatrix} \qquad (c) \begin{bmatrix} 23 & 69 & 188 \\ 69 & 45 & 202 \\ 188 & 202 & 68 \end{bmatrix} \qquad (d) \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 4 & 8 & 12 \end{bmatrix}$$

解. (a) 可以. 矩阵的特征多项式为 $(\lambda-3)(\lambda-5)$, 有两个单特征值 $3,5$, 所以该矩阵可以对角化.

(b) 不可以. 矩阵的特征多项式为 $(\lambda-100)^2$, 只有一个特征值 $\lambda=100$, 其代数重数为 2. 我们来计算此特征值的几何重数, 即其特征子空间 $V=\{\mathbf{x}|(A-100I_2)\mathbf{x}=\mathbf{0}\}$ 的维数, 解得 $V=\{\begin{bmatrix} x_1 \\ 0 \end{bmatrix}|x_1\in\mathbf{R}\}$, 可知 $\lambda=100$ 的几何重数等于 1, 小于代数重数. 由此可知所考虑的矩阵不可对角化.

(c) 可以. 这是实对称方阵, 我们证明了实对称方阵都可以正交相似于对角矩阵.

(d) 可以. 记此矩阵为 A, 注意到它的三个行向量成比例, 则 $\det(A)=0$. 由此可知 $\lambda=0$ 是 A 的特征值, 对应的特征子空间 $V_0=\{A\mathbf{x}=\mathbf{0}\}=\{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}|x_1+2x_2+3x_3=0\}$ 是 2 维的, 故 $\lambda=0$ 的几何重数为 2, 代数重数至少为 2. 结合 A 的所有特征值 (计重数) 的总和等于 $\operatorname{trace}(A)=17$, 可得 A 还有一个单特征值 17, 且特征值 0 的代数重数等于 2. 这样, A 的所有特征值都是半单的, 所以 A 可以对角化.

或者, 注意到 $\operatorname{rank}(A)=1$, 由维数公式知 $\dim(\mathcal{N}(A))=2$, 即特征子空间 V_0 的维数为 2, 可知特征值 $\lambda=0$ 的几何重数等于 2.

□

例 9.7.2 (8 分). 判断以下实对称阵是否正定，并简单说明理由。

(a) $\begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ -2 & -2 & 0 \\ 3 & 3 & 3 \end{bmatrix}\begin{bmatrix} 1 & -2 & 3 \\ 0 & -2 & 3 \\ 0 & 0 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 7 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$

解. (a) 正定, 因为其顺序主子式都大于零.

(b) 不正定, 该矩阵可写成 $X=AB$, 则 $\operatorname{rank}(X)\le\min\{\operatorname{rank}(A),\operatorname{rank}(B)\}\le 2$. 由于正定矩阵都是满秩的, 所以 X 不正定.

(c) 正定. 该矩阵可写成 $X=C^TC$, 我们证明过: C^TC 正定当且仅当 C 列满秩. 注意到本小问中的 C 是上三角矩阵, 其行列式等于 $1\cdot(-2)\cdot 3\neq 0$, 故 C 满秩. 结合这两点可知 X 正定.

(d) 不正定. 该矩阵可写成 $A=C^TBC$, 其中 C 是可逆矩阵, 我们断言: A 正定当且仅当

B 正定. 这是由于, A 给出的二次函数为

$$Q(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle = \langle \mathbf{x}, C^T BC\mathbf{x} \rangle = \langle C\mathbf{x}, BC\mathbf{x} \rangle,$$

可换元 $\mathbf{y} = C\mathbf{x}$, 则有 $Q(\mathbf{x}) = \langle \mathbf{y}, B\mathbf{y} \rangle$, 可表示为由 B 给出的二次型. 注意到, 由于 C 可逆, 当 $\mathbf{x}$ 取遍非零向量时 $\mathbf{y}$ 也取遍非零向量, 反之亦然. 结合正定矩阵的定义, 即可知 A 正定当且仅当 B 正定.

本小问中 B 是对角阵, 有负的对角元, 故 B 不正定, 从而 A 也不正定. □

例 9.7.3 (10 分)**.** 设 $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$, 分别找出 $N(A)$, $N(A^T)$, $C(A)$, $C(A^T)$ 的一组基。

解. 将 A 写成 $A = BC$. 注意到 B 是下三角矩阵且对角元都非零, 可知 B 可逆.

(1) 由 B 可逆, 有 $\mathcal{N}(A) = \mathcal{N}(C)$, 而 C 是阶梯形矩阵, x_1, x_2, x_4 是主变量, 可得 C 的零空间的一组基

$$(-1, -1, 1, 0, 0)^T, \quad (0, 0, 0, -1, 1)^T.$$

(2) 由 B 可逆, 它将 $\mathcal{R}(C) = C(C)$ 线性同构成 $\mathcal{R}(A) = C(A)$. 由前一小问的结论 $\dim(\mathcal{N}(C)) = 2$, 利用维数公式有 $\dim(\mathcal{R}(C)) = 5 - 2 = 3$. 显然如下三个向量

$$(1, 0, 0)^T, \quad (0, 1, 0)^T, \quad (1, 1, 1)^T$$

是 $\mathcal{R}(C)$ 的一组基. 它们被线性映射 B 映成 $\mathcal{R}(A)$ 的一组基:

$$(1, 2, 4)^T, \quad (0, 3, 5)^T, \quad (1, 5, 15)^T.$$

(实际上, 由 $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(C)) = 3$ 知 A 是满射, 则 $\mathbf{R}^3$ 的任何一组基都是 $\mathcal{R}(A)$ 的基)

(3) 由 $\mathcal{N}(A^T) = \mathcal{R}(A)^\perp$ 可知 $\mathcal{N}(A^T) = \{\mathbf{0}\}$, 空集是它的基.

(4) 由 $A^T = C^T B^T$ 以及 B 可逆, 有 $\mathcal{R}(A^T) = \mathcal{R}(C^T)$. 显然 C^T 的三个列向量线性无关, 构成 $\mathcal{R}(A^T) = \mathcal{R}(C^T)$ 的一组基:

$$(1, 0, 1, 1, 1)^T, \quad (0, 1, 1, 1, 1)^T, \quad (0, 0, 0, 1, 1)^T.$$

□

例 9.7.4 (5 分)**.** 设 $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}$. 求 P 的特征多项式, 并说明理由。

解. 记 $P = A(A^TA)^{-1}A^T$,由于 A 列满秩, 可知 $\dim(\mathcal{R}(A)) = 2$, 且 P 是向 A 的列空间 $\mathcal{R}(A)$ 的正交投影矩阵. 我们证明过: 正交投影矩阵满足 $P^T = P$ 且 $P^2 = P$, 由前者可知 P 可对角化, 其特征值都半单; 由后者可知 P 的特征值 λ 都满足 $\lambda^2 = \lambda$, 从而 λ 只能取 0 与 1. 考虑 0 的特征子空间 V_0, 有 $\dim(V_0) = \dim(\mathcal{N}(A)) = 3 - \dim(\mathcal{R}(A)) = 1$, 这表明特征值 0 的几何重数等于 1. 这样, P 的所有特征值为 0 与 1, 代数重数分别为 1 与 2. 由此可得 P 的特征多项式为 $\lambda(\lambda-1)^2$. □

例 9.7.5 (16 分)**.** 设

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

等号右边的第一个矩阵记为 Q，第二个矩阵记为 R.

(1) (2 分) 验证 $Q^TQ = I$.

(2) (6 分) 求到 $C(A)$ 的投影矩阵。

(3) (8 分) 设 $\boldsymbol{b} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. 求 $A\boldsymbol{x} = \boldsymbol{b}$ 的最小二乘解。

解. (1) 即验证 Q 的列向量构成单位正交基.

(2) C 是对角元非零的上三角矩阵, 从而是可逆矩阵. 由此可知 $A = QC$ 的像集 $\mathcal{R}(A)$ 等于 Q 的像集 $\mathcal{R}(Q)$. 前一小问验证了 Q 的列向量构成 $\mathcal{R}(Q) = \mathcal{R}(A)$ 的单位正交基, 则到 $\mathcal{R}(A)$ 的正交投影矩阵为 $P_A = QQ^T$, 它的矩阵元等于 Q 的各个行向量之间的内积. 直接计算可得

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}.$$

(3) $\mathbf{x}$ 是最小二乘解当且仅当 $A\mathbf{x} = P_A\mathbf{b}$, 即 $QR\mathbf{x} = QQ^T\mathbf{b}$. 在后面这个方程两边左乘以 Q^T, 可将它同解变形为:

$$Q^TQR\mathbf{x} = Q^TQQ^T\mathbf{b} \iff R\mathbf{x} = Q^T\mathbf{b}.$$

直接求解可得最小二乘解为

$$\mathbf{x} = (4 + 2\sqrt{2}, -2, 2)^T.$$

□

例 9.7.6 (6 分)**.** 已知: 整数 $1653, 2581, 3451, 4582$ 可以被 29 整除. 证明下面的四阶行列式值被 29 整除.

$$\begin{vmatrix} 1 & 6 & 5 & 3 \\ 2 & 5 & 8 & 1 \\ 3 & 4 & 5 & 1 \\ 4 & 5 & 8 & 2 \end{vmatrix}$$

证明: 对矩阵做三次列倍加变换, 对每个 $1 \leq i \leq 3$, 分别将第 4 列加上第 i 列的 10^{4-i} 倍, 矩阵的行列式不变, 可得

$$\det(A) = \begin{vmatrix} 1 & 6 & 5 & 1653 \\ 2 & 5 & 8 & 2581 \\ 3 & 4 & 5 & 3451 \\ 4 & 5 & 8 & 4582 \end{vmatrix}$$

再对第四列做 Laplace 展开, 结合条件即得 $\det(A)$ 能被 29 整除.

□

例 9.7.7 (6 分)**.** 解关于 x 的方程

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & -2 & 4 & -8 \\ 1 & x & x^2 & x^3 \end{vmatrix} = 0.$$

解. 题述矩阵为范德蒙矩阵 $V(x_1, x_2, x_3, x_4)$, 其行列式等于

$$\det V(x_1, x_2, x_3, x_4) = \prod_{i<j}(x_j - x_i).$$

由此可得要求解的方程为:

$$(2-1)(-2-1)(x-1)(-2-2)(x-2)(x+2) = 0,$$

解为 $x = 1, 2, -2$. □

例 9.7.8 (6 分)**.** 定义 $M_2(\mathbb{R})$ 上线性变换 $T: M_2(\mathbb{R}) \to M_2(\mathbb{R})$ 满足

$$T(A) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} A \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

求 T 在基 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 下的矩阵。

解. 将题述基的成员依次记为 $\mathbf{v}_1, \cdots, \mathbf{v}_4$, 则任何矩阵 $A = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ 在此基下的表示为 $A = x\mathbf{v}_1 + y\mathbf{v}_2 + z\mathbf{v}_3 + w\mathbf{v}_4$. 直接计算矩阵乘法, 有

$$T(\mathbf{v}_1) = T(\mathbf{v}_3) = 2\mathbf{v}_1 + 2\mathbf{v}_3, \quad T(\mathbf{v}_2) = T(\mathbf{v}_4) = \mathbf{v}_2 + \mathbf{v}_4.$$

由此可得 T 在此基底下的表示矩阵为

$$\begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

□

例 9.7.9 (20 分)**.** 设 $A = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix}$.

(a) (10 分) 求 A 的奇异值分解 $A = U\Sigma V^T$，其中 U 是 3 阶正交阵，V 是 2 阶正交阵。

(b) (2 分) 应用 (a) 写出 A 的四个基本子空间的一组标准正交基。

(c) (8 分) 设 $M = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$. 若 $A\boldsymbol{v} = \sigma\boldsymbol{u}$, 其中 $\boldsymbol{u}, \boldsymbol{v}$ 是奇异向量 (singular vector), σ 是奇异值 (singular value), 证明 $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}$ 是 M 的特征向量，并由此应用 A 的奇异向量给出 5 阶正交阵 Q，使得 Q^TMQ 是对角阵.

解. (a) 直接计算可知 $A^TA = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$, 其特征多项式为 $(\lambda - 5)^2 - 1$. 解得 A^TA 有两个特征值 6 与 4, 相应的特征向量分别为

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, 1)^T, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}}(1, -1)^T.$$

由奇异值分解定理, 令

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}}A\mathbf{v}_1 = \frac{1}{\sqrt{3}}(1,1,1)^T, \quad \mathbf{u}_2 = \frac{1}{\sqrt{4}}A\mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,0,-1)^T.$$

添加 $\mathbf{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1)$ 将 $\mathbf{u}_1, \mathbf{u}_2$ 扩充为 $\mathbf{R}^3$ 的单位正交基 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. 令 $\Sigma = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$, 以及

$$V = [\mathbf{v}_1, \mathbf{v}_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad U = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix},$$

即得到了 A 的奇异值分解 $A = U\Sigma V^T$.

(b) 由 (a) 中的计算以及奇异值分解定理, 可得: $\mathcal{N}(A) = \{\mathbf{0}\}$, 基是空集; $\mathcal{R}(A) = \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\mathbf{u}_1, \mathbf{u}_2$ 是基; $\mathcal{N}(A^T) = \mathrm{span}\{\mathbf{u}_3\}$, $\mathbf{u}_3$ 是基; $\mathcal{R}(A^T) = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, $\mathbf{v}_1, \mathbf{v}_2$ 是基.

(c) 若 $\boldsymbol{u}, \boldsymbol{v}$ 是属于奇异值 σ 的奇异向量, 则有 $A\boldsymbol{v} = \sigma\boldsymbol{u}, A^T\boldsymbol{u} = \sigma\boldsymbol{v}$, 由此可得

$$M\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} A\boldsymbol{v} \\ A^T\boldsymbol{u} \end{bmatrix} = \sigma\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix},$$

可知 $\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}$ 是 M 的特征向量.

设 M 对应的线性映射为 $f: \mathbf{R}^{3+2} \to \mathbf{R}^{3+2}$, 并把 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{v}_1, \mathbf{v}_2$ 视为 $\mathbf{R}^{3+2}$ 中的向量, 即

$$\mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ 0 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

它们构成 $\mathbf{R}^{3+2}$ 的单位正交基, 且 f 的作用为

$$f(\mathbf{u}_1) = \sqrt{6}\mathbf{v}_1, f(\mathbf{u}_2) = \sqrt{4}\mathbf{v}_2, f(\mathbf{u}_3) = \mathbf{0}, f(\mathbf{v}_1) = \sqrt{6}\mathbf{u}_1, f(\mathbf{v}_2) = \sqrt{4}\mathbf{u}_2,$$

由此可得

$$f(\frac{\mathbf{u}_1 \pm \mathbf{v}_1}{\sqrt{2}}) = \pm\sqrt{6} \cdot \frac{\mathbf{u}_1 \pm \mathbf{v}_1}{\sqrt{2}}, \quad f(\frac{\mathbf{u}_2 \pm \mathbf{v}_2}{\sqrt{2}}) = \pm\sqrt{4} \cdot \frac{\mathbf{u}_2 \pm \mathbf{v}_2}{\sqrt{2}}, \quad f(\mathbf{u}_3) = \mathbf{0}.$$

选取 $\frac{\mathbf{u}_1 \pm \mathbf{v}_1}{\sqrt{2}}, \frac{\mathbf{u}_2 \pm \mathbf{v}_2}{\sqrt{2}}, \mathbf{u}_3$ 作为 $\mathbf{R}^{3+2}$ 的基, 即可对角化 M. 具体的说, 令

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix},$$

则可得 $Q^T M Q = \mathrm{diag}\{\sqrt{6}, -\sqrt{6}, \sqrt{4}, -\sqrt{4}, 0\}$.

□

例 9.7.10 (10 分)**.** 在以下两题中选且仅选一道题完成。

(1) $C: 3x_1^2 + 4x_1x_2 + 6x_2^2 = 1$ 是实平面上哪种二次曲线，椭圆、双曲还是抛物线? 若 C 是椭圆，请算出它的长、短轴长，以及长、短轴所在的直线方程；若 C 是双曲线，请算出它的虚、实轴长以及虚、实轴所在的直线方阵，以及两条渐近线方程；若 C 是抛物线，请算出它的顶点以及对称轴方程。

(2) 令 $A = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$. 求 4 阶正交阵 Q 和对角阵 Λ 使得 $Q^T A Q = \Lambda$.

解. (1) 曲线 C 的定义方程可以写成

$$1 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

. 记上式右边二次型对应的实对称矩阵为 A, 其特征多项式为

$$p_A(\lambda) = (\lambda - 3)(\lambda - 6) - (-2)(-2) = (\lambda - 2)(\lambda - 7),$$

特征值为 $\lambda = 2$ 或 $\lambda = 7$. 通过求解方程组 $(A - \lambda I_2)\mathbf{x} = \mathbf{0}$ 可找出 A 的特征向量: $\frac{1}{\sqrt{5}}(-2, 1)$ 是 $\lambda = 2$ 的特征向量; $\frac{1}{\sqrt{5}}(1, 2)$ 是 $\lambda = 7$ 的特征向量. 令 $Q = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix}$, 则有 $Q^T A Q = \mathrm{diag}\{2, 7\}$. 换元

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Q^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Q^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

则 C 的定义方程为

$$\begin{aligned}1 &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} Q \begin{bmatrix} 2 & \\ & 7 \end{bmatrix} Q^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 2 & \\ & 7 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.\end{aligned}$$

注意到, Q 是正交矩阵, 从 (x_1, x_2) 平面到 (y_1, y_2) 平面的变换是等距同构, 保持一切几何量. 在 (y_1, y_2) 平面中, C 的定义方程是 $2y_1^2 + 7y_2^2 = 1$, 它是椭圆, 并可进一步确定 C 的所有几何量, 比如长短轴长度. 另外, 还可以求出 C 的焦点, 长短轴的直线方程, 再用 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Q \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ 变换到 (x_1, x_2) 平面中去.

(2) 即要找出实对称矩阵 A 的谱分解. 注意到 A 保持子空间 $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_3\}$ 与 $\mathrm{span}\{\mathbf{e}_2, \mathbf{e}_4\}$. 只需分别计算 A 在这两个子空间上的谱分解.

在 $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_3\}$ 上 A 的表示矩阵为 $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, 特征值为 $0, 1$, 对应的特征向量分别为 $\mathbf{e}_1 - \mathbf{e}_3, \mathbf{e}_1 + \mathbf{e}_3$.

在 $\mathrm{span}\{\mathbf{e}_2, \mathbf{e}_4\}$ 上 A 的表示矩阵为 $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$, 特征值为 $-\frac{1}{3}, 1$, 对应的特征向量分别为 $\mathbf{e}_2 - \mathbf{e}_4, \mathbf{e}_2 + \mathbf{e}_4$.

结合起来, 如下四个特征向量

$$\frac{1}{\sqrt{2}}(\mathbf{e}_1 - \mathbf{e}_3), \quad \frac{1}{\sqrt{2}}(\mathbf{e}_1 + \mathbf{e}_3), \quad \frac{1}{\sqrt{2}}(\mathbf{e}_2 - \mathbf{e}_4), \quad \frac{1}{\sqrt{2}}(\mathbf{e}_2 + \mathbf{e}_4)$$

构成 $\mathbf{R}^4$ 的单位正交基, 可用它们给出 A 的谱分解. 令

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & -\frac{1}{3} & \\ & & & 1 \end{bmatrix},$$

即有 $Q^T A Q = \Lambda$.

□

例 9.7.11 (5 分)**.** 设 $A \in M_{m\times n}(\mathbb{R})$, A 的算子范数 (operator norm) 是

$$\|A\| = \max_{\substack{\boldsymbol{v}\in\mathbb{R}^n \\ \|\boldsymbol{v}\|=1}} \|A\boldsymbol{v}\| = \max_{\substack{\boldsymbol{v}\in\mathbb{R}^n \\ \boldsymbol{v}\neq\mathbf{0}}} \frac{\|A\boldsymbol{v}\|}{\|\boldsymbol{v}\|}.$$

试证:

$$\|A\| = \max_{\substack{\boldsymbol{u}\in\mathbb{R}^m, \boldsymbol{v}\in\mathbb{R}^n \\ \|\boldsymbol{u}\|=\|\boldsymbol{v}\|=1}} \boldsymbol{u}^T A \boldsymbol{v}.$$

证明: 对 $\mathbf{R}^m$ 中给定的向量 $\mathbf{y}$, 对任何单位向量 $\mathbf{u} \in \mathbf{R}^m$, 由 Cauchy-Schwartz 不等式有

$$\mathbf{u}^T\mathbf{y} \leq ||\mathbf{u}|| \cdot ||\mathbf{y}|| = ||\mathbf{y}||.$$

当 $\mathbf{y} = \mathbf{0}$ 时, 任取 $\mathbf{u}$; 当 $\mathbf{y} \neq \mathbf{0}$ 时, 取 $\mathbf{u} = \frac{\mathbf{y}}{||\mathbf{y}||}$, 可使上述不等式成立等号. 这样就有

$$\max_{||\mathbf{u}||=1} \mathbf{u}^T\mathbf{y} = ||\mathbf{y}||.$$

由此可得

$$\max_{||\mathbf{u}||=||\mathbf{v}||=1} \mathbf{u}^T A\mathbf{v} = \max_{||\mathbf{v}||=1} \left(\max_{||\mathbf{u}||=1} \mathbf{u}^T A\mathbf{v} \right) = \max_{||\mathbf{v}||=1} ||A\mathbf{v}|| = ||A||.$$

□

9.8 习题课题目

例 9.8.1. 设 A 是一个 $m \times n$ 实矩阵, $\mathbf{b} \in \mathbf{R}^m$. 证明: $A\mathbf{x} = \mathbf{b}$ 的全部最小二乘解的集合是

$$\{A^+\mathbf{b} + (I_n - A^+A)\mathbf{y} | \mathbf{y} \in \mathbf{R}^n\}.$$

证明: 我们证明过:

(1) $\mathbf{x}$ 是最小二乘解当且仅当 $A\mathbf{x} = P_A\mathbf{b}$;

(2) $\mathbf{x}_0 = A^+\mathbf{b}$ 是模长最小的最小二乘解;

(3) $A^+A = V_rV_r^T$ 是向 $\text{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_r\}) = \mathcal{N}(A)^\perp$ 的正交投影.

一方面, 由前两点可知全部最小二乘解为

$$A^+\mathbf{b} + \mathcal{N}(A) = A^+\mathbf{b} + \text{span}(\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}).$$

另一方面, 记 $\mathbf{R}^n$ 中的元素 $\mathbf{y} = \sum_{i=1}^n y_i\mathbf{v}_i$, 利用前述第 3 个性质可得

$$(I_n - A^+A)\mathbf{y} = \sum_{i=r+1}^n y_i\mathbf{v}_i,$$

即有 $(I_n - A^+A)(\mathbf{R}^n) = \text{span}(\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\})$.

结合这两方面, 就完成了证明. □

例 9.8.2. 设 A 是 n 阶实对称矩阵. 证明: A 的奇异值与 A 的特征值相同当且仅当 A 是半正定矩阵.

证明: 设实对称矩阵 A 的全部 n 个特征值为 $\lambda_1, \dots, \lambda_n$, 则 $A^TA = A^2$ 的全部特征值为 $\lambda_1^2, \dots, \lambda_n^2$. 依定义, A 的奇异值为 A^TA 的特征值的非负平方根, 可知 A 的奇异值为 $|\lambda_1|, \dots, |\lambda_n|$.

这样, A 的奇异值与 A 的特征值相同当且仅当每个 λ_i 都非负 $(1 \le i \le n)$, 当且仅当 A 是半正定的. □

例 9.8.3. 设 $m \times n$ 实矩阵 A 的奇异值分解为 $A = U\Sigma V^T$, 满足 $U = [\mathbf{u}_1, \dots, \mathbf{u}_m]$ 是 m 阶正交矩阵, $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ 是 n 阶正交矩阵, $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}_{m\times n}$, 其中 $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$, $r = \mathrm{rank}(A)$. 求 $A - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$ 的奇异值分解.

解. 对每个 $i \ge 2$, $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$ 作用在 $\mathbf{v}_i$ 上都等于零; $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$ 作用在 $\mathbf{v}_1$ 上等于 $\sigma_1 \mathbf{u}_1$. 由此可得 $A - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$ 的奇异值分解为:

$$A - \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = U \begin{bmatrix} 0 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \end{bmatrix} V^T$$

. □

例 9.8.4. 设 $\mathbf{v} \in \mathbf{R}^n, \mathbf{u} \in \mathbf{R}^m$ 且 $||\mathbf{u}|| = ||\mathbf{v}|| = 1$. 求 $A = \mathbf{u}\mathbf{v}^T$ 的奇异值分解.

解. 将 $\mathbf{v}$ 扩充为 $\mathbf{R}^n$ 的单位正交基 $\mathbf{v}_1 = \mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n$; 将 $\mathbf{u}$ 扩充为 $\mathbf{R}^m$ 的单位正交基 $\mathbf{u}_1 = \mathbf{u}, \mathbf{u}_2, \dots, \mathbf{u}_m$. 对每个 $i \ge 2$, A 作用在 $\mathbf{v}_i$ 上都等于零; A 作用在 $\mathbf{v}_1$ 上等于 $\mathbf{u}_1$. 由此可得 A 的奇异值分解为:

$$A = U \begin{bmatrix} 1 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \end{bmatrix} V^T.$$

□

例 9.8.5. (1) 设 $A, B \in M_{m\times n}(\mathbf{R})$, $A, B, A+B$ 的最大奇异值分别是 r, s, t. 证明: $r + s \ge t$.

(2) 设 $A \in M_{m\times n}(\mathbf{R}), B \in M_{n\times p}(\mathbf{R})$, A, B, AB 的最大奇异值分别是 r, s, t. 证明: $rs \ge t$.

(3) 设 $A \in M_n(\mathbf{R})$ 的奇异值都是正数, σ_1 是最大奇异值, σ_r 是最小奇异值, λ 是 A 的任意实特征值. 证明: $\sigma_r \le |\lambda| \le \sigma_1$.

证明: 我们证明了如下命题: A 的谱范数等于 A 的最大奇异值 σ_1, 即有

$$\sigma_1 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{||A\mathbf{x}||}{||\mathbf{x}||}.$$

(1) 由前述命题, 存在非零向量 $\mathbf{x}$, 使得 $t = \frac{||(A+B)\mathbf{x}||}{||\mathbf{x}||}$, 由此可得

$$t = \frac{||(A+B)\mathbf{x}||}{||\mathbf{x}||} \le \frac{||A\mathbf{x}||}{||\mathbf{x}||} + \frac{||B\mathbf{x}||}{||\mathbf{x}||} \le r + s.$$

(2) 不妨设 $t > 0$. 由前述命题, 存在非零向量 $\mathbf{x}$ 使得 $t = \frac{||(AB)\mathbf{x}||}{||\mathbf{x}||}$, 则有 $B\mathbf{x} \neq \mathbf{0}$, 且进一步有

$$t = \frac{||A(B\mathbf{x})||}{||B\mathbf{x}||} \cdot \frac{||B\mathbf{x}||}{||\mathbf{x}||} \le r \cdot s.$$

(3) 设 $\mathbf{v}$ 是属于 λ 的实特征向量. 由 A 的奇异值都是正数, 在奇异值分解中 $r = n$. 设 $\mathbf{v} = \sum\limits_{i=1}^{n} x_i \mathbf{v}_i$, 则有 $A\mathbf{v} = \sum\limits_{i=1}^{n} x_i \sigma_i \mathbf{u}_i$, 由此可得

$$|\lambda| = \frac{||A\mathbf{v}||}{||\mathbf{v}||} = \frac{\sqrt{\sum_{i=1}^n x_i^2 \sigma_i^2}}{\sqrt{\sum_{i=1}^n x_i^2}} \in [\sigma_n, \sigma_1].$$

□

例 9.8.6. 考虑由所有定义在区间 $[-1,1]$ 上的不超过二次的多项式组成的线性空间 $P_2[x] = \{y(x)|y(x) = a + bx + cx^2\}$. 已知 $w_1(x), w_2(x), w_3(x) \in P_2[x]$ 且满足

$$\begin{aligned} &w_1(-1) = 1, \quad w_1(0) = 0, \quad w_1(1) = 0, \\ &w_2(-1) = 0, \quad w_2(0) = 1, \quad w_2(1) = 0, \\ &w_2(-1) = 0, \quad w_2(0) = 0, \quad w_2(1) = 1. \end{aligned}$$

(1) 证明: $w_1(x), w_2(x), w_3(x)$ 构成 $P_2[x]$ 的一组基.

(2) 取 $v_1(x) = 1, v_2(x) = x, v_3(x) = x^2$. 求从 v_1, v_2, v_3 到 w_1, w_2, w_3 的过渡矩阵, 以及从 w_1, w_2, w_3 到 v_1, v_2, v_3 的过渡矩阵.

(3) 考虑线性映射 $D : P_2[x] \to P_2[x]$, 它定义为 $D(f) = f'$, 即为求导算子. 求 D 在基 w_1, w_2, w_3 下的表示矩阵. (将表示矩阵记为 D)

(4) 定义 $P_2[x]$ 上的内积为:

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx.$$

在此内积下, 求 w_1, w_2, w_3 的 Gram-Schimdt 正交化.

解. (1) 如果实数 k_1,k_2,k_3 满足 $k_1w_1(x)+k_2w_2(x)+k_3w_3(x)=0$, 则有

$$\begin{bmatrix} w_1(-1) & w_2(-1) & w_3(-1) \\ w_1(0) & w_2(0) & w_3(0) \\ w_1(1) & w_2(1) & w_3(1) \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

由于上述方程组的系数矩阵为 I_3, 则上述方程组只有零解, 这表明 $w_1(x),w_2(x),w_3(x)$ 构成 $P_2[x]$ 的一组基.

(2) 设从 v_1,v_2,v_3 到 w_1,w_2,w_3 的过渡矩阵为 T, 即有

$$[w_1(x),w_2(x),w_3(x)]=[v_1(x),v_2(x),v_3(x)]T.$$

将上式分别在点 $x=-1,0,1$ 赋值, 可得

$$\begin{aligned} &[w_1(-1),w_2(-1),w_3(-1)]=[v_1(-1),v_2(-1),v_3(-1)]T, \\ &[w_1(0),w_2(0),w_3(0)]=[v_1(0),v_2(0),v_3(0)]T, \\ &[w_1(1),w_2(1),w_3(1)]=[v_1(1),v_2(1),v_3(1)]T. \end{aligned}$$

用矩阵的语言表示即为

$$\begin{bmatrix} w_1(-1) & w_2(-1) & w_3(-1) \\ w_1(0) & w_2(0) & w_3(0) \\ w_1(1) & w_2(1) & w_3(1) \end{bmatrix}=\begin{bmatrix} v_1(-1) & v_2(-1) & v_3(-1) \\ v_1(0) & v_2(0) & v_3(0) \\ v_1(1) & v_2(1) & v_3(1) \end{bmatrix}T,$$

解出

$$T=\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}^{-1}=\begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}.$$

从 w_1,w_2,w_3 到 v_1,v_2,v_3 的过渡矩阵为 T^{-1}.

(3) D 在 v_1,v_2,v_3 下的表示矩阵为

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

由此可知 D 在基 w_1,w_2,w_3 下的表示矩阵为

$$D=T^{-1}AT=\begin{bmatrix} -\frac{3}{2} & 2 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -2 & \frac{3}{2} \end{bmatrix}.$$

(4) 由第 (2) 小问的结论, 有

$$w_1(x) = \frac{1}{2}(x^2 - x), \quad w_2(x) = 1 - x^2, \quad w_3(x) = \frac{1}{2}(x^2 + x).$$

之后耐心的做 Gram-Schimdt 正交化. □

例 9.8.7. 设 3 阶实对称矩阵 A 有三个特征值 $3, 3, -3$. 已知属于特征值 (-3) 的特征向量为 $(1, -2, 1)^T$. 求矩阵 A 及 A^{-1}.

证明: 设 A 对应的线性映射为 $f: \mathbf{R}^3 \to \mathbf{R}^3$. 令 $\mathbf{q}_1 = \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$, 它是 A 的特征子空间 V_{-3} 的单位正交基; 取特征子空间 V_3 的单位正交基 $\mathbf{q}_2, \mathbf{q}_3$, 则 $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ 是 $\mathbf{R}^3$ 的单位正交基, 且有

$$f(\mathbf{q}_1) = -3\mathbf{q}_1, \quad f(\mathbf{q}_2) = 3\mathbf{q}_2, \quad f(\mathbf{q}_3) = 3\mathbf{q}_3.$$

对任何 $\mathbf{x} \in \mathbf{R}^3$, 有 $\mathbf{x} = \mathbf{q}_1\langle \mathbf{q}_1, \mathbf{x}\rangle + \mathbf{q}_2\langle \mathbf{q}_2, \mathbf{x}\rangle + \mathbf{q}_3\langle \mathbf{q}_3, \mathbf{x}\rangle$, 结合上式可得

$$f(\mathbf{x}) = -3\mathbf{q}_1\langle \mathbf{q}_1, \mathbf{x}\rangle + 3\mathbf{q}_2\langle \mathbf{q}_2, \mathbf{x}\rangle + 3\mathbf{q}_3\langle \mathbf{q}_3, \mathbf{x}\rangle = (-3\mathbf{q}_1\mathbf{q}_1^T + 3\mathbf{q}_2\mathbf{q}_2^T + 3\mathbf{q}_3\mathbf{q}_3^T)\mathbf{x},$$

即有 $A = -3\mathbf{q}_1\mathbf{q}_1^T + 3\mathbf{q}_2\mathbf{q}_2^T + 3\mathbf{q}_3\mathbf{q}_3^T$. 类似的, $I_3 = \mathbf{q}_1\mathbf{q}_1^T + \mathbf{q}_2\mathbf{q}_2^T + \mathbf{q}_3\mathbf{q}_3^T$. 结合起来可得

$$A = 3I_3 - 6\mathbf{q}_1\mathbf{q}_1^T.$$

用同样的方法, 可得

$$A^{-1} = -\frac{1}{3}\mathbf{q}_1\mathbf{q}_1^T + \frac{1}{3}\mathbf{q}_2\mathbf{q}_2^T + \frac{1}{3}\mathbf{q}_3\mathbf{q}_3^T = \frac{1}{3}I_3 - \frac{2}{3}\mathbf{q}_1\mathbf{q}_1^T.$$

□