线性代数笔记整理V2.2

陈鸿峥

2019.04*

目录

*Build 20190415

1 矩阵与线性方程组

1.1 基本概念

在此仅罗列一些基本的知识点的名字，具体内容本文不作赘述.

1. 解线性方程组（高斯消元法）
2. 方程组无解、唯一解、多解的判断，与主元列、主元位置的关系
   解的**存在性**：看每一列是否都有主元位置
   解的**唯一性**：看有无自由变元（$A\mathbf{x}=\mathbf{0}$是否有**非平凡解**同样看自由变元）
3. **齐次方程**$A\mathbf{x}=\mathbf{0}$解的参数表示
   非齐次方程$A\mathbf{x}=\mathbf{b}$解的形式为$\mathbf{x}=\mathbf{p}+t\mathbf{v}, t\in\mathbb{R}$，即非齐通解=齐通解+非齐特解，但要注意首先得有解
   这两种方程的具体比较见2.2节
4. 线性相关与线性无关
   要判定一组向量是线性相关还是线性无关，即判断$A\mathbf{x}=\mathbf{0}$是否有非平凡解
5. 基本矩阵运算（加减乘除、求逆、转置）

定义 1 (基本行变换/初等矩阵变换)**.** *替换、交换、数乘*

定义 2 (初等(elementary)矩阵)**.** *对单位矩阵做一次基本行变换所得的矩阵.*

下面是线性无关线性相关一些常见的结论.

定理 1. $S=\{\mathbf{v}_1,\dots,\mathbf{v}_p\}, |S|>1$

(*a*) *S线性相关，当且仅当至少有一向量可以表示为其他向量的线性组合（注意***不是***每一个）*

(b) 若S线性相关，$\mathbf{v}_1 \neq 0$，那么$\exists \mathbf{v}_j, j > 1$可以被表示为$\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$的线性组合

(c) 若S的子集S'线性相关，则S线性相关；若S线性无关，则S'线性无关（注意关系不能颠倒）

(d) 初等行变换不改变**列**之间的线性相关关系

分析. 记$A = \begin{bmatrix}\mathbf{v}_1, \ldots, \mathbf{v}_n\end{bmatrix}$，$B \sim A$为$A$的阶梯形

(a) 必要性：即$\exists c_1, \ldots, c_n$不全为0，使得$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$.
不妨设$c_1 \neq 0$，则$\mathbf{v}_1 = (-\dfrac{c_2}{c_1})\mathbf{v}_2 + \cdots + (-\dfrac{c_n}{c_1})\mathbf{v}_n$，证毕.
充分性：不妨设$\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$，移项得$(-1)\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$.

(b) 用((a))的结论，设$\mathbf{v}_p = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1} + c_{p+1}\mathbf{v}_{p+1} + \cdots + c_n\mathbf{v}_n, p > 1$，找出最大的$j$使得$c_j \neq 0$（$\mathbf{v}_p$的系数$c_p = -1$也计入），进而$c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j = \mathbf{0}$，接下来证明方法与((a))相同，$\mathbf{v}_j$即为所求.

(c) 因两者互为逆否命题，故只用证前者，而前者通过不断扩充S'并令扩充部分的权重全为0即可.

(d) $A\mathbf{x} = \mathbf{0}$与$B\mathbf{x} = \mathbf{0}$有相同的解，故线性相关关系不会改变.

1.2 矩阵的逆

如无特殊说明，在本节中我们讨论的矩阵均为方阵.

定义 3 (逆)**.** A为$n \times n$的矩阵，若$\exists C\ s.t.\ CA = I_n, AC = I_n$，则称$A$可逆.
注意：在定义中要求左乘右乘都满足才说其可逆.

定理 2. A是$m \times n$矩阵，若存在$n \times m$矩阵C和D使得$CA = I_n$，$AD = I_m$，则$m = n$且$C = D$

分析. 这个定理告诉我们只有方阵才会存在逆，下面来证明.
1° 证明若$CA = I_n$，则$m \geq n$
因为$CA\mathbf{x} = I_n\mathbf{x} = \mathbf{x} = \mathbf{0}$，故$A\mathbf{x} = \mathbf{0}$只有平凡解，没有自由变元，因此$m \geq n$.
2° 证明若$AD = I_m$，则$m \leq n$
因为$AD\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$，也即$\forall \mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$都有解$\mathbf{x} = D\mathbf{b}$，因此$m \leq n$.
3° 由上分析知，$m = n$.
考虑乘积CAD，有$CAD = (CA)D = I_nD = D = C(AD) = CI_m = C$，故$C = D$.

定理 3 (二阶矩阵的逆)**.**

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \Longrightarrow \quad A^{-1} = \frac{1}{\det A}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \det A \neq 0$$

命题 1. 若$A = BC$，B可逆，则$\begin{bmatrix} B & A \end{bmatrix} \sim \begin{bmatrix} I & C \end{bmatrix}$

分析.

$$EB = I \implies E = B^{-1}$$

$$EA = C \implies B^{-1}A = C \iff A = BC$$

算法 1 (求逆)**.**

$$\begin{bmatrix} A & I \end{bmatrix} \sim \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

分析. *在命题1中令*$A \leftarrow I$, $B \leftarrow A$, $C \leftarrow A^{-1}$*即可.*

若令$A^{-1} = \begin{bmatrix} \mathbf{x}_1, \ldots, \mathbf{x}_n \end{bmatrix}$，那么也可以将上述求逆过程看成是解下面的这个方程组

$$A\mathbf{x}_1 = \mathbf{e}_1, \quad A\mathbf{x}_2 = \mathbf{e}_2, \quad \ldots, \quad A\mathbf{x}_n = \mathbf{e}_n$$

其说明的是矩阵A乘上A^{-1}的第i列，将得到I_n的第i列，也即$\mathbf{e}_i$.

虽然只是变换了视角，但对于一些问题来说这种视角的变换是重要的.

定理 4 (可逆阵定理)**.** 对于$n \times n$矩阵A，下列说法均等价

- (a) A可逆
- (b) $A \sim I_n$
- (c) A有n个主元位置
- (d) $A\mathbf{x} = \mathbf{0}$只有平凡解
- (e) A的列线性无关
- (f) $\mathbf{x} \mapsto A\mathbf{x}$是单射
- (g) $A\mathbf{x} = \mathbf{b}, \forall b \in \mathbb{R}^n$至少有一解
- (h) A的列张成$\mathbb{R}^n$
- (i) $\mathbf{x} \mapsto A\mathbf{x}$是满射
- (j) $\exists\, n \times n\; C\, s.t.\, CA = I$
- (k) $\exists\, n \times n\; D\, s.t.\, AD = I$
- (l) A^{T}可逆
- (m) A的列构成$\mathbb{R}^n$的一组基
- (n) $\operatorname{Col} A = \mathbb{R}^n$
- (o) $\dim \operatorname{Col} A = n$
- (p) $\operatorname{rank} A = n$（满秩）
- (q) $\operatorname{Nul} A = \{\mathbf{0}\}$
- (r) $\dim \operatorname{Nul} A = 0$
- (s) 0**不是**A的特征值
- (t) $\det A \neq 0$
- (u) $(\operatorname{Col} A)^{\perp} = \operatorname{Nul} A^{\mathrm{T}} = \{\mathbf{0}\}$
- (v) $(\operatorname{Nul} A)^{\perp} = \operatorname{Row} A = \mathbb{R}^n$
- (w) $\operatorname{Row} A = \mathbb{R}^n$
- (x) A有n个非零奇异值

算法 2 (LU分解)**.** A为$m \times n$的矩阵，则A一定可以被分解为$A = LU$，其中L为下三角矩阵，U为上三角矩阵.以下为步骤

1. 初等行变换（**只进行替换**）将A变成行阶梯形U，并记录每次$E_1, \dots, E_p$
2. 则$L = (E_p \cdots E_1)^{-1}$，相当于将每次初等行变换逆过来，并将矩阵直接重合即可

注意到LU分解与方阵求逆的方法基本相同，但它适用于所有矩阵，算是矩阵分解里最简单的一种了.用LU分解解方程，加法乘法计算量会降低不少（因为只需回代）.

命题 2 (消去律)**.** 若A可逆，$AB = AC$，则$B = C$.

命题 3. 若AB可逆，则A，B都可逆. 若A**或**B不可逆，则AB不可逆.

分析. $C = AB \implies I = BC^{-1}A \implies A^{-1} = BC^{-1}$，$B$同理. 而后面一句为前面一句的逆否命题，故证毕.

定理 5. 上（下）三角矩阵的逆仍为上（下）三角

分析. 不妨设A为下三角矩阵，则$a_{ij} = 0, i \in \{1, \dots, n-1\}, j \in \{i+1, \dots, n\}$.
若$j \neq n-1$，注意到$a_{(n-1)n} = 0$一定被包含在A_{ji}中，且一定在A_{ji}的对角线上（最右下角）;
若$j = n-1$，$a_{(n-1)(n-3)} = 0$一定被包含在A_{ji}中，且也在A_{ji}对角线上（右下角倒数第二个元素）.
又注意到在$a_{ij} = 0, i \in \{1, \dots, n-1\}, j \in \{i+1, \dots, n\}$的前提下，$A_{ji}$一定为下三角矩阵，由上分析知对角线上一定存在0元素，故代数余子式C_{ji}为0. 进而伴随矩阵为下三角，原矩阵的逆为下三角.

1.3 分块矩阵

需要知道矩阵乘法另外的表示方法

$$AB = A\Big[\mathbf{b}_1, \ldots, \mathbf{b}_n\Big] = \Big[A\mathbf{b}_1, \ldots, A\mathbf{b}_n\Big]$$

类似地，分块矩阵有

$$AB = \Big[\mathbf{a}_1, \ldots, \mathbf{a}_n\Big]\begin{bmatrix}\mathbf{b}_1\\ \vdots \\ \mathbf{b}_n\end{bmatrix} = \sum_{i=1}^{n}\mathbf{a}_i\mathbf{b}_i$$

注意区别正常的矩阵乘法（行乘列得到某一位置上的元素），分块矩阵乘法是列乘行得到矩阵.

分块矩阵的转置除了每一个矩阵转置外，大的矩阵内部也需要转，如

$$A = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} \implies A^{\mathrm{T}} = \begin{bmatrix}A_{11}^{\mathrm{T}} & A_{21}^{\mathrm{T}}\\ A_{12}^{\mathrm{T}} & A_{22}^{\mathrm{T}}\end{bmatrix}$$

分块矩阵求逆并不是每一块分别求逆简单合并即可，而是需要将逆设出来，乘开后通过对比系数来求. 如下面定理6所示.

定理 6.

$$A = \begin{bmatrix}A_{11} & A_{12}\\ O & A_{22}\end{bmatrix} \implies A^{-1} = \begin{bmatrix}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1}\\ O & A_{22}^{-1}\end{bmatrix}$$

*特别地，当A_{12}为0矩阵，该分块矩阵的逆可以变成各个矩阵分别求逆. 实际上对于***对角矩阵***，其k次方都是主对角线上的元素直接取k次幂，求逆看成是-1次幂当然也成立，见推论1.*

分析. 设$B = \begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix}$，且

$$\begin{bmatrix}A_{11} & A_{12}\\ O & A_{22}\end{bmatrix}\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix} = \begin{bmatrix}I_p & O\\ O & I_q\end{bmatrix}$$

乘开后对比系数有

$$\begin{aligned}A_{11}B_{11} + A_{12}B_{21} &= I_p\\ A_{11}B_{12} + A_{12}B_{22} &= O\\ A_{22}B_{21} &= O\\ A_{22}B_{22} &= I_q\end{aligned}$$

由下往上回代即得结果.

推论 1. *$A = \begin{bmatrix}B & O\\ O & C\end{bmatrix}$，$B, C$均为方阵，则$A$可逆，当且仅当$B, C$都可逆.若$A$可逆，则$A^{-1} = \begin{bmatrix}B^{-1} & O\\ O & C^{-1}\end{bmatrix}$*

如求以下这个矩阵的逆就可以划分为对角矩阵后，直接对各块求逆合并.

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 5 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 7 & 8 \\ 0 & 0 & 0 & 5 & 6 \end{bmatrix}$$

当然，分块也是需要技巧的，常常需要先观察矩阵的性质，再做出合理划分，见例1.

例 1.

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 1 & 1 & & 0 \\ \vdots & & & \ddots & \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}, \text{证明} B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & & \cdots & -1 & 1 \end{bmatrix} \text{为} A \text{的逆}$$

分析. *递归定义分块矩阵：* $A_{k+1} = \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{v} & A_k \end{bmatrix}$, $B_{k+1} = \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{w} & B_k \end{bmatrix}$*，并用数学归纳法可证.*

定理 7. *两个上（下）三角矩阵的乘积仍为上（下）三角*

分析. *不妨设两个矩阵* A, B *都为下三角，且* $A = [a_{ij}], B = [b_{ij}], C = AB = [c_{ij}]$*，其中*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

当 $i < j$ *时，* $a_{ij} = b_{ij} = 0$*，故（自行画图观察可得以下事实）*

$$c_{ij} = \underbrace{a_{i1}}_{\neq 0}\underbrace{b_{1j}}_{=0} + \cdots + \underbrace{a_{ii}}_{\neq 0}\underbrace{b_{ij}}_{=0} + \underbrace{a_{i(i+1)}}_{=0}\underbrace{b_{(i+1)j}}_{=0} + \cdots + \underbrace{a_{ij}}_{=0}\underbrace{b_{jj}}_{\neq 0} + \cdots + \underbrace{a_{in}}_{=0}\underbrace{b_{nj}}_{\neq 0} = 0$$

定理 8 (舒尔(Schur)补)**.** *若* A_{11} *可逆，则*

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & O \\ X & I \end{bmatrix} \begin{bmatrix} A_{11} & O \\ O & S \end{bmatrix} \begin{bmatrix} I & Y \\ O & I \end{bmatrix}$$

其中 $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ *称为* A_{11} *的舒尔补，* $X = A_{21}A_{11}^{-1}$*，* $Y = A_{11}^{-1}A_{12}$*.*

分析. *类似定理6的方法，乘开后通过对比系数可得.*

1.4 行列式

命题 4 (行列式的基本运算)**.**

1. $\det A^{\mathrm{T}} = \det A$

2. $\det AB = (\det A)(\det B) = \det BA$
3. $\det A^{-1} = \dfrac{1}{\det A}$
4. $\det A^k = (\det A)^k$（如果把矩阵的逆看成-1次幂，那就是3式）
5. $\det rA = r^n \det A$

命题 5 (行列式基本行变换)**.** 三种基本行变换与矩阵相同，但替换不改变它的值，交换乘(-1)，数乘乘k.

定理 9 (拉普拉斯(Laplace)展开)**.** A为$n \times n$的矩阵，称$C_{ij} = (-1)^{i+j} \det A_{ij}$为$(i, j)$的代数余子式，其中$A_{ij}$是$A$删除第$i$行和第$j$列后剩下的矩阵.
对第i行展开有

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}$$

对第j列展开有

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij}$$

特别地，若A是三角矩阵，则$\det A =$主对角线上元素的积（因对第一行或第一列展开，形式不变）.

定理 10 (克莱姆(Cramer)法则)**.** $A\mathbf{x} = \mathbf{b}$，则$\mathbf{x}_i = \dfrac{\det A_i(\mathbf{b})}{\det A}$，其中

$$\det A_i(\mathbf{b}) = \begin{array}{c} \qquad\qquad\qquad \text{col } i \\ \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} \end{array}$$

分析. 直接构造

$$AI_i(\mathbf{b}) = A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{a}_1 & \cdots & A\mathbf{b} & \cdots & A\mathbf{a}_n \end{bmatrix}$$

左右两边取det得

$$\det AI_i(\mathbf{b}) = (\det A)(\det I_i(\mathbf{b})) = (\det A)\mathbf{x}_i = \det A_i(\mathbf{b})$$

定理 11. 若A可逆，则$A^{-1} = \dfrac{1}{\det A} \operatorname{adj} A$，其中共轭/伴随矩阵$(\operatorname{adj} A)_{ij} = C_{ji}$
注意：共轭矩阵元素下标与代数余子式下标不同.

分析. 求逆算法换个角度思考，记$B = A^{-1}$的第j列为$\mathbf{b}_j$，则$A\mathbf{b}_j = \mathbf{e}_j$，
又记第i个元素为b_{ij}，则由克莱姆法则，$b_{ij} = \dfrac{\det A_i(\mathbf{e}_j)}{\det A}$，分子

$$\det A_i(\mathbf{e}_j) = (-1)^{j+i} \det A_{ji} = C_{ji}$$

（对第i列进行展开，除了第j行为1，其余行均为0），得证.

定理 12 (行列式的线性性)**.** 记$T : \mathbb{R}^n \to \mathbb{R}$，$T(\mathbf{x}) = \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$，则$T$为线性变换.
注：这里的线性性只是针对行列式中的一列可变来说的，若整个行列式都可变，则不一定有$\det(A+B) = \det A + \det B$.

定理 13 (线性变换与体积变化). 线性变换$T:\mathbb{R}^n\to\mathbb{R}^n$的标准矩阵为$A$，$S$是一块有限的区域，那么

$$T(S)\text{的面积或体积}=|\det A|\cdot S\text{的面积或体积}$$

由此定理立得椭圆面积为πab.

定理 14 (解析几何). *1.* 相异两点$A(x_1,y_1),B(x_2,y_2)\in\mathbb{R}^2$，则直线$AB$的方程为$\det\begin{bmatrix}1&x&y\\1&x_1&y_1\\1&x_2&y_2\end{bmatrix}=0$

2. 过点$A(x_1,y_1)\in\mathbb{R}^2$，斜率为$k$的直线方程为$\det\begin{bmatrix}1&x&y\\1&x_1&y_1\\0&1&k\end{bmatrix}=0$

3. $A(x_1,y_1),B(x_2,y_2),C(x_3,y_3)\in\mathbb{R}^2$，三角形$ABC$的面积为$\dfrac{1}{2}\left|\det\begin{bmatrix}x_1&y_1&1\\x_2&y_2&1\\x_3&y_3&1\end{bmatrix}\right|$

分析. 实际上还有很多类似的结论，在此不作赘述. 均是做基本行变换后展开即得证.

定理 15 (范德蒙(Vandermonde)矩阵).

$$V=\begin{bmatrix}1&\alpha_1&\alpha_1^2&\dots&\alpha_1^{n-1}\\1&\alpha_2&\alpha_2^2&\dots&\alpha_2^{n-1}\\1&\alpha_3&\alpha_3^2&\dots&\alpha_3^{n-1}\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&\alpha_m&\alpha_m^2&\dots&\alpha_m^{n-1}\end{bmatrix}$$

令首行为变量，其他行为常量可得$n-1$次多项式，常被用于多项式插值. 当$m=n$时它的行列式

$$\det(V)=\prod_{1\le i<j\le n}(\alpha_j-\alpha_i)$$

分析. 从第$n-1$列到第1列，每一列乘上$-\alpha_n$后加到后一列：

$$\det(V_n)=\begin{vmatrix}1 & \alpha_1-\alpha_n & \alpha_1^2-\alpha_1\alpha_n & \cdots & \alpha_1^{n-2}-\alpha_1^{n-3}\alpha_n & \alpha_1^{n-1}-\alpha_1^{n-2}\alpha_n\\ 1 & \alpha_2-\alpha_n & \alpha_2^2-\alpha_2\alpha_n & \cdots & \alpha_2^{n-2}-\alpha_2^{n-3}\alpha_n & \alpha_2^{n-1}-\alpha_2^{n-2}\alpha_n\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 1 & \alpha_{n-1}-\alpha_n & \alpha_{n-1}^2-\alpha_{n-1}\alpha_n & \cdots & \alpha_1^{n-2}-\alpha_{n-1}^{n-3}\alpha_n & \alpha_{n-1}^{n-1}-\alpha_{n-1}^{n-2}\alpha_n\\ 1 & 0 & 0 & \cdots & 0 & 0\end{vmatrix}$$

$$=(-1)^{n+1}\begin{vmatrix}\alpha_1-\alpha_n & \alpha_1(\alpha_1-\alpha_n) & \cdots & \alpha_1^{n-3}(\alpha_1-\alpha_n) & \alpha_1^{n-2}(\alpha_1-\alpha_n)\\ \alpha_2-\alpha_n & \alpha_2(\alpha_2-\alpha_n) & \cdots & \alpha_2^{n-3}(\alpha_2-\alpha_n) & \alpha_2^{n-2}(\alpha_2-\alpha_n)\\ \vdots & \vdots & & \vdots & \vdots\\ \alpha_{n-1}-\alpha_n & \alpha_{n-1}(\alpha_{n-1}-\alpha_n) & \cdots & \alpha_1^{n-2}(\alpha_{n-1}-\alpha_n) & \alpha_{n-1}^{n-2}(\alpha_{n-1}-\alpha_n)\end{vmatrix} \quad \text{按最后一行展开}$$

$$=(-1)^{n+1}(\alpha_1-\alpha_n)(\alpha_2-\alpha_n)\cdots(\alpha_{n-1}-\alpha_n)\begin{vmatrix}1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-2}\\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-2}\\ 1 & \alpha_3 & \alpha_3^2 & \dots & \alpha_3^{n-2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \alpha_{n-1} & \alpha_{n-1}^2 & \dots & \alpha_{n-1}^{n-2}\end{vmatrix}$$

$$=(\alpha_n-\alpha_1)(\alpha_n-\alpha_2)\cdots(\alpha_n-\alpha_{n-1})\det(V_{n-1})=\prod_{1\le i<j\le n}(\alpha_j-\alpha_i)$$

除了懂得用拉普拉斯展开外，也要懂得通过基本行变换（或列变换）来计算行列式，如例2.

例 2.

$$A=\begin{bmatrix}1 & 1 & 1 & \dots & 1\\ 1 & 2 & 2 & \dots & 2\\ 1 & 2 & 3 & \dots & 3\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 2 & 3 & \dots & n\end{bmatrix}\Longrightarrow \det A=1$$

分析. 将每行乘(-1)后加到下一行，最后只剩对角线上元素全为1.

2 向量空间

2.1 基本定义

定义 4 (向量空间/线性空间). 给定一个域F[1]，若对一个非空集合V满足以下十条公理，则称V为F上的向量空间：

1. 定义两种运算 $+$ 和 $\cdot$ $\langle V,+,\cdot\rangle$ 加法是$V\times V$到V的一个映射，数乘是$F\times V$到V的一个映射
   注：在说明一个集合构成一个向量空间时，一定要指出域是什么，向量加法是怎么定义的，数乘是怎么定义的，否则一切都只是空中楼阁.

[1]参见https://en.wikipedia.org/wiki/Field_(mathematics)

2. 封闭性(对两种运算封闭)

3. V的基本性质，设$u,v,w\in V$，c,d为常数

   (a) 零元 $u+0=u$

   (b) 单位元 $1\cdot u=u$

   (c) 逆元 $u+(-u)=0$

   (d) 加法交换律 $u+v=v+u$ （暂不定义两向量相乘）

   (e) 结合律 $(u+v)+w=u+(v+w),(cd)u=c(du)$

   (f) 左右分配律 $c(u+v)=cu+cv,(c+d)u=cu+du$

要注意，这里的向量空间实际上是很抽象的，F,V可以是一些很奇怪的东西.

如度为n的多项式所构成的集合$\mathbb{P}_n$，有

$$\mathbf{p}(t)=a_0+a_1t+a_2t+\cdots+a_nt^n$$

的形式（这里的t虽然可以取任意值，但在$\{1,t,\ldots,t^n\}$这组标准基下区别不同多项式的是前面的系数，也即坐标向量. 故考虑其线性性时，不是考虑$\mathbf{p}(u+v)$，而是考虑$\mathbf{p}(t)+\mathbf{q}(t)$），这个集合满足以上所有公理，也被称为向量空间.

甚至于不用显式地表示出来的也可以算是向量空间，如定义在$[a,b]$上实值函数f的集合，同样是一个向量空间.

2.2 四个基本子空间

定义 5 (子空间)**.** 称H是V的子空间，当$\mathbf{u},\mathbf{v}\in H$，有：

1. $\mathbf{0}\in H$
2. $\mathbf{u}+\mathbf{v}\in H$
3. $c\mathbf{u}\in H$

注意：$\mathbb{R}^2$并不是$\mathbb{R}^3$的子空间，向量元素数目都不同，但可以通过补0来达到目的.

定义 6 (基本子空间)**.** A为$m\times n$的矩阵，则列空间$\operatorname{Col}A$、行空间$\operatorname{Row}A$、零空间$\operatorname{Nul}A$、左零空间$\operatorname{Nul}A^{\mathrm{T}}$称为$A$的基本子空间

我们先来探究A的零空间与列空间：

同时我们发现：

1. $\operatorname{Nul}A=\{\mathbf{0}\}$（$A$的列线性无关）当且仅当$\mathbf{x}\mapsto A\mathbf{x}$是一一映射/单射
   也说明A的每一列都为主元列(n个)，并且$m\geq n$，无自由变元.

2. $\operatorname{Col}A=\mathbb{R}^m$当且仅当$\mathbf{x}\mapsto A\mathbf{x}$将$\mathbb{R}^n$映上/满射到$\mathbb{R}^m$
   也说明A的每一行都有主元位置(m个)，并且$m\leq n$，**可能**存在自由变元.

需要注意：

$\mathrm{Nul}\,A$	$\mathrm{Col}\,A$
$\{\mathbf{x}:\mathbf{x}\in\mathbb{R}^n, A\mathbf{x}=\mathbf{0}\}$	$\{\mathbf{b}:\exists\,\mathbf{x}\in\mathbb{R}^n, A\mathbf{x}=\mathbf{b}\}$
隐性(implicit)定义	显性(explicit)定义
$\mathbb{R}^n$的子空间	$\mathbb{R}^m$的子空间
考虑有n个变量	考虑有m个方程
dim =方程自由变量个数	dim =主元列数
基得解出来才知道	基为**原矩阵**的主元列

1. 当A不是方阵时，$\mathrm{Nul}\,A$与$\mathrm{Col}\,A$是两个完全不同的空间；A是方阵，则有可能存在非零向量同时属于这两个空间
2. 基本行变换不会影响列之间的相关关系，但**会改变**其列空间
3. 基本行变换**不会改变**其行空间，阶梯形的**非零行**形成其行空间的一组基，也是原矩阵行空间的一组基

   分析. *B是由A经过行变换得来的矩阵，则B的行是A的行的线性组合，因此B的行的线性组合也一定是A的行的线性组合，进而$\mathrm{Row}\,B\subset\mathrm{Row}\,A$；而行操作又是可逆的，相反的操作可以说明$\mathrm{Row}\,A\subset\mathrm{Row}\,B$，因此$\mathrm{Row}\,A=\mathrm{Row}\,B$.*
   若B是阶梯形，它的非零行一定线性无关，因为没有一个非零行是它下面非零行的线性组合（定理1(c)的逆否命题），证毕.

定理 16.

$$(\mathrm{Row}\,A)^{\perp}=\mathrm{Nul}\,A,(\mathrm{Col}\,A)^{\perp}=\mathrm{Nul}\,A^{\mathrm{T}}$$

分析. *若$\mathbf{x}\in\mathrm{Nul}\,A$，由矩阵乘法，$A$每一行与$\mathbf{x}$都正交，而$A$的行张成$\mathrm{Row}\,A$，因此$\mathrm{Nul}\,A\subset(\mathrm{Row}\,A)^{\perp}$.*
若$\mathbf{x}\in(\mathrm{Row}\,A)^{\perp}$，则$\mathbf{x}$与$A$的每一行都正交，进而$\mathbf{x}\in\mathrm{Nul}\,A$，因此$(\mathrm{Row}\,A)^{\perp}\subset\mathrm{Nul}\,A$.
故$(\mathrm{Row}\,A)^{\perp}=\mathrm{Nul}\,A$. 最后令$A\leftarrow A^{\mathrm{T}}$代入即可得第二个结论.

关于四个基本子空间维度之间的关系，见定理21.

2.3 线性变换

定义 7 (线性变换[2]). *若$T:V\to W$ 满足*

1. $T(\mathbf{u}+\mathbf{v})=T(\mathbf{u})+T(\mathbf{v})$

2. $T(c\mathbf{u})=cT(\mathbf{u})$

则称T为线性变换.
注意：这里并不需要零元的定义，即$T(\mathbf{0})=\mathbf{0}\,(*)$，因线性变换保持数乘不变性，这已经可以推出 $(*)$式

线性变换研究的是两个向量空间，则更加抽象. 如对一个实值连续函数求导，也可以算是一个线性变换.

[2]有些地方说线性映射(linear mapping)是从一个向量空间V到另一个向量空间W的映射且保持加法运算和数量乘法运算，而线性变换(linear transformation)是线性空间V到其自身的线性映射. 但从国外的课本上看，这两者并没有区分得特别清楚.

定理 17 (标准(standard)矩阵). 对于线性变换$T:\mathbb{R}^n\to\mathbb{R}^m$，存在唯一矩阵$A$使得

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

其中A是一个$m\times n$的矩阵，且

$$A=\begin{bmatrix}T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n)\end{bmatrix}$$

称A为T的标准矩阵.

分析. 1° 存在性：

$$\begin{aligned}T(\mathbf{x})&=T(x_1\mathbf{e}_1+\cdots+x_n\mathbf{e}_n)\\&=x_1T(\mathbf{e}_1)+\cdots+x_nT(\mathbf{e}_n)\\&=\begin{bmatrix}T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n)\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\\&=A\mathbf{x}\end{aligned}$$

2° 唯一性：设存在另一矩阵B使得$T(\mathbf{x})=B\mathbf{x},\forall\mathbf{x}\in\mathbb{R}^n$，故$A\mathbf{x}=B\mathbf{x},\forall\mathbf{x}\in\mathbb{R}^n$，令$\mathbf{x}=\mathbf{e}_i$即得$A=B$

这里需要注意几点：

1. 由矩阵乘向量的线性性知，所有的矩阵变换都是线性变换.
2. 线性变换着重于映射的性质，而矩阵变换则描述了该映射是怎么实施的.
3. 标准矩阵的唯一性是在确定了基的情况下才说的，如上面的定理是标准基，非标准基的情况请见定理40.

定义 8 (单射与满射). 线性变换$T:V\to W$满足：

1. $\forall\mathbf{b}\in W$**至多**可以在V中找到一个原像$\mathbf{x}$，则称T单射（即$T(\mathbf{u})=T(\mathbf{v})\implies\mathbf{u}=\mathbf{v}$）
2. $\forall\mathbf{b}\in W$**至少**可以在V中找到一个原像$\mathbf{x}$，则称T满射
3. $\forall\mathbf{b}\in W$在V中都**有且只有**一个原像$\mathbf{x}$，则称T双射/一一映射，称V和W**同构**，记为$V\cong W$.

定义 9 (核空间与值域空间). $T(\mathbf{x})=\mathbf{0}$的解为T的核空间，$T(\mathbf{x}),\forall\mathbf{x}$的所有可能值构成$T$的值域空间，分别对应着矩阵中的零空间和列空间.

定理 18. 线性变换$T:V\to W$，$S=\{\mathbf{v}_1,\dots,\mathbf{v}_p\}\in V$，$S'=\{T(\mathbf{v}_1),\dots,T(\mathbf{v}_p)\}\in W$，$U$是$V$的子空间

1. 若S线性相关，则S'线性相关；若S'线性无关，则S线性无关
2. T是单射，若S'线性相关，则S线性相关；若S线性无关，则S'线性无关
3. 若S为U的一组基，则S'张成$T(U)$
4. $V\cong W$，若S是V的线性无关组（生成集/基），则S'是W的线性无关组（生成集/基）
5. $V\cong W\iff\dim V=\dim W$
6. $T(U)$是W的子空间

7. $\dim T(U) \le \dim U$，当T为单射时取等

8. **核是定义域的子空间，值域空间是对映域的子空间**

分析. 1. 不妨设$\mathbf{v}_p = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1}$，则$T(\mathbf{v}_p) = T(c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1}) = c_1T(\mathbf{v}_1) + \cdots + c_{p-1}T(\mathbf{v}_{p-1})$，说明$S'$也线性相关. 后者为前者的逆否命题，易知成立.

2. 不妨设$T(\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_{p-1}T(\mathbf{v}_{p-1})$，则$c_1T(\mathbf{v}_1) + \cdots + c_{p-1}T(\mathbf{v}_{p-1}) = T(c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1}) = T(\mathbf{v}_p)$. 又因$T$是单射，故$\mathbf{v}_p = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1}$，$S$线性相关.

3. $\mathbf{y} \in T(U), \exists T(\mathbf{x}) = \mathbf{y}$，因$\mathbf{x} \in U$，所以$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$，因$T$为线性，故$\mathbf{y} = T(\mathbf{x}) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p)$，即可说明.（注意这里只能说明$T(U)$的基在$S'$内，而不能说明$S'$就是基）

4. 结合1,2,3即得证.

5. 结合4，S为V的基，S'为W的基，都含有p个向量，故维度相同.

6. (a) $\mathbf{0} \in U, T(\mathbf{0}) = \mathbf{0} \in T(U)$

   (b) $\forall \mathbf{v}_1, \mathbf{v}_2 \in T(U), \exists\, \mathbf{u}_1, \mathbf{u}_2 \in U\; s.t.\; \mathbf{v}_1 = T(\mathbf{u}_1),\, \mathbf{v}_2 = T(\mathbf{u}_2),$
   $\because \mathbf{u}_1 + \mathbf{u}_2 \in U \quad \therefore T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2) = \mathbf{v}_1 + \mathbf{v}_2 \in T(U)$

   (c) $\forall \mathbf{v} \in T(U), \exists\, \mathbf{u} \in U\; s.t.\; \mathbf{v} = T(\mathbf{u}),$
   $\because c\mathbf{u} \in U \quad \therefore T(c\mathbf{u}) = cT(\mathbf{u}) = c\mathbf{v} \in T(U)$

7. 由定理20(d)知$\dim T(U) \le \dim U$. 若S为U的一组基，则S'张成$T(U)$，又本题的题2证明了S'线性无关，故S'为$T(U)$的一组基，故$\dim T(U) = p = \dim U$

8. 记T的核为$\ker T$，显然$\mathbf{0} \in \ker T$. 若$\mathbf{u}, \mathbf{v} \in \ker T$，即$T(\mathbf{u}) = \mathbf{0}, T(\mathbf{v}) = \mathbf{0}$，则$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) = \mathbf{0} \implies \mathbf{u} + \mathbf{v} \in \ker T, T(c\mathbf{u}) = cT(\mathbf{u}) = \mathbf{0} \implies c\mathbf{u} \in \ker T$，故核是定义域的子空间.
   上面6中令$U \leftarrow V$，即得值域空间是对映域的子空间.

2.4 基、维度与秩

定义 10 (基). H是向量空间V的子空间，$\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$满足以下两个条件:

1. $\mathcal{B}$为线性无关集

2. $\mathrm{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_p\} = H$

则称$\mathcal{B}$是H的基. 换言之，基是**极大线性无关组**，也是**极小生成集**.

由条件2知$\mathbf{b}_1, \ldots, \mathbf{b}_p$都得在$H$中，这个显然的结论看似没用，但却是$\mathcal{B}$成为基的重要条件（在后文证明施密特正交化（算法4）时也需要说明这一点）.

如$\mathbf{v}_1 = (1,0,1), \mathbf{v}_2 = (0,1,1), \mathbf{v}_3 = (0,1,0)$，令$H = \{(s,t,t) : s,t \in \mathbb{R}\}$，有

$$\begin{bmatrix} s \\ t \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (t-s)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

这里$\mathbf{v}_1, \mathbf{v}_3 \notin H$，其实$H$只是$\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$的子集，故$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$不能称为一组基.

定理 19 (生成集(Spanning set))**.** $S=\{\mathbf{v}_1,\dots,\mathbf{v}_p\}$是$V$中的集合，令$H=\operatorname{Span}\{\mathbf{v}_1,\dots,\mathbf{v}_p\}$，

(*a*) 若$\mathbf{v}_k$是S中其余向量的线性组合，则将$\mathbf{v}_k$移除，S仍然张成H

(*b*) 若$H\neq\{\mathbf{0}\}$，则S的某个子集是H的基

分析. 不妨设$\mathbf{v}_p$可以被其他向量线性表示，则

$$\mathbf{v}_p=a_1\mathbf{v}_1+\cdots+a_{p-1}\mathbf{v}_{p-1}$$

将上式代入下式，可知将$\mathbf{v}_p$移除后，H中的向量$\mathbf{x}$依然可以被其余$p-1$个向量线性表示，即S仍然张成H

$$\mathbf{x}=c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p=(c_1+a_1c_p)\mathbf{v}_1+\cdots+(c_{p-1}+a_{p-1}c_p)\mathbf{v}_{p-1}$$

不断重复(a)的步骤，删除线性相关的向量，那么剩下的向量一定都线性无关，且可张成H，故形成一组基

定义 11 (维度)**.** 向量空间V的一组基中向量的个数定义为V的维度，记为$\dim V$

定义 12 (秩)**.** 对于矩阵A，它的秩记为

$$\operatorname{rank}A=\dim\operatorname{Col}A$$

一般地，对于一组向量集合，它的秩是极大线性无关组中向量的数目

定理 20. 以下是关于基和维度的一些结论. 设$S=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}\in V$, $|S|\geq 1$，且S为V的基

(*a*) 在V中任取$p(p>n)$个向量，一定线性相关

(*b*) 在V中任取$p(p<n)$个向量，一定不能张成V

(*c*) V的所有基都包含n个向量

分析. 由假设知$\dim V=n$，先将S与这p个向量改写成坐标向量后（见*2.5*节），构造$n\times p$矩阵A，A的每一列即为这p个向量. 由*2.2*节的讨论知，若$n<p$，则A一定存在自由变元，A的列一定线性相关；若$n>p$，则不能保证每一行都有主元位置，进而不能张成$\mathbb{R}^n$，对应地也就不能张成V. 因此，V的所有基都包含n个向量. 直接推论是：每一个$\mathbb{R}^n$的基都必须包含n个向量.

(*d*) H是V的一个子空间，则H内任一线性无关集都可以被拓展为H的一组基，且$\dim H\leq\dim V$（基的存在性）

分析. 若$H=\{\mathbf{0}\}$，显然$\dim H=0\leq\dim V$；
否则，令$S'=\{\mathbf{u}_1,\dots,\mathbf{u}_k\}$是$H$的一个线性无关集. 若$S'$张成$H$，则$S'$是$H$的一组基；否则，$\exists\mathbf{u}_{k+1}\in H-\operatorname{Span}S'$，则$\{\mathbf{u}_1,\dots,\mathbf{u}_k,\mathbf{u}_{k+1}\}$将线性无关. 继续这个过程，可以不断扩展$S'$直至其能张成$H$，也即成为$H$的一组基. 但由*((a))*，$S'$中向量的数量不能超过$\dim V$. 综上，$\dim H\leq\dim V$.

(*e*) 任一p个元素的线性无关集或是可张成V的集合自动成为V的基

分析. 由*((d))*，有p个元素的线性无关集S'可以被拓展成V的基，但这组基一定得含p个元素，因为$\dim V = p$，因此S'已经是V的基. 假设S''有p个元素且可张成V，因为$V \neq \{\mathbf{0}\}$，生成集定理告诉我们S''的某一个子集S^*一定是V的基，又因$\dim V = p$，所以S^*一定包含p个向量，也即$S^* = S''$.

(*f*) H是V的n维子空间，则$H = V$

分析. 若$\dim V = \dim H = 0$，则$V = H = \{\mathbf{0}\}$;
若$\dim V = \dim H > 0$，则H存在一组基S'包含n个向量，那么由*((e))*，S'也是V的基，因此$H = V = \operatorname{Span} S'$.

(*g*) $\mathbb{P}$是无限维空间，$C(\mathbb{R})$为所有实值连续函数构成的空间，也是无限维的

分析. 反证法，假设$\dim \mathbb{P} = k < \infty$，$\mathbb{P}_n$是$\mathbb{P}$的子空间，$\dim \mathbb{P}_{k-1} = k$，因此$\dim \mathbb{P}_{k-1} = \dim \mathbb{P}$，进而$\mathbb{P}_{k-1} = \mathbb{P}$，这显然不对，如$\mathbf{p}(t) = t^k$在$\mathbb{P}$内，但不在$\mathbb{P}_{k-1}$内，矛盾.
因$\mathbb{P}$是$C(\mathbb{R})$的子空间，若$C(\mathbb{R})$为有限维，则由*((d))*，$\mathbb{P}$也是有限维的，矛盾.

定理 21 (秩定理)**.**

$$\operatorname{rank} A + \dim \operatorname{Nul} A = n$$

$$\dim \operatorname{Row} A + \dim \operatorname{Nul} A^{\mathrm{T}} = m$$

分析. 由前面*2.2*节的讨论知道，列空间的维度为主元列数，加上零空间的维度为自由变量个数就等于总的列数.

例 3. 若一个非齐次线性方程组有6个方程8个未知数，已知其有一个有2个自由变元的解，则无论方程右侧的常数是什么，该方程组总有解.

分析. $\dim \operatorname{Nul} A = 2, \operatorname{rank} = 8 - 2 = 6, \operatorname{Col} A \in \mathbb{R}^6 \implies \operatorname{Col} A = \mathbb{R}^6$

例 4. $A\mathbf{x} = \mathbf{b}, \forall \mathbf{b} \in \mathbb{R}^m$都有解 $\iff A^{\mathrm{T}}\mathbf{x} = \mathbf{0}$只有平凡解

分析. $\operatorname{rank} A = \dim \operatorname{Col} A = \dim \operatorname{Row} A = m \implies \dim \operatorname{Nul} A^{\mathrm{T}} = 0$

定理 22 (秩的进阶定理)**.** A是$m \times n$矩阵，B是$n \times p$矩阵，记$\operatorname{rank} A = r = r_1$，$\operatorname{rank} B = r_2$，有以下结论

(*a*) $\operatorname{rank} AB \leq \operatorname{rank} A, \operatorname{rank} AB \leq \operatorname{rank} B$

分析. 设$\mathbf{y} \in \operatorname{Col} AB$，则$\exists \mathbf{x}\, s.t.\, \mathbf{y} = AB\mathbf{x}$，而$AB\mathbf{x} = A(B\mathbf{x}) = \mathbf{y}$，因此$\mathbf{y} \in \operatorname{Col} A$，故$\operatorname{Col} AB$为$\operatorname{Col} A$的子空间，由定理*20((d))*知$\operatorname{rank} AB \leq \operatorname{rank} A$.
又由秩定理，$\operatorname{rank} AB = \operatorname{rank}(AB)^{\mathrm{T}} = \operatorname{rank} B^{\mathrm{T}} A^{\mathrm{T}} \leq \operatorname{rank} B^{\mathrm{T}} = \operatorname{rank} B$.

(*b*) P是$m \times m$可逆矩阵，Q是$n \times n$可逆矩阵，则$\operatorname{rank} PA = \operatorname{rank} AQ = \operatorname{rank} A$

分析. 由(a)有$\operatorname{rank} PA \leq \operatorname{rank} A = \operatorname{rank}(P^{-1}P)A = \operatorname{rank} P^{-1}(PA) \leq PA \implies \operatorname{rank} PA = \operatorname{rank} A$.
$\operatorname{rank} AQ = \operatorname{rank}(AQ)^{\mathrm{T}} = \operatorname{rank} Q^{\mathrm{T}} A^{\mathrm{T}} = \operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} A$.

(*c*) 若$AB = O$，则$\operatorname{rank} A + \operatorname{rank} B \leq n$

分析. 由$AB=O$知，B的每一列都在$\operatorname{Nul} A$中，故$\operatorname{Col} B$是$\operatorname{Nul} A$的子空间，故$\operatorname{rank} B \le \dim \operatorname{Nul} A$.
又由秩定理$n = \operatorname{rank} A + \dim \operatorname{Nul} A \ge \operatorname{rank} A + \operatorname{rank} B$.

(*d*) $\operatorname{rank}(A+B) \le \operatorname{rank} A + \operatorname{rank} B$

分析. 引理：一定存在秩分解$A=CR$，其中C是$m\times r$矩阵，R是$r\times n$矩阵.
将A,B秩分解得$A=C_1R_1, B=C_2R_2$，构造$m\times(r_1+r_2)$矩阵$C=\begin{bmatrix} C_1 & C_2 \end{bmatrix}$，则

$$A+B = C_1R_1 + C_2R_2 = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = CR$$

因C有r_1+r_2列，故$\operatorname{rank} C \le r_1+r_2$，同理$R$有$r_1+r_2$行，故$\operatorname{rank} R \le r_1+r_2$，进而$\operatorname{rank}(A+B) \le r_1+r_2 = \operatorname{rank} A + \operatorname{rank} B$.

(*e*) A的秩为r当且仅当A包含一个$r\times r$的子矩阵，且没有更大的方阵是可逆的

分析. 1° 证明A一定有$m\times r$且秩为r的子矩阵A_1
令A_1包含A的r个主元列，因这些列线性无关，故A_1即为所求.
2° 证明A_1一定有$r\times r$且可逆的子矩阵A_2
$\operatorname{rank} A_1 = \dim \operatorname{Row} A_1 = r$，令$A_2$包含$A_1$的$r$个线性无关的行，则$A_2$即为所求，且为方阵故可逆.

(*f*) 若$C=\begin{bmatrix} A & O \\ O & B \end{bmatrix}$，则$\operatorname{rank} C = \operatorname{rank} A + \operatorname{rank} B$

分析. 将A,B都写成阶梯形，因分块矩阵其余位置都为0，故C此时也为阶梯形，C的秩就等于A,B主元列之和.

2.5 坐标系统

定理 23 (向量唯一表示定理)**.** $\mathcal{B}=\{b_1,\cdots,b_p\}$是向量空间$V$的基，则$\forall \mathbf{x}\in V$，存在唯一$c_1,\ldots,c_n$使得$\mathbf{x}=c_1\mathbf{b}_1+\cdots+c_p\mathbf{b}_p$

分析.

$$0=\mathbf{x}-\mathbf{x}=(c_1-d_1)\mathbf{b}_1+\cdots+(c_p-d_p)\mathbf{b}_p$$

由基线性无关，知上述方程只有平凡解.

定义 13 (坐标向量)**.** $\mathcal{B}=\{b_1,\cdots,b_p\}$是$H$的基，$\forall \mathbf{x}\in H$, $\mathbf{x}=c_1\mathbf{b}_1+\cdots+c_p\mathbf{b}_p$，则$\mathbf{x}$关于$\mathcal{B}$的坐标向量定义为

$$[\mathbf{x}]_{\mathcal{B}} := \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

坐标系统最大的用处是可以将向量空间V中奇奇怪怪的东西转换成$\mathbb{R}^n$中我们熟悉的向量，求解坐标的过程实质上又是解线性方程组，进而有如下定理.

定理 24 (坐标变换矩阵). 对于V的一组基$\mathcal{B}=\{\mathbf{b}_1,\cdots,\mathbf{b}_n\}$，存在唯一$n\times n$矩阵$P_{\mathcal{B}}$使得

$$\mathbf{x}=P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}},\forall\mathbf{x}\in V$$

其中$P_{\mathcal{B}}=\begin{bmatrix}\mathbf{b}_1,\ldots,\mathbf{b}_n\end{bmatrix}$称为坐标变换矩阵. 移项即可解出坐标,

$$[\mathbf{x}]_{\mathcal{B}}=P_{\mathcal{B}}^{-1}\mathbf{x}$$

定理 25 (换基). $\mathcal{B}=\{\mathbf{b}_1,\ldots,\mathbf{b}_n\},\mathcal{C}=\{\mathbf{c}_1,\ldots,\mathbf{c}_n\}$都是向量空间$V$的基，那么存在唯一的$n\times n$矩阵$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$,使得

$$[\mathbf{x}]_{\mathcal{C}}=\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{x}]_{\mathcal{B}}$$

其中$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}=\begin{bmatrix}[\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}}\end{bmatrix}$，称为过渡矩阵.

算法 3 (在$\mathbb{R}^n$中换基). 要求$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$，首先要将$\mathcal{B}$与$\mathcal{C}$之间的关系找到，即用$\mathcal{C}$中的向量表示$\mathcal{B}$中的向量.
设坐标变换矩阵$P_{\mathcal{C}}=\begin{bmatrix}\mathbf{c}_1,\ldots,\mathbf{c}_n\end{bmatrix}$，进而$\mathbf{b}_i=P_{\mathcal{C}}[\mathbf{b}_i]_{\mathcal{C}}$，故

$$P_{\mathcal{B}}=P_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$$

（注意矩阵下标）由命题*1*，可得以下过程

$$\begin{bmatrix}P_{\mathcal{C}} & P_{\mathcal{B}}\end{bmatrix}\sim\begin{bmatrix}I & \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\end{bmatrix}$$

或者$P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}=\mathbf{x}=P_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}}\implies[\mathbf{x}]_{\mathcal{C}}=P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}\implies\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}=P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$.

定理 26. 关于坐标变换$T:\mathbf{x}\mapsto[\mathbf{x}]_{\mathcal{B}}$，其中$\mathcal{B}=\{\mathbf{b}_1,\ldots,\mathbf{b}_n\}$，有如下结论.

1. T是一个双射的线性变换，$\mathbf{x}$所处的空间V与$[\mathbf{x}]_{\mathcal{B}}$所处的空间$\mathbb{R}^n$同构
2. $S=\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$线性无关当且仅当$S'=\{[\mathbf{u}_1]_{\mathcal{B}},\ldots,[\mathbf{u}_p]_{\mathcal{B}}\}$线性无关；$S$线性相关当且仅当$S'$线性相关

分析.

1. 若$[\mathbf{u}]_{\mathcal{B}}=[\mathbf{w}]_{\mathcal{B}}=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$，由坐标向量定义$\mathbf{u}=\mathbf{w}=c_1\mathbf{b}_1+\cdots+c_n\mathbf{b}_n$，而$\mathbf{u},\mathbf{w}$均为$V$中任意元素，故$T$为单射. 令$\mathbf{y}=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$，$\mathbf{u}=y_1\mathbf{b}_1+\cdots+y_n\mathbf{b}_n$，则由定义$[\mathbf{u}]_{\mathcal{B}}=\mathbf{y}$，因$\mathbf{y}$为$\mathbb{R}^n$任意向量，故$T$为满射.
2. 由定理*18*易知.

3 内积空间

3.1 基本定义

定义 14 (内积空间). 线性空间V上的内积是一个函数$\langle\cdot,\cdot\rangle : V\times V\to\mathbb{R}$，$\forall u,v,w\in V, c$为常数，满足

1. $\langle\mathbf{u},\mathbf{v}\rangle=\langle\mathbf{v},\mathbf{u}\rangle$
2. $\langle\mathbf{u}+\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{u},\mathbf{w}\rangle+\langle\mathbf{v},\mathbf{w}\rangle$
3. $\langle c\mathbf{u},\mathbf{v}\rangle=c\langle\mathbf{u},\mathbf{v}\rangle$
4. $\langle\mathbf{u},\mathbf{u}\rangle\geq 0$，且$\langle\mathbf{u},\mathbf{u}\rangle=0$当且仅当$u=0$

定义了内积的向量空间称为内积空间.

定义 15 (长度/模/范数). V是一个内积空间，则对于某一向量$v\in V$的长度为$\|v\|:=\sqrt{\langle v,v\rangle}$.

注：如无特殊说明，下文在谈论到内积时都是指最简单不加权的内积，即$\langle\mathbf{u},\mathbf{v}\rangle=\mathbf{u}\cdot\mathbf{v}=\mathbf{u}^{\mathrm{T}}\mathbf{v}$.

3.2 正交

定义 16 (正交). 若$\langle u,v\rangle=0$，则称u,v正交.
注：与高中的定义不同，这里我们可以说$\mathbf{0}$与任意向量正交.

下面这些定理看似简单，因为有二维三维空间的几何直观作为基础，但拓展到高维空间这些定理还是否会成立呢？我们不得而知，因此还应从公理出发一一证明.

定理 27 (毕达哥拉斯(Pythagoras)定理). u,v正交当且仅当$\|\mathbf{u}\|^2+\|\mathbf{v}\|^2=\|\mathbf{u}+\mathbf{v}\|^2$

定义 17. 设W是向量空间V的子空间，V中与所有与W正交的向量构成的集合称为W的正交补，记为$W^{\perp}$.

以下有两条性质

1. $\mathbf{x}\in W^{\perp}\iff\mathbf{x}$与张成$V$的一组向量中的每一个都正交
2. $W^{\perp}$是V的子空间

定理 28 (正交必定线性无关). $S=\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$是$V$的一个由**非零**向量构成的正交集，那么$S$线性无关，因此构成Span S的基.
注：一定要注意前提条件!!!

分析.

$$\mathbf{0}=c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_p\mathbf{u}_p$$

$$\begin{aligned}0=\langle\mathbf{0},\mathbf{u}_1\rangle&=\langle c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_p\mathbf{u}_p,\mathbf{u}_1\rangle\\&=\langle c_1\mathbf{u}_1,\mathbf{u}_1\rangle+\langle c_2\mathbf{u}_2,\mathbf{u}_1\rangle+\cdots+\langle c_p\mathbf{u}_p,\mathbf{u}_1\rangle\\&=c_1\langle\mathbf{u}_1,\mathbf{u}_1\rangle+c_2\langle\mathbf{u}_2,\mathbf{u}_1\rangle+\cdots+c_p\langle\mathbf{u}_p,\mathbf{u}_1\rangle\\&=c_1\langle\mathbf{u}_1,\mathbf{u}_1\rangle\end{aligned}$$

因为$\mathbf{u}_1\neq\mathbf{0}$，所以$c_1=0$. 类似地，可以推出$c_i=0, i=2,\ldots,p$，因此$S$线性无关.

定理 29 (正交分解). $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$是$V$的子空间$W$的一组正交基，对于$\mathbf{y}\in V$，都可以唯一表示成$\mathbf{y}=\hat{\mathbf{y}}+\mathbf{z}$，其中$\hat{\mathbf{y}}\in W,\mathbf{z}\in W^{\perp}$，且

$$\hat{\mathbf{y}}=c_1\mathbf{u}_1+\cdots+c_p\mathbf{u}_p$$

各项为

$$\hat{\mathbf{y}}_i=\operatorname{proj}_{u_i}\hat{\mathbf{y}}=c_i\mathbf{u}_i=\frac{\langle\hat{\mathbf{y}},\mathbf{u}_i\rangle}{\langle\mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i=\frac{\langle\hat{\mathbf{y}},\mathbf{u}_i\rangle}{\|\mathbf{u}_i\|}\frac{\mathbf{u}_i}{\|\mathbf{u}_i\|},i=1,\ldots,p\qquad(*)$$

注：与选的基没有关系，只和W有关.

分析. 类似定理*28*的证明有$\langle\hat{\mathbf{y}},\mathbf{u}_i\rangle=\langle c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_p\mathbf{u}_p,\mathbf{u}_i\rangle=c_i\langle\mathbf{u}_i,\mathbf{u}_i\rangle$，移项即得$(*)$式.

$$\begin{aligned}\langle\mathbf{z},\mathbf{u}_1\rangle=\langle\mathbf{y}-\hat{\mathbf{y}},\mathbf{u}_1\rangle&=\langle\mathbf{y},\mathbf{u}_1\rangle-\langle\frac{\langle\hat{\mathbf{y}},\mathbf{u}_1\rangle}{\langle\mathbf{u}_1,\mathbf{u}_1\rangle},\mathbf{u}_1\rangle\mathbf{u}_1-0-\cdots-0\\&=\langle\mathbf{y},\mathbf{u}_1\rangle-\langle\mathbf{y},\mathbf{u}_1\rangle=0\end{aligned}$$

类似地可证明$\mathbf{z}$与W的每一个基$\mathbf{u}_i$都正交，进而$\mathbf{z}\in W^{\perp}$.
为证明唯一性，不妨设存在另一表示方法$\mathbf{y}=\hat{\mathbf{y}}'+\mathbf{z}'$，其中$\hat{\mathbf{y}}'\in W$，$\mathbf{z}'\in W^{\perp}$那么$\hat{\mathbf{y}}+\mathbf{z}=\hat{\mathbf{y}}'+\mathbf{z}'$，即$\hat{\mathbf{y}}-\hat{\mathbf{y}}'=\mathbf{z}'-\mathbf{z}$. 因为$\mathbf{z},\mathbf{z}'\in W^{\perp}$，所以$\mathbf{z}'-\mathbf{z}\in W^{\perp}$，故$\hat{\mathbf{y}}-\hat{\mathbf{y}}'\in W^{\perp}$，而$\hat{\mathbf{y}}-\hat{\mathbf{y}}'\in W$，推出$\langle\hat{\mathbf{y}}-\hat{\mathbf{y}}',\hat{\mathbf{y}}-\hat{\mathbf{y}}'\rangle=0$，即$\hat{\mathbf{y}}=\hat{\mathbf{y}}'$，同时$\mathbf{z}'=\mathbf{z}$.

从$(*)$的最后一个等式可以看出，实际上正交分解与我们高中所学的向量知识是一致的，即$\mathbf{y}$在$\mathbf{u}_i$上的投影的长度再乘上该方向上的单位向量.

定理 30 (最佳估计). W为V的子空间，$\mathbf{y}\in V$，则

$$\|\mathbf{y}-\hat{\mathbf{y}}\|\le\|\mathbf{y}-\mathbf{v}\|,\forall v\in W$$

分析.

$$\mathbf{y}-\mathbf{v}=(\mathbf{y}-\hat{\mathbf{y}})+(\hat{\mathbf{y}}-\mathbf{v})$$

因为$\hat{\mathbf{y}}-\mathbf{v}\in W,\mathbf{y}-\hat{\mathbf{y}}\in W^{\perp}$，所以由毕达哥拉斯定理

$$\|\mathbf{y}-\mathbf{v}\|^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\mathbf{v}\|^2$$

即得结论.

定理 31 (柯西(Cauchy)不等式).

$$\|\mathbf{u}\|\|\mathbf{v}\|\ge|\langle\mathbf{u},\mathbf{v}\rangle|$$

分析.

$$\|\operatorname{proj}_{\mathbf{u}}\mathbf{v}\|=\frac{\langle\mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|}\le\|\mathbf{v}\|$$

定理 32 (三角不等式).

$$\|\mathbf{u}\|+\|\mathbf{v}\|\ge\|\mathbf{u}+\mathbf{v}\|$$

分析. 平方后用柯西不等式.

命题 6. A是$m\times n$矩阵，$A\mathbf{x}=\mathbf{0}$当且仅当$A^{\mathrm{T}}A\mathbf{x}=\mathbf{0}$，进而$\operatorname{Nul}A=\operatorname{Nul}A^{\mathrm{T}}A$.
$A^{\mathrm{T}}A$可逆当且仅当A的列线性无关. $\operatorname{rank}A^{\mathrm{T}}A=n-\dim\operatorname{Nul}A^{\mathrm{T}}A=\operatorname{Nul}A=\operatorname{rank}A$

分析. $\mathbf{x}^{\mathrm{T}}A^{\mathrm{T}}A\mathbf{x}=\mathbf{0}$

3.3 正交矩阵

定理 33. U是$m\times n$矩阵，则U每一列都**单位正交**当且仅当$U^{\mathrm{T}}U=I$

分析.

$$U^{\mathrm{T}}U=\begin{bmatrix}\mathbf{u}_1^{\mathrm{T}}\\ \vdots\\ \mathbf{u}_n^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}=\begin{bmatrix}\mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1 & \cdots & \mathbf{u}_1^{\mathrm{T}}\mathbf{u}_n\\ \vdots & \ddots & \vdots\\ \mathbf{u}_n^{\mathrm{T}}\mathbf{u}_1 & \cdots & \mathbf{u}_n^{\mathrm{T}}\mathbf{u}_n\end{bmatrix}=\begin{bmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{bmatrix}=I$$

定理 34. U是$m\times n$矩阵且每一列都单位正交，$x,y\in\mathbb{R}^n$

1. $\|U\mathbf{x}\|=\|\mathbf{x}\|$
2. $(U\mathbf{x})\cdot(U\mathbf{y})=\mathbf{x}\cdot\mathbf{y}$
3. $(U\mathbf{x})\cdot(U\mathbf{y})=0 \iff \mathbf{x}\cdot\mathbf{y}=0$

分析. $(U\mathbf{x})\cdot(U\mathbf{y})=(U\mathbf{x})^{\mathrm{T}}(U\mathbf{y})=\mathbf{x}^{\mathrm{T}}U^{\mathrm{T}}U\mathbf{y}=\mathbf{x}\cdot\mathbf{y}$， 1,3同理可证.

定义 18 (正交矩阵)**.** *若U为**方阵**且每一列都**单位正交**，则称U为正交矩阵.*
注：从定义上看似乎应该叫单位正交矩阵(orthonormal matrix)比较合适，但实际上在线性代数中正交矩阵(orthogonal matrix)已经包含了单位正交的意思.

由以上定理和定义可知下面几条显然的结论.

命题 7.
1. *U为方阵则一定可逆，因$U^{\mathrm{T}}U=I=U^{-1}U=I$*
2. *U,V均为正交矩阵，则UV也为正交矩阵*
3. *U中的列交换一下得到V，则V也为正交矩阵*

命题 8. *$\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$是$V$的子空间$W$的一组单位正交基，则$\operatorname{proj}_W\mathbf{y}=UU^{\mathrm{T}}\mathbf{y},\forall\mathbf{y}\in V$*

算法 4 (施密特(Schmidt)正交化)**.** *设$\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$是$W$的一组基，定义*

$$\begin{aligned}\mathbf{u}_1&=\mathbf{v}_1\\ \mathbf{u}_2&=\mathbf{v}_2-\operatorname{proj}_{\mathbf{u}_1}\mathbf{v}_2\\ &\vdots\\ \mathbf{u}_p&=\mathbf{v}_p-\sum_{j=1}^{p-1}\operatorname{proj}_{\mathbf{u}_j}\mathbf{v}_p\end{aligned}$$

则$\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$是$W$的一组正交基，且

$$\operatorname{Span}\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}=\operatorname{Span}\{\mathbf{u}_1,\ldots,\mathbf{u}_k\},\forall\, 1\le k\le p$$

注：此过程说明了结合向量空间基的存在性可说明正交基的存在性.

分析. 设$V_k = \mathrm{Span}\{\mathbf{v}_1,\dots,\mathbf{v}_k\}, W_k = \mathrm{Span}\{\mathbf{u}_1,\dots,\mathbf{u}_k\}, 1 \le k \le p$，则$\mathbf{u}_{k+1} = \mathbf{v}_{k+1} - \mathrm{proj}_{W_k}\mathbf{v}_{k+1}$，进而$\mathbf{u}_{k+1} \in W_{k+1}^{\perp}$，即每一$\mathbf{u}_i$均与前面的$\mathbf{u}_j, 1 \le j \le i-1$正交（也就保证了两两正交）. 而$\mathbf{v}_i \in V_i$，由向量加减法的封闭性，$\mathbf{u}_i \in V_i$，故$\mathbf{u}_i \in V_p$（因$V_k \subset V_{k+1}$）. 又正交一定线性无关，且$V$已有一组$p$个向量的基，故$\{\mathbf{u}_1,\dots,\mathbf{u}_k\}$是$V_k$的一组正交基，进而$V_k = W_k$.

定理 35 (QR分解). *若$m \times n$矩阵A的***列线性无关***，则A可被分解为$A = QR$，其中Q是$m \times n$矩阵且它的列形成$\mathrm{Col}\,A$的***单位正交基***，R是$n \times n$可逆上三角矩阵且对角线上元素全为正数.*
注：在下面的证明可以看出$A = QR$，单位正交不是必须的，但可能因为具有某种性质，所以才要强调单位正交.

分析. 施密特正交化将A的列转为正交基$\{\mathbf{u}_1,\dots,\mathbf{u}_n\}$，则$Q = \begin{bmatrix}\mathbf{u}_1,\dots,\mathbf{u}_n\end{bmatrix}$.
由算法4的分析知$\mathbf{x}_k \in V_k = W_k$，故存在$r_{k1},\dots,r_{kk}$，使得

$$\mathbf{x}_k = r_{k1}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + \cdots + 0 \cdot \mathbf{u}_n$$

假设$r_{kk} \ge 0$（否则给r_{kk}和$\mathbf{u}_k$同乘-1），令

$$\mathbf{r}_k = \begin{bmatrix} r_{k1} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

为R的列，显然R为上三角，即可满足$\mathbf{x}_k = Q\mathbf{r}_k$.

算法 5 (QR分解). *1. 先将A的列向量单位正交化，得到Q*
2. 通过$R = IR = Q^{\mathrm{T}}QR = Q^{\mathrm{T}}A$计算得$R$

4 特征值与特征向量

4.1 基本定义

定义 19. A为$n \times n$的矩阵，若$\exists \lambda, \mathbf{x} \ne \mathbf{0}\ s.t.\ A\mathbf{x} = \lambda\mathbf{x}$，则

1. λ称为A的**特征值**，$\mathbf{x}$称为关于λ的**特征向量**
2. $\mathrm{Nul}(A - \lambda I)$称为$A$关于$\lambda$的**特征空间**
3. $\det(A - \lambda I)$称为A的**特征多项式**
4. λ的**代数重数**是它在特征多项式中作为根的重数，**几何重数**是它对应的特征空间的维度

需要注意：

1. 特征向量不能为$\mathbf{0}$，但特征值可以为0
2. A不可逆，则0是一个特征值（充要条件）；$A-\lambda I$不可逆，A才有特征值

定理 36. $\lambda_1,\cdots,\lambda_r$是$n\times n$矩阵$A$不同的特征值，则其对应的特征向量$\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$线性无关.

分析. 反证，假设$\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$线性相关.令$p$是最小的下标使得$\mathbf{v}_{p+1}$是之前向量的线性组合，即

$$c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p=\mathbf{v}_{p+1} \tag{1}$$

左右同乘A，得

$$c_1A\mathbf{v}_1+\cdots+c_pA\mathbf{v}_p=A\mathbf{v}_{p+1}$$

进而

$$c_1\lambda_1\mathbf{v}_1+\cdots+c_p\lambda_p\mathbf{v}_p=\lambda_{p+1}\mathbf{v}_{p+1} \tag{2}$$

将*(1)*式同乘λ_{p+1}，并与*(2)*式相减得

$$c_1(\lambda_1-\lambda_{p+1})\mathbf{v}_1+\cdots+c_p(\lambda_p-\lambda_{p+1})\mathbf{v}_p=\mathbf{0}$$

由于$\{\mathbf{v}_1,\cdots,\mathbf{v}_p\}$线性无关，则它们的权重$c_i(\lambda_i-\lambda_{p+1})=0,\ i=1,\ldots,p$.而$A$的特征值均不同，故$\lambda_i-\lambda_{p+1}\neq 0$，因而$c_i=0,\ i=1,\ldots,p$.
但由*(1)*又推出$\mathbf{v}_{p+1}=\mathbf{0}$，与特征向量的定义矛盾，故假设不成立.

命题 9. 特征值与矩阵运算

1. λ为可逆矩阵A的特征值，则λ^{-1}为A^{-1}的特征值
2. λ是A的特征值，当且仅当λ是A^T的特征值

分析. 前者给$A\mathbf{x}=\lambda\mathbf{x}$左乘$A^{-1}$后显然. 后者$\det(A-\lambda I)=\det(A-\lambda I)^T=\det(A^T-\lambda I)$特征多项式相同

定义 20 (相似性)**.** 矩阵A,B满足$A=P^{-1}BP$，则称A,B相似.
注：相似性不同于行等价，行变换**会改变**矩阵的特征值，但相似性不会变

4.2 对角化

定理 37. $n\times n$矩阵A可对角化当且仅当A有n个线性无关的特征向量.

分析.
1° 必要性：A可对角化，即$A=PDP^{-1}$，D为对角矩阵，也即$AP=PD$.设$P=\begin{bmatrix}\mathbf{v}_1,\ldots,\mathbf{v}_n\end{bmatrix}$，$D=\begin{bmatrix}\lambda_1\mathbf{e}_1,\ldots,\lambda_n\mathbf{e}_n\end{bmatrix}$（注意这里$\mathbf{v}_i$只是$P$的列，还不是特征向量；同样$\lambda_i$也还不是$A$的特征值）.则

$$AP=\begin{bmatrix}A\mathbf{v}_1,\ldots,A\mathbf{v}_n\end{bmatrix}=PD=\begin{bmatrix}\lambda_1\mathbf{v}_1,\ldots,\lambda_n\mathbf{v}_n\end{bmatrix}$$

进而$A\mathbf{v}_i=\lambda_i\mathbf{v}_i$，$\lambda_i$为$A$的特征值，又$\mathbf{v}_i\neq 0$，$\mathbf{v}_i$为$\lambda_i$对应的特征向量.
由于P可逆，故P的列，也即A的特征向量线性无关.

2° 充分性：令$P=\begin{bmatrix}\mathbf{v}_1,\ldots,\mathbf{v}_n\end{bmatrix}$，其中$\mathbf{v}_i$为$\lambda_i$对应的特征向量，$D=\begin{bmatrix}\lambda_1\mathbf{e}_1,\ldots,\lambda_n\mathbf{e}_n\end{bmatrix}$，其中$\lambda_i$为$A$的特征向量.那么

$$AP=\begin{bmatrix}A\mathbf{v}_1,\ldots,A\mathbf{v}_n\end{bmatrix}=\begin{bmatrix}\lambda_1\mathbf{v}_1,\ldots,\lambda_n\mathbf{v}_n\end{bmatrix}=PD$$

因为$\mathbf{v}_i$线性无关，故P可逆，进而$A=PDP^{-1}$.
注：从以上构造可以知道λ相同并不影响对角矩阵的构造，但是$\mathbf{v}$相同则无法保证线性无关.

定理 38. 若$n\times n$矩阵A有n个不同的特征值，则A可对角化.

分析. 从定理*36*可知，$\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$线性无关，进而由定理*37*，$A$可对角化.

算法 6 (对角化)**.**

1. 找到A的特征值
2. 找到A的线性无关特征向量
3. 构造P，其中$P=\begin{bmatrix}\mathbf{v}_1,\ldots,\mathbf{v}_n\end{bmatrix}$
4. 构造D，其中$D=\begin{bmatrix}\lambda_1\mathbf{e}_1,\ldots,\lambda_n\mathbf{e}_n\end{bmatrix}$

注：对角化不是唯一的，会随着特征值摆放位置、特征向量的不同而改变. 可逆与可对角化没有必然联系.

定理 39. A是$n\times n$矩阵，$\lambda_1,\cdots,\lambda_p$是$A$不同的特征值，则

1. $\forall 1\le k\le p$，λ_k的几何重数$\le\lambda_k$的代数重数
2. A可对角化，当且仅当不同特征值的几何重数之和为n，而这当且仅当对于每一个λ_k，其代数重数都等于几何重数
3. 若A可对角化，$\mathcal{B}_k$是λ_k对应的特征空间的基，那么$\mathcal{B}_1,\ldots,\mathcal{B}_p$形成$\mathbb{R}^n$的一组特征向量基

定理 40 (线性变换矩阵)**.** $\mathcal{B}=\{\mathbf{b}_1,\ldots,\mathbf{b}_n\}$为向量空间$V$的基，$\mathcal{C}=\{\mathbf{b}c_1,\ldots,\mathbf{b}c_m\}$为向量空间$W$的基，线性变换$T:V\to W$，存在唯一$m\times n$矩阵$M$使得

$$[T(\mathbf{x})]_{\mathcal{C}}=M[\mathbf{x}]_{\mathcal{B}},\forall\mathbf{x}\in V$$

其中

$$M=\begin{bmatrix}[T(\mathbf{b}_1)]_{\mathcal{C}} & [T(\mathbf{b}_2)]_{\mathcal{C}} & \cdots & [T(\mathbf{b}_n)]_{\mathcal{C}}\end{bmatrix}$$

称为T关于基$\mathcal{B}$和$\mathcal{C}$的矩阵 *(the matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$)*.
特别地，当线性空间和基取特殊值时，可以得到我们之前求得的一些矩阵.

1. T的标准矩阵 *(standard matrix for T)*
   $A=\begin{bmatrix}T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n)\end{bmatrix}$，当$V=W=\mathbb{R}^n,\mathcal{B}=\mathcal{C}=\mathcal{E}$（$\mathbb{R}^n$中的标准基）
2. T的$\mathcal{B}$矩阵 *(the matrix for T relative to $\mathcal{B}$, or the $\mathcal{B}$-matrix for T)*
   $[T]_{\mathcal{B}}=\begin{bmatrix}[T(\mathbf{b}_1)]_{\mathcal{B}},\ldots,[T(\mathbf{b}_n)]_{\mathcal{B}}\end{bmatrix}$，当$V=W,\mathcal{B}=\mathcal{C}$

3. 坐标变换矩阵 *(change-of-coordinates matrix)*，右乘坐标可以将坐标变换成具体的$\mathbf{x}$

$P_{\mathcal{B}} = \begin{bmatrix}\mathbf{b}_1, \dots, \mathbf{b}_n\end{bmatrix}$，当$V = W = \mathbb{R}^n, \mathcal{C} = \mathcal{E}$，$T(\mathbf{x}) = \mathbf{x}$

4. 过渡矩阵/$\mathcal{B}$到$\mathcal{C}$的坐标变换矩阵 *(change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$)*

$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix}[\mathbf{b}_1]_{\mathcal{C}}, \dots, [\mathbf{b}_n]_{\mathcal{C}}\end{bmatrix}$，当$V = W$，$T(\mathbf{x}) = \mathbf{x}$

分析. 1° 存在性：

$$\begin{aligned}
[T(\mathbf{x})]_{\mathcal{C}} &= [T(x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n)]_{\mathcal{C}} \\
&= x_1[T(\mathbf{b}_1)]_{\mathcal{C}} + \cdots + x_n[T(\mathbf{b}_n)]_{\mathcal{C}} \\
&= \begin{bmatrix}[T(\mathbf{b}_1)]_{\mathcal{C}}, \dots, [T(\mathbf{b}_n)]_{\mathcal{C}}\end{bmatrix}\begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix} \\
&= \begin{bmatrix}[T(\mathbf{b}_1)]_{\mathcal{C}}, \dots, [T(\mathbf{b}_n)]_{\mathcal{C}}\end{bmatrix}[\mathbf{x}]_{\mathcal{B}} \\
&= M[\mathbf{x}]_{\mathcal{B}}
\end{aligned}$$

2° 唯一性：设存在另一矩阵M'使得$[T(\mathbf{x})]_{\mathcal{C}} = M'[\mathbf{x}]_{\mathcal{B}}$，故$M[\mathbf{x}]_{\mathcal{B}} = M'[\mathbf{x}]_{\mathcal{B}}$，令$\mathbf{x} = \mathbf{b}_i$即得$M = M'$

命题 10. 若$A = PDP^{-1}$，D是$n \times n$的对角矩阵，若$\mathcal{B}$是$\operatorname{Col} P$的一组基，则D是线性变换$\mathbf{x} \mapsto A\mathbf{x}$的$\mathcal{B}$-矩阵.

分析.

$$\begin{aligned}
[T]_{\mathcal{B}} &= \begin{bmatrix}[T(\mathbf{b}_1)]_{\mathcal{B}}, \dots, [T(\mathbf{b}_n)]_{\mathcal{B}}\end{bmatrix} \\
&= \begin{bmatrix}[A\mathbf{b}_1]_{\mathcal{B}}, \dots, [A\mathbf{b}_n]_{\mathcal{B}}\end{bmatrix} \\
&= \begin{bmatrix}P^{-1}A\mathbf{b}_1, \dots, P^{-1}A\mathbf{b}_n\end{bmatrix} \quad \text{坐标变换} \\
&= P^{-1}A\begin{bmatrix}\mathbf{b}_1, \dots, \mathbf{b}_n\end{bmatrix} \\
&= P^{-1}AP = D
\end{aligned}$$

求解方法$\begin{bmatrix}P & AP\end{bmatrix} \sim \begin{bmatrix}I & P^{-1}AP\end{bmatrix}$. 由上面分析知$D$不需是对角矩阵，只要$A$与$D$相似即可.

如下图所示[3]，一个线性变换T对于标准基（或其他基）的矩阵为A，为了更清楚地通过矩阵看出这个线性变换的效果，就将A对角化：$A = PDP^{-1}$.这其实相当于先把标准基换成由特征向量组成的基(P^{-1}的意义)，于是每一个基向量在经过T变换后都只是乘了个常数(D的意义)，最后再把由特征向量组成的基换回标准基(P的意义). 因此对角化其实是用一组比标准基更好的基来描述线性变换，也就是由特征向量组成的基. 至于更好的基，由特征向量组成的规范正交基（谱定理）描述的则更好.

定义 21 (迹(trace))**.** A为$n \times n$的方阵，A主对角线上元素之和称为A的迹，定义为

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$$

定理 41. 若A, B均为$n \times n$的方阵，则$\operatorname{tr}(AB) = \operatorname{tr}(BA)$

[3]图根据课本重绘，话摘自线性变换的矩阵为什么要强调在这组基下？ - 匡世珉的回答 - 知乎 https://www.zhihu.com/question/22218306/answer/88697757

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "乘A"] \arrow[d, "乘P^{-1}"'] & A\mathbf{x} \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "乘D"] & {[A\mathbf{x}]_{\mathcal{B}}} \arrow[u, "乘P"']
\end{tikzcd}$$

分析.

$$\operatorname{tr}(AB)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}b_{ji}=\sum_{j=1}^{n}\sum_{i=1}^{n}a_{ji}b_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{n}b_{ij}a_{ji}=\operatorname{tr}(BA)$$

定理 42 (相似性一些定理)**.** 已知A与B相似，

1. 若A可逆，则A^{-1}与B^{-1}相似
2. 则A^k与B^k相似
3. A与C相似，则B与C相似
4. A可对角化，则B也可对角化
5. $A=PBP^{-1}$，$\mathbf{x}$是A关于λ的特征向量，则$P^{-1}\mathbf{x}$是B关于λ的特征向量
6. $\operatorname{rank}A=\operatorname{rank}B$

5 对称矩阵与二次型

5.1 谱分解

定义 22 (对称矩阵)**.** $A^{\mathrm{T}}=A$（显然得是方阵）

定理 43. A为对称矩阵，任两个不同特征空间的特征向量一定正交

分析. 设$\mathbf{v}_1,\mathbf{v}_2$为不同特征值$\lambda_1,\lambda_2$对应的特征向量

$$\begin{aligned}\lambda_1\mathbf{v}_1\cdot\mathbf{v}_2&=(\lambda_1\mathbf{v}_1)^{\mathrm{T}}\mathbf{v}_2=(A\mathbf{v}_1)^{\mathrm{T}}\mathbf{v}_2\\&=\mathbf{v}_1^{\mathrm{T}}A^{\mathrm{T}}\mathbf{v}_2=\mathbf{v}_1^{\mathrm{T}}(A\mathbf{v}_2)\\&=\lambda_2\mathbf{v}_1\cdot\mathbf{v}_2\end{aligned}$$

进而$(\lambda_1-\lambda_2)\mathbf{v}_1\cdot\mathbf{v}_2=0$，而$\lambda_1,\lambda_2$不同，故$\mathbf{v}_1\cdot\mathbf{v}_2=0$

定义 23 (正交对角化)**.** 若存在对角矩阵D和正交矩阵P使得$A=PDP^{\mathrm{T}}=PDP^{-1}$，则称$A$可正交对角化

算法 7 (正交对角化)**.**

1. 先将矩阵一般对角化，暂不用求P^{-1}
2. 然后将P的列施密特**单位**正交化，使P变为正交矩阵
3. $P^{\mathrm{T}}=P^{-1}$直接求得P的逆

定理 44 (谱定理)**.** A为$n\times n$的对称矩阵，则

1. A有n个实特征值（包含重数）

2. $\dim W_{\lambda_i} = \lambda_i$的代数重数

3. 特征空间相互正交，对应不同特征值的特征向量都正交

4. A可以被正交对角化（**充要条件**）

注：正交矩阵不一定可以被正交对角化，正交对角化一定会产生正交矩阵，要理清关系.

定理 45 (谱分解). A为对称矩阵，记正交矩阵P的列为$\mathbf{u}_i$，则$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathrm{T}} + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^{\mathrm{T}} + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^{\mathrm{T}}$

定义 24 (投影矩阵). A是对称矩阵且$A^2 = A$，则称其为投影矩阵

命题 11. 若$A = \mathbf{u}\mathbf{u}^{\mathrm{T}}$，其中$\mathbf{u} \in \mathbb{R}^n$为**单位**向量，则

1. A是投影矩阵

2. $A\mathbf{x} = \operatorname{proj}_{\mathbf{u}} \mathbf{x}$

3. $\mathbf{u}$是A的特征向量

5.2 奇异值分解

定理 46. $A^{\mathrm{T}}A$的特征值均为非负数

分析. $\forall \mathbf{u}_i, \|A\mathbf{u}_i\|^2 = (A\mathbf{u}_i)^{\mathrm{T}} A\mathbf{u}_i = \mathbf{u}_i^{\mathrm{T}}(A^{\mathrm{T}}A)\mathbf{u}_i = \mathbf{u}_i^{\mathrm{T}} \lambda_i \mathbf{u}_i = \lambda_i \mathbf{u}_i \cdot \mathbf{u}_i = \lambda_i \|\mathbf{u}_i\| \geq 0 \implies \lambda_i \geq 0$，故记$\sigma_i = \sqrt{\lambda_i} = \|A\mathbf{u}_i\|$为$A$的奇异值，注意这里的$\lambda_i$都为$A^{\mathrm{T}}A$的特征值.

定义 25 (奇异值). A的奇异值是$A^{\mathrm{T}}A$的特征值的算术平方根，记为$\sigma_i := \sqrt{\lambda_i}, \forall 1 \leq i \leq n$，且呈递减序列.

定理 47. A为$m \times n$矩阵，$A^{\mathrm{T}}A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$，假设有$r$个非零特征值，不妨设$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$，则$\{A\mathbf{u}_1, \ldots, A\mathbf{u}_r\}$是$\operatorname{Col} A$正交基，且$\operatorname{rank} A = r$.

分析. $\|A\mathbf{u}_i\| = \sqrt{\lambda_i} \neq 0, i = 1, \ldots, r$
正交性：$\forall i \neq j, (A\mathbf{u}_i) \cdot (A\mathbf{u}_j) = \mathbf{u}_i^{\mathrm{T}} A^{\mathrm{T}} A \mathbf{u}_j = \lambda_j \mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = 0$
任取$\mathbf{v} \in \operatorname{Col} A$，即$\exists \mathbf{x} \in \mathbb{R}^n \; s.t. \; A\mathbf{x} = \mathbf{v}$，其中$\mathbf{x} = c_1 \mathbf{u}_1 + \cdots + c_n \mathbf{u}_n$，则$\mathbf{v} = c_1 A\mathbf{u}_1 + \cdots + c_n A\mathbf{u}_n = c_1 A\mathbf{u}_1 + \cdots + c_r A\mathbf{u}_r + 0 + \cdots + 0$，即说明$\mathbf{v}$可由$\{A\mathbf{u}_1, \ldots, A\mathbf{u}_r\}$线性表示，故证毕.

定理 48 (奇异值分解(SVD)). A是$m \times n$的矩阵，$\operatorname{rank} A = r$，则存在$m \times n$矩阵$\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$，其中$D$的对角线是前$r$个$A$的奇异值，且$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$，那么存在$m \times m$正交矩阵$U$和$n \times n$正交矩阵$V$使得$A = U\Sigma V^{\mathrm{T}}$.

分析. 设$A^{\mathrm{T}}A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$，其有$r$个非零特征值，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$.
$\{A\mathbf{u}_1, \ldots, A\mathbf{u}_r\}$是$\operatorname{Col} A$正交基，将其标准化，得$\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$，其中$\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i} A\mathbf{v}_i$，即$A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.
将$\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$拓展为$\mathbb{R}^m$的正交基$\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$，令

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}, V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$

那么U,V均为正交矩阵，而且

$$AV = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_r & \mathbf{0}\cdots\mathbf{0} \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0}\cdots\mathbf{0} \end{bmatrix}$$

进而

$$\begin{aligned} U\Sigma &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} D & O \\ O & O \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_r & \mathbf{0}\cdots\mathbf{0} \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0}\cdots\mathbf{0} \end{bmatrix} \\ &= AV \end{aligned}$$

因为V为正交矩阵，故$U\Sigma V^{\mathrm{T}} = AVV^{\mathrm{T}} = A$.

算法 8 (奇异值分解)**.** 分为以下几个步骤

1. 对$A^{\mathrm{T}}A$正交对角化
2. 建立起Σ和V
3. 建立起U

5.3 二次型

定义 26 (二次型)**.** $Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$，其中$A$是对称矩阵

将二次型合并为矩阵的写法，平方项放对角线，交叉项取一半对称写. 以三元二次型为例，观察下面各个元素的去向.

$$\begin{aligned} Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + (a_{12}+a_{21})x_1x_2 + (a_{13}+a_{31})x_1x_3 + (a_{23}+a_{32})x_2x_3 \end{aligned}$$

定理 49 (主轴定理)**.** $\mathbf{x} = P\mathbf{y}$，存在P将二次型$\mathbf{x}^{\mathrm{T}}A\mathbf{x}$转为$\mathbf{y}^{\mathrm{T}}D\mathbf{y}$，其中$P$的列（$A$的单位特征向量）称为$\mathbf{x}^{\mathrm{T}}A\mathbf{x}$的主轴.

算法 9 (标准型变换)**.** 将矩阵A**正交对角化**后，得到$A = PDP^{-1}$，做变换$\mathbf{x} = P\mathbf{y}$，可得$\mathbf{x}^{\mathrm{T}}A\mathbf{x} = \mathbf{y}^{\mathrm{T}}D\mathbf{y}$，由于$D$为对角矩阵，由上面的分析即可知变换后的二次型不存在交叉项，这种形式称为**标准型**. 又$\mathbf{x} = P\mathbf{y}$，令$\mathbf{y} = \mathbf{e}_i$（变换后的主轴在坐标轴上），可得$\mathbf{x} = \mathbf{u}_i$，即原来的主轴为$P$的列，也即$A$的特征向量.

定义 27. 对于二次型$Q(\mathbf{x})$

1. 若$Q(\mathbf{x}) > 0\ \forall \mathbf{x} \neq \mathbf{0}$，正定 $\iff \forall\lambda > 0$；$Q(\mathbf{x}) \geq 0\ \forall \mathbf{x} \neq \mathbf{0}$，半正定
2. 若$Q(\mathbf{x}) < 0\ \forall \mathbf{x} \neq \mathbf{0}$，负定 $\iff \forall\lambda < 0$；$Q(\mathbf{x}) \leq 0\ \forall \mathbf{x} \neq \mathbf{0}$，半负定
3. $Q(\mathbf{x})$正负值都有，不定

分析. 对二次型进行标准型变换使其没有交叉项，$Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x} = \mathbf{y}^{\mathrm{T}}D\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$，进而得证.

命题 12. *1.* B为$m\times n$矩阵，$B^{\mathrm{T}}B$为半正定；若B为方阵且可逆，则$B^{\mathrm{T}}B$正定.

分析. $\mathbf{x}^{\mathrm{T}}(B^{\mathrm{T}}B)\mathbf{x}=\|B\mathbf{x}\|\geq 0$，可逆则$B\mathbf{x}=\mathbf{0}$只有平凡解.

2. A为方阵且正定，那么存在一个正定方阵B使得$A=B^{\mathrm{T}}B$

命题 13. A,B均为方阵，且特征值均为正数，则$A+B$的特征值也全为正数.

分析. $\mathbf{x}^{\mathrm{T}}A\mathbf{x}>0,\mathbf{x}^{\mathrm{T}}B\mathbf{x}>0\implies\mathbf{x}^{\mathrm{T}}(A+B)\mathbf{x}>0$

5.4 带约束的最值问题

定理 50. 二次型Q，$\min\limits_{\|\mathbf{x}\|=1}Q(\mathbf{x})=\lambda_{\min}$，$\max\limits_{\|\mathbf{x}\|=1}Q(\mathbf{x})=\lambda_{\max}$

分析. 化为标准型后直接放缩即可.

定理 51. A为$m\times n$矩阵，$\min\limits_{\|\mathbf{x}\|=1}\|A\mathbf{x}\|=\lambda_{\min}$，$\max\limits_{\|\mathbf{x}\|=1}\|A\mathbf{x}\|=\lambda_{\max}$

分析. $\|A\mathbf{x}\|^2=(A\mathbf{x})^{\mathrm{T}}A\mathbf{x}=\mathbf{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\mathbf{x}$进而转化为二次型. 实际上$\|A\mathbf{x}\|\leq\lambda\|x\|$.

6 仿射空间

由于这一章涉及的概念太多，且比较难找到对应的中文术语，故在专业术语后都会标上对应的英文名称.

6.1 仿射组合与凸组合

定义 28 (仿射组合与凸组合)**.** $\mathbb{R}^n$中的向量$\mathbf{v}_1,\dots,\mathbf{v}_p$的线性组合$c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p$，若满足权重$c_1+\cdots+c_p=1$，则该线性组合称为仿射组合，集合$S$所有点仿射组合的集合称为仿射包(affine hull)或仿射生成(affine span)集，记为$\operatorname{aff}S$. 在仿射组合的基础上满足权重$c_i\geq 0,i=1,\dots,p$，则该线性组合称为凸组合，集合S所有点凸组合的集合称为凸包(convex hull)，记为$\operatorname{conv}S$.

由定义得知，$S=\{\mathbf{v}_1\}$时，$\operatorname{aff}S$和$\operatorname{conv}S$均为一个点$\mathbf{v}_1$；$S=\{\mathbf{v}_1,\mathbf{v}_2\}$时，$\operatorname{aff}S$是通过$\mathbf{v}_1,\mathbf{v}_2$的直线（定比分点），$\operatorname{conv}S$则是线段$\overline{\mathbf{v}_1\mathbf{v}_2}$；$S=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$时，$\operatorname{aff}S$是通过$\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$的平面，$\operatorname{conv}S$则是三角形$\triangle\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$；以此类推.

定义 29 (仿射(affine)与凸性(convex))**.** 对于集合S，若$\forall\mathbf{u},\mathbf{v}\in S,(1-t)\mathbf{u}+t\mathbf{v}\in S,\forall t\in\mathbb{R}$，则称$S$是仿射集. 若$\forall\mathbf{u},\mathbf{v}\in S,\overline{\mathbf{uv}}\in S$，则称$S$是凸集.

定理 52. 以下是关于仿射集和凸集的一些定理.

1. S是仿射集当且仅当$S=\operatorname{aff}S$；S是凸集当且仅当$S=\operatorname{conv}S$
2. 仿射集的子集是仿射集，凸集的子集是凸集

定义 30. $\mathbb{R}^n$的一个集合S经向量$\mathbf{p}$的平移(translate)变为$S+\mathbf{p}=\{\mathbf{s}+\mathbf{p}|\mathbf{s}\in S\}$. $\mathbb{R}^n$的一个面(flat)[4]是$\mathbb{R}^n$子空间的一个平移. 若一个面是另一个的平移，则两个面平行. 一个面的维度是它对应的平行子空间的维度. S的维度是包含S最小的面的维度. $\mathbb{R}^n$中的线是一维的面，$\mathbb{R}^n$中的超平面(hyperplane)是$n-1$维的面.
注：由于子空间必须包含零点，故经过平移的子空间就不是子空间了，就称为面. 真(proper)面即不包含自己的面.

定义 31 (齐次(homogeneous)坐标)**.** 对于$\mathbf{v}\in\mathbb{R}^n$，其标准齐次坐标为$\tilde{\mathbf{v}}=\begin{bmatrix}\mathbf{v}\\1\end{bmatrix}\in\mathbb{R}^{n+1}$

引入齐次坐标的好处有以下几点.

1. 合并矩阵加法乘法运算
2. 引入无穷远点等同于其他点

定义 32 (仿射无关)**.** *若存在$c_1,\ldots,c_p$不全为0，使得$c_1+\cdots+c_p=0, c_1\mathbf{v}_1+\cdots+c_p\mathbf{v}_p=\mathbf{0}$，则称$\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}\in\mathbb{R}^n$仿射相关，否则称为仿射无关.*

定理 53. *$S=\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}\in\mathbb{R}^n, p\geq 2$，以下命题都等价*

1. *S仿射相关*
2. *S其中一个点是其他点的仿射组合*
3. *$\{\mathbf{v}_2-\mathbf{v}_1,\ldots,\mathbf{v}_p-\mathbf{v}_1\}\in\mathbb{R}^n$线性相关*
4. *$\{\tilde{\mathbf{v}}_1,\ldots,\tilde{\mathbf{v}}_p\}\in\mathbb{R}^{n+1}$线性相关*

定理 54 (向量唯一表示定理（仿射组合）)**.** *$S=\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$是$\mathbb{R}^n$的仿射无关集，那么$\forall\mathbf{p}\in\operatorname{aff}S$有唯一向量表示*

$$\mathbf{p}=c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k \quad 且 \quad c_1+\cdots+c_k=1$$

其中$c_1,\ldots,c_p$称为$\mathbf{p}$的质心(barycentric)坐标或仿射坐标.

分析. 相当于解$\tilde{\mathbf{p}}=c_1\tilde{\mathbf{v}}_1+\cdots+c_k\tilde{\mathbf{v}}_k$，通过高斯消元初等行变换可得. 广泛应用于图像渐变、透视关系.

定理 55. *1. 一个非空子集S是仿射集当且仅当它是一个面*

2. *S仿射无关，$\mathbf{p}\in\operatorname{aff}S$，那么$\mathbf{p}\in\operatorname{conv}S$当且仅当$\mathbf{p}$的质心坐标全部非负*
3. *$\operatorname{conv}S$是所有包含S的凸集的交*

定理 56 (Caratheodory)**.** *S是$\mathbb{R}^n$的一个非空子集，那么$\operatorname{conv}S$的每一个点都可以被表示成S中小于等于$n+1$个点的凸组合*

6.2 超平面

定义 33 (线性泛函(linear functional))**.** *线性变换$f:\mathbb{R}^n\to\mathbb{R}$是一个线性泛函. $\forall d\in\mathbb{R}, [f:d]:=\{\mathbf{x}\in\mathbb{R}^n: f(\mathbf{x})=d\}$. 零泛函是使得$f(\mathbf{x})=0,\forall\mathbf{x}\in\mathbb{R}^n$的变换，其他均称为非零.*

[4] https://en.wikipedia.org/wiki/Flat_(geometry)

定理 57. $\mathbb{R}^n$的子集H是一个超平面当且仅当$H=[f:d]$，其中f是非零泛函，$d\in\mathbb{R}$. 因此，若H是一个超平面，则存在一个非零向量$\mathbf{n}$和$d\in\mathbb{R}$使得$H=\{\mathbf{x}:\mathbf{n}\cdot\mathbf{x}=d\}$

定义 34 (拓扑概念)**.** $\forall\mathbf{p}\in\mathbb{R}^n,\delta>0$，中心为$\mathbf{p}$，半径为$\delta$的开球(open ball)记为$B(\mathbf{p},\delta):=\{\mathbf{x}:\|\mathbf{x}-\mathbf{p}\|<\delta\}$. 给定集合$S\in\mathbb{R}^n$，若$\exists\delta>0\,s.t.\,B(\mathbf{p},\delta)\subset S$，则称$\mathbf{p}$是$S$的内点(interior point). 若每一个中心在$\mathbf{p}$的开球都与S和S的补相交，则$\mathbf{p}$是S的边界点(boundary point). 若S不包含任一边界点，则称S是开的(open)；若S包含所有边界点，则称S是闭的(closed)；否则既不开也不闭. 若$\exists\delta>0\,s.t.\,S\subset B(\mathbf{0},\delta)$，则称$S$是有界的(bounded). S同时是闭的，又是有界的，则称S是紧的(compact).（欧氏空间上有界闭集等于紧集）

定理 58. 开集的凸包是开的，紧集的凸包是紧的，但闭集的凸包不一定是闭的

定义 35. 称超平面$H=[f:d]$将集合A,B分隔(seperate)，若满足以下其一

1. $f(A)\leq d$且$f(B)\geq d$
2. $f(A)\geq d$且$f(B)\leq d$

定理 59. A,B均为非空凸集，A是紧集，B是闭集，那么存在一个超平面H严格分割A,B，当且仅当$A\cap B=\varnothing$

定理 60. A,B均为非空紧集，那么那么存在一个超平面H严格分割A,B，当且仅当$(\mathrm{conv}\,A)\cap(\mathrm{conv}\,B)=\varnothing$

6.3 多胞形

定义 36 (多胞形(polytopes))**.** $\mathbb{R}^n$中的多胞形是有限集合的凸包. $\mathbb{R}^2$中，多胞体就是简单的多边形；$\mathbb{R}^3$，则是多面体.

定义 37. S是$\mathbb{R}^n$紧致的凸集，F是S的非空子集，若$F\neq S$，$\exists H=[f:d]\,s.t.\,F=S\cap H$以及$f(S)\leq d$或$f(S)\geq d$，则称$F$为$S$的面(face). H称为S的支撑超平面(supporting hyperplane). 若$\dim F=k$，则F称为S的k-面. 若P是k维的多胞体，称P为k-多胞体. P的0-面称为顶点(vertex/vertices)，1-面称为边(edge)，$k-1$-面是S的facet.

定义 38 (端点)**.** S是凸集，若$\mathbf{x},\mathbf{y}\in S,\mathbf{p}\in\overline{\mathbf{xy}}$，则$\mathbf{p}=\mathbf{x}$或$\mathbf{p}=\mathbf{y}$称为$S$的端点(extreme point)，$S$的所有端点称为$S$的轮廓(profile).

定义 39 (最小表示)**.** 若多胞形$P=\mathrm{conv}\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$且$\mathbf{v}_i\notin\mathrm{conv}\{\mathbf{v}_j:j\neq i\},\forall i=1,\ldots,k$，则称$\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$是$P$的最小表示(minimal representation).

定理 61. $M=\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$是多胞形$P$的最小表示，则以下说法等价

1. $\mathbf{p}\in M$

2. $\mathbf{p}$是P的顶点

3. $\mathbf{p}$是P的端点

定理 62. S是非空紧致凸集，则S是它轮廓（S的端点）的凸包

定理 63. f是一个定义在非空紧致凸集S上的线性泛函，则存在S的端点$\hat{\mathbf{v}},\hat{\mathbf{w}}$使得

$$f(\hat{\mathbf{v}})=\max_{\mathbf{v}\in S}f(\mathbf{v})\quad f(\hat{\mathbf{w}})=\min_{\mathbf{v}\in S}f(\mathbf{v})$$

注：线性规划极值点取得依据

定义 40 (单纯形(simplex))**.** 单纯形是有限仿射无关向量构成的集合的凸包. $\mathbb{R}^2$中，多胞体就是简单的多边形；$\mathbb{R}^3$，则是多面体.

$$\begin{aligned}
&0\text{-单纯形}S^0:\ \text{一个点}\{\mathbf{v}_1\}\\
&\qquad\vdots\\
&k\text{-单纯形}S^k:\ \operatorname{conv}(S^{k-1}\cup\{\mathbf{v}_{k+1}\})\text{，其中}\mathbf{v}_{k+1}\notin S^{k-1}
\end{aligned}$$

定义 41 (超立方体(hypercube))**.** $I_i=\overline{\mathbf{0}\mathbf{e}_i}$，向量和(vector sum)[5]$C^k=I_1+I_2+\cdots+I_k$称为f$k$-维超立方体

定理 64 (欧拉(Euler)公式)**.** 记$f_k(P)$为n-维多胞形P的k-维面的数目，则

$$\sum_{k=0}^{n-1}(-1)^k f_k(P)=1+(-1)^{n-1}$$

特别地，当$n=3$时，有$v-e+f=2$，其中v,e,f分别为顶点、边、面的数量.

6.4 曲线和表面

定义 42 (贝塞尔(Bézier)曲线)**.** 三阶贝塞尔曲线

$$\mathbf{x}(t)=\begin{bmatrix}\mathbf{p}_0 & \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3\end{bmatrix}\begin{bmatrix}1 & -3 & 3 & -1\\ 0 & 3 & -6 & 3\\ 0 & 0 & 3 & -3\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ t\\ t^2\\ t^3\end{bmatrix}=GM_B\mathbf{u}(t)$$

其中G为四个控制点构成的几何(geometry)矩阵，M_B为贝塞尔基底矩阵. 换种形式表示

$$\mathbf{x}(s)=\mathbf{u}(s)^{\mathrm{T}}M_B^{\mathrm{T}}G^{\mathrm{T}}=\begin{bmatrix}(1-s)^3 & 3s(1-s)^2 & 3s^2(1-s) & s^3\end{bmatrix}\begin{bmatrix}\mathbf{p}_0\\ \mathbf{p}_1\\ \mathbf{p}_2\\ \mathbf{p}_3\end{bmatrix}$$

[5] $A+B=\{\mathbf{c}:\mathbf{c}=\mathbf{a}+\mathbf{b},\forall\mathbf{a}\in A,\mathbf{b}\in B\}$

定义 43 (贝塞尔表面)**.**

$$GM_B\mathbf{u}(t)\begin{bmatrix}\mathbf{p}_{11}&\mathbf{p}_{12}&\mathbf{p}_{13}&\mathbf{p}_{14}\\\mathbf{p}_{21}&\mathbf{p}_{22}&\mathbf{p}_{23}&\mathbf{p}_{24}\\\mathbf{p}_{31}&\mathbf{p}_{32}&\mathbf{p}_{33}&\mathbf{p}_{34}\\\mathbf{p}_{41}&\mathbf{p}_{42}&\mathbf{p}_{43}&\mathbf{p}_{44}\end{bmatrix}\begin{bmatrix}(1-t)^3\\3t(1-t)^2\\3t^2(1-t)\\t^3\end{bmatrix}$$

进而，

$$\mathbf{x}(s,t)=\mathbf{u}(s)^{\mathrm{T}}M_B^{\mathrm{T}}GM_B\mathbf{u}(t),0\leq s,t\leq 1$$

7 拓展

7.1 线性模型与最小二乘

定理 65 (最小二乘)**.** A为$m\times n$的矩阵，$\mathbf{b}\in\mathbb{R}^m$，$A\mathbf{x}=\mathbf{b}$的最小二乘解是$\hat{\mathbf{x}}\in\mathbb{R}^n$使得

$$\|\mathbf{b}-A\hat{\mathbf{x}}\|\leq\|\mathbf{b}-A\mathbf{x}\|$$

则$\hat{\mathbf{b}}=\mathrm{proj}_{\mathrm{Col}\,A}\mathbf{b}$，解与$A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$相同.
若A可被QR分解（定理*35*），则$\hat{\mathbf{x}}=R^{-1}Q^{\mathrm{T}}\mathbf{b}$.

分析. $\mathbf{a}_j\cdot(\mathbf{b}-A\hat{\mathbf{x}})=\mathbf{a}_j^{\mathrm{T}}(\mathbf{b}-A\hat{\mathbf{x}})=0\implies A^{\mathrm{T}}(\mathbf{b}-A\hat{\mathbf{x}})=\mathbf{0}\implies A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$

算法 10 (最小二乘估计)**.** 直接计算$A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$即可.

定义 44 (一般线性模型)**.** $\mathbf{y}=X\boldsymbol{\beta}+\varepsilon$，$X$为设计矩阵，$\boldsymbol{\beta}$为参数向量，$\mathbf{y}$为观测向量，$\varepsilon$为残差向量，满足这种形式的方程称为线性模型. 使$\varepsilon$得长度最小化，即找出$X\boldsymbol{\beta}=\mathbf{y}$的最小二乘解.

以下是一些例子.

1. 直线拟合

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}1&x_1\\1&x_2\\\vdots&\vdots\\1&x_n\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}$$

2. 抛物线拟合

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}1&x_1&x_1^2\\1&x_2&x_2^2\\\vdots&\vdots&\vdots\\1&x_n&x_n^2\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix}$$

3. 多重回归

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

7.2 马尔可夫链

定义 45 (马尔可夫(Markov)链)**.** 具有非负分量的数值且相加等于1的向量$\mathbf{x}_0, \mathbf{x}_1, \ldots$称为**概率向量**，各列向量均为概率向量的方阵$P$称为**随机矩阵**，则**马尔可夫链**为$\mathbf{x}_{k+1} = P\mathbf{x}_k, k = 0, 1, 2, \ldots$，其中$\mathbf{x}_k$称为**状态向量**.

定理 66 (马尔可夫链收敛定理[6])**.** 马尔可夫链$\{\mathbf{x}_{k+1}\}$一定会收敛至**平衡向量**$\mathbf{q}$，其中$\mathbf{q}$满足$P\mathbf{q} = \mathbf{q}$.

7.3 复数特征值

定理 67. 当A为实矩阵时，它的复特征值成对出现.

分析. $A\bar{\mathbf{x}} = \bar{A}\bar{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}$

定理 68. A为2×2实矩阵，有复特征值$\lambda = a - b\mathrm{i}(b \neq 0)$及对应$\mathbb{C}^2$中的复特征向量$\mathbf{v}$，则

$$A = PCP^{-1}\text{，其中 } P = \begin{bmatrix} \Re\mathbf{v} & \Im\mathbf{v} \end{bmatrix}, C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

分析. 证明利用了结论：A是实矩阵，则$A(\Re\mathbf{x}) = \Re A\mathbf{x}$，$A(\Im\mathbf{x}) = \Im A\mathbf{x}$. $\Re\mathbf{x}$与$\Im\mathbf{x}$线性无关

7.4 离散动力系统

若A可对角化，有n个线性无关的特征向量$\mathbf{v}_1, \ldots, \mathbf{v}_n$和对应的特征值$\lambda_1, \ldots, \lambda_n$，且由大到小排列. 由于$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$为$\mathbb{R}^n$的基. 故任一初始向量可唯一表示为$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$，则

$$\mathbf{x}_k = A^k\mathbf{x}_0 = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n$$

线性动力系统中，只有原点才可能是吸引子($|\lambda| < 1$)或者排斥子($|\lambda| > 1$)，但非线性系统中可能存在多个吸引子或排斥子，其可用雅可比矩阵的特征值定义. 若特征值正负都有，则原点为鞍点.

7.5 估计特征值

幂算法、逆幂法、QR算法

若A可对角化，特征向量$\mathbf{v}_1, \ldots, \mathbf{v}_n$是$\mathbb{R}^n$的基，且

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

[6]具体情况比较复杂

（注意第一个符号为严格大）其中λ_1称为主特征值

$$\mathbf{x}_k = A^k\mathbf{x}_0 = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n$$

假设$c_1 \neq 0$，左右同除λ_1^k，可知$(\lambda_1)^{-k}A^k\mathbf{x} \to c_1\mathbf{v}_1 (k \to \infty)$

7.6 傅里叶级数

$\forall n \geq 1, \{1, \cos t, \cos 2t, \cdots, \cos nt, \sin t, \sin 2t, \cdots, \sin nt\}$，定义内积$\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\,\mathrm{d}t$，知这个集合为正交集

定理 69 (傅里叶(Fourier)级数)**.**

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \sin kt)$$

当$k \geq 1$时，

$$a_k = \frac{\langle f, \cos kt\rangle}{\langle \cos kt, \cos kt\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\,\mathrm{d}t,\ b_k = \frac{\langle f, \sin kt\rangle}{\langle \sin kt, \sin kt\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\,\mathrm{d}t$$

常数项，

$$\frac{\langle f, 1\rangle}{\langle 1, 1\rangle} = \frac{a_0}{2}$$

7.7 统计学应用

定义 46 (平均值)**.** $p \times N$的观测矩阵$A = \begin{bmatrix}\mathbf{X}_1, \ldots, \mathbf{X}_N\end{bmatrix}$（即有$N$个样本，每个样本有$p$个维度的信息），则其样本均值为$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$. 平均偏差形式为$B = \begin{bmatrix}\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_N\end{bmatrix}$，其中$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$.

定义 47 (方差)**.** 协方差矩阵为$S = \frac{1}{N-1}BB^{\mathrm{T}}$（$BB^{\mathrm{T}}$正定，故$S$正定），其中元素$s_{ii}$称为$x_i$（某一行）的方差，$s_{ij}, i \neq j$称为$x_i$与$x_j$的协方差，而总方差$\mathrm{TVar}\, B = \mathrm{tr}\, S$

定义 48 (主成分分析)**.** 假设$A = \begin{bmatrix}\mathbf{X}_1, \ldots, \mathbf{X}_N\end{bmatrix}$为平均偏差形式，找到$p \times p$正交矩阵$P = \begin{bmatrix}\mathbf{u}_1, \ldots, \mathbf{u}_p\end{bmatrix}$使得$\mathbf{X} = P\mathbf{Y}$，其中$y_1, \ldots, y_p$都线性无关且方差递减. $\mathbf{u}_1, \ldots, \mathbf{u}_p$为数据的主成分，第一主成分是$S$最大的特征值对应的特征向量.
注：主成分分析等同于正交回归

分析. $\mathbf{X}_k = P\mathbf{Y}_k \implies \mathbf{Y}_k = P^{-1}\mathbf{X}_k = P^{\mathrm{T}}\mathbf{X}_k, k = 1, \ldots, N \implies S = PDP^{\mathrm{T}}, P^{\mathrm{T}}SP = D$
可以验证对于任意正交矩阵P，$\mathbf{Y}, \ldots \mathbf{Y}_N$的协方差矩阵都是$P^{\mathrm{T}}SP$

变量的正交变换$\mathbf{X} = P\mathbf{Y}$不改变数据的总方差，且$\mathrm{TVar}\,\mathbf{X} = \mathrm{TVar}\,\mathbf{Y} = \mathrm{tr}\, D = \lambda_1 + \cdots + \lambda_p$，商$\lambda_j / \mathrm{tr}\, S$表明$y_j$的占比，用于降维