线性代数期末小班辅导讲义

李子钰

2020 年 12 月 26 日

1 正交相关的问题

1.1 基本概念

回忆向量的内积，正交补，勾股定理，正交分解等.

1.2 投影与最小二乘

定理 1. 1. $A^TAx = A^Tb$ 一定有解；

2. $C(A)^\perp = N(A^T), N(A)^\perp = C(A^T)$；

3. x_0 为 $Ax = \beta$ 的最小二乘解等价于 x_0 为方程组 $A^TAx = A^T\beta$ 的正常解.

注 1. 1. 在 A^TA 可逆的时候，我们可以算得到 A 列空间的投影矩阵为 $A(A^TA)^{-1}A^T$，最小二乘解为 $(A^TA)^{-1}A^T\beta$.

2. 给定一个子空间 V，我们计算到这个子空间 V 上的投影矩阵，只要取 V 中的一组基，拼成矩阵

$$A = (v_1, \cdots, v_n),$$

然后化归到前一条；

3. 如果 A^TA 不可逆，你求最小二乘解的时候就只能硬解方程

$$A^TAx = A^T\beta.$$

4. 如果要求求最优最小二乘解，你还需要找

$$A^TAx = A^T\beta.$$

在 $C(A^TA)$ 中的解.

学习投影矩阵的关键不近要熟悉它的代数表达，更要清楚它的几何描述.

例 1. 给定向量 $\alpha = (1,2,3)^T$，P 是把向量投影到 α 所在直线的投影矩阵，求 P 的列空间与零空间.

解答. 列空间是 $(1,2,3)^T$ 生成的子空间，零空间是与它正交的空间，先取 $\beta = (2,-1,0)$，然后再取 $\beta \times \alpha$ 即可得张成.

1.3 正交化

回忆什么是正交向量组，标准正交向量组和标准正交基.

定理 2. 非零的正交向量组一定线性无关.

注 2. 我们可以使用把正交向量组

$$v_1, \cdots, v_r$$

以列向量的方式拼成矩阵 A 的方式去研究它们:

1. A 会是一个列满秩矩阵;

2. A^TA 是可逆对角矩阵，且对角线上都是 v_i 的范数;

3. 如果 v_i 还是标准正交的，那么 A^TA 是单位矩阵;

4. A 是正交矩阵，等价于 $A^TA = I$ 等价于 $v_1, \cdots, v_n$ 是 $\mathbb{R}^n$ 的一组标准正交基.

评论 1. 我们还介绍了几种特殊的矩阵:

1. 二维的旋转;

2. 二维的反射，形如 $I_2 - 2\omega\omega^T$，这里 ω 是一个单位向量;

3. 取 V 的一组标准正交基 $\underline{v_1, \cdots, v_r}$，则 $\underline{P_v = \sum_{i=1}^r v_iv_i^T}$.

$P_v = QQ^T = QIQ^T$
$= \sum_{i=1}^{r} v_iv_i^T$

说明标准正交基存在的方式是使用 Gram-Schmidt 正交化：

1. 先把 v_1 单位化，变为 v_1'；
2. 让 v_2 减去在 v_1' 方向的分量，再规范化，变为 v_2'；
3. 让 v_3 减去在 v_1', v_2' 张成平面的分量，即分别减去在这两个方向的分量，再规范化，变为 v_3'；
4. 一直这样做下去.

这里记公式是记不住的，但是把思路记下来就可以把公式补出来. 它有一个直接的推论就是：

定理 3. *任意可逆实矩阵都有唯一 QR 分解.*

2 行列式

2.1 综述

行列式的定义有着鲜明的几何意义，它可以从多个角度引入. 其中最直观且最初等理解还是把它看成单形的体积.

除此之外，并且常用的还有公理化定义法和逆序定义法，它们均能在不同情形下给我们带来方便.

注 3. *在线性代数中，行列式的应用主要体现在这样几个方面：*

1. $\boxed{\text{构造矩阵的逆}}$，*这需要借助于伴随矩阵.*
2. $\boxed{\text{判断向量组的相关性}}$. *给定 n 维 n 个数组向量*[1]*，把它们拼成一个矩阵 A，则这些数组向量线性无关等价于 $|A| \neq 0$.*
3. *判断矩阵的*$\boxed{\text{可逆性}}$[2]. *设 A 为 n 阶方阵，于是 A 可逆的充要条件是 $|A| \neq 0$.*
4. *求方程组的解.*
   (a) *设 A 为 n 阶矩阵，则 $AX = 0$ 只有零解等价于 $|A| \neq 0$;*

[1]这个对抽象向量也成立，只要取定一组基底，把它们等价为数组就可以.

[2]或者说叫满秩性.

特征值 求特征值 特征方程，取特殊值，tr(A)，|A|与λ的关系

(b) 设 A 为 n 阶矩阵，则 $AX=b$ 有唯一一个解等价于 $|A|\neq 0$.

5. 设矩阵 A 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_m$，则 $|A|=\lambda_1\lambda_2\cdots\lambda_m$.

注 4. 关于求解线性方程组的问题 (克拉默法则)，还有一些补充说明：

1. 只有方程数与变量数相等的时候才能用克拉默法则；
2. 齐次线性方程组有非零解与有无穷多解等价；
3. 系数行列式为 0 时，系数矩阵可能为 0，也可能不为 0，但不确定是哪一种情况，这个时候只能通过秩来判断.

例 2. 求解方程组

$$x_1+x_2+x_3=1$$
$$a_1x_1+a_2x_2+a_3x_3=0$$
$$a_1^2x_1+a_2^2x_2+a_3^3x_3=0$$

这里 a_1,a_2,a_3 互不相等.

2.2 计算技巧

2.2.1 $D+\alpha\beta^T$ 型

$(x-a)^{n-1}[x+(n-1)a]$

例 3. 计算 $D_n=\begin{vmatrix} x & a & \cdots & a \\ a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & x \end{vmatrix}$.

例 4. 计算 $D_n=\begin{vmatrix} x_1 & a & \cdots & a \\ a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & x_n \end{vmatrix}$.

例 5. 计算 $D_n=\begin{vmatrix} x_1 & a & \cdots & a \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ b & \cdots & b & x_n \end{vmatrix}$.

$\prod(x_i-a)$

$\prod(x_i-a)\left(1+\sum\frac{a}{x_i-a}\right)$.

例 6. 计算 $D_n = \begin{vmatrix} x_1 - a & x_2 & \cdots & x_n \\ x_1 & x_2 - a & \cdots & x_n \\ \vdots & \vdots & \ddots & x_n \\ x_1 & x_2 & \cdots & x_n - a \end{vmatrix}$.

2.2.2 三线型

例 7. 计算 $D_n = \begin{vmatrix} a & -1 & 0 & 0 \\ 1 & a & -1 & 0 \\ 0 & 1 & a & -1 \\ 0 & 0 & 1 & a \end{vmatrix}$.

2.2.3 范德蒙型

例 8. 计算 $V_n = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$.

注 5. 考场上出现的一般需要我们发现有范德蒙行列式的形式.

3 特征值与特征向量

3.1 特征值与特征向量的基本概念

回忆特征值，特征向量和特征方程，矩阵相似等概念与基本性质.

注 6. *1.* 矩阵的迹对应特征值的和，矩阵的行列式对应特征值的积；

2. 因此矩阵 A 可逆等价于特征值均非零；

3. 回忆矩阵特征值与矩阵多项式的特征值的关系.

定理 4. *1.* 属于不同特征值的特征向量线性无关；

2. 属于不同特征值的特征向量的线性组合不是特征向量；

3. 对称矩阵属于不同特征值的特征向量一定正交.

4. 实对称矩阵的特征值都是实数；

5. 实对称矩阵一定可以正交相似对角化.

3.2 对角化理论

3.2.1 对角化判定

注 7. 1. 对角化等价于有 n 个线性无关的特征向量；

2. 对角化等价于每个特征值的几何重数等于于代数重数，即特征值在特征多项式里出现的次数等于它对应的线性无关的向量个数；

3. 有 n 个不同特征值的矩阵一定可以对角化；

4. 实对称矩阵一定可以对角化；

5. 秩为 1 的矩阵一定可以对角化.

例 9. α, β 是三维正交的单位列向量，$A = \alpha\beta^T + \beta\alpha^T$，判断 A 是否可以对角化.

解答. $\alpha + \beta$ 与 $\alpha - \beta$ 都是 A 的特征向量，它们对应的特征值分别是 1 和 -1，并且由于 A 的秩小于等于 2，因此 0 也是 A 的特征值，从而 A 一定可以对角化。.

3.2.2 对角化过程

一般矩阵对角化；

1. 先从特征方程 $|\lambda I - A| = 0$ 接触 A 的特征值 $\lambda_1, \cdots, \lambda_m$；

2. 再由 $(\lambda_i I - A)X = 0$，求出 A 的线性无关的特征向量；

3. 把这些特征向量拼成 P，即可得 $P^{-1}AP$ 为对角矩阵，注意对角线上的特征值位置与你拼成 P 的方式有对应关系.

实对称矩阵对角化：

1. 先从特征方程 $|\lambda I - A| = 0$ 接触 A 的特征值 $\lambda_1, \cdots, \lambda_m$；

2. 再由 $(\lambda_i I - A)X = 0$，求出 A 的线性无关的特征向量；

3. 把这些特征向量先经过对角化，再经过规范化拼成正交矩阵 Q，即可得 $Q^{-1}AQ$ 为对角矩阵.

3.3 奇异值

定义 1. 任意 $m\times n$ 实矩阵 A，我们都可以把它写成矩阵乘积

$$A = U\Sigma V^T,$$

其中 U 是 m 阶正交矩阵，V 是 n 阶正交矩阵，Σ 是 $m\times n$ 阶矩阵，形如

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}_{m\times n},$$

并且 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$，$r = r(A)$.

注 8. 如果 $U = (u_1, \cdots, u_m)$，$V = (v_1, \cdots, v_n)$，则有

$$A = \sigma_1 u_1 v_1^T + \cdots \sigma_r u_r v_r^T.$$

事实上，我们计算奇异值分解，主要也是计算 u_i 与 v_i.

算法 1. 奇异值分解的计算过程：

1. 先计算非零奇异值. 这时候只要计算 A^TA 的特征值，然后开根号就可以.
2. 计算右奇异特征向量矩阵 V，V 的列向量会是 A^TA 的特征向量的规范正交化.
3. 用 $u_i = \dfrac{Av_i}{\sigma_i}$ 来计算左奇异向量.

注 9. 真正计算的时候可能会遇到一些问题：

1. 特征值出现 0 怎么办？不用管它，正常算右奇异向量，但是写奇异值时别把 0 写进去；
2. 如果发现右奇异向量比左奇异向量多怎么办？那不挺好，我们就取前面一部分用 $u_i = \dfrac{Av_i}{\sigma_i}$ 计算就可以.
3. 如果发现左奇异向量比右奇异向量多怎么办？用用 $u_i = \dfrac{Av_i}{\sigma_i}$ 算完之后，只能继续规范正交化.
4. 正是因为有前面那一点，因此我们常常取个转置使得列数大于等于行数.

4 二次型

4.1 理论

二次型可以粗糙地理解为多元二次多项式

$$f(x_1,\ldots,x_n)=X^TAX,$$

标准二次型是指只含有平方项不含有交叉项的二次型. 一般地，对于一个二次型 $f(X)=X^TAX$，经过可逆的线性变换 $X=PY$，我们都可以把它化为

$$f(X)=X^TAX=Y^T(P^TAP)Y=l_1y_1^2+\cdots+l_my_m^2,$$

这个称作二次型的标准化. 显然，标准二次型等价于系数矩阵 A 为对角矩阵，二次型标准化其实就是对称 A 的对角化. 注意这里我们是用 P^TAP 对角化，这个叫做矩阵的合同变换，类似矩阵的相似变换，我们也可以定义合同关系，这也是一种等价关系.

二次型的标准型不一定唯一，但是正负系数的个数是唯一确定的，这就是惯性定理，我们把正负系数的个数分别称作正负惯性指数，由此可以看出矩阵 A 与 B 合同的充要条件是 A 与 B 特征值中正数负数与零的个数相同.

正定二次型来源于对形如

$$f(x_1,x_2,x_3)=x_1^2+x_2^2+x_3^2$$

的式子的观察，它有如下特点：

1. 对于任意 $x_1.x_2,x_3$，有 $f(x_1,x_2,x_3)\geq 0$；
2. $f(x_1,x_2,x_3)=0$ 当且仅当 $x_i=0,\forall i$.

这类二次型不仅只有标准的这些，例如

$$f(x_1,x_2.x_3)=x_1^2-2x_1x_2+2x_2^2+x_3^2,$$

而对这类二次型的研究会帮助我们解决一些极值相关的问题，因此我们定义了正定二次型. 对于任意 $X\neq 0$，总有 $X^TAX>0$，就称 X^TAX 为正定二次型，称 A 为正定矩阵.

一般地，对于正定二次型，我们有这样的等价说法：

1. 二次型 X^TAX 为正定二次型的充要条件是 A 的特征值全为正数.
2. 二次型 X^TAX 为正定二次型的充要条件是 A 的顺序主子式全都大于 0，即
$$a_{11}>0,\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}>0,\cdots,|A|>0.$$
3. 对称矩阵 A 为正定矩阵的充要条件是存在可逆矩阵 B，使得 $A=B^TB$.
4. 对称矩阵 A 为正定矩阵的充要条件是 A 与单位矩阵 I 合同.
5. 对称矩阵 A 为正定矩阵的充要条件是 A 的正惯性指数为 n.
6. 对称矩阵 A 为正定矩阵的充要条件是做 LU 分解后，U 对角线上的元素全是正数.

注 10. *使用 LU 分解我们可以说明对称矩阵一定可以合同于对角型.*

注 11. *把 $>$ 改成 $\geq$ 可以得到半正定的定义，注意半正定的等价条件和正定基本一致，除了顺序主子式那一条需要改为全部的主子式.*

4.2 应用

与考试相关的主要应用是研究二次曲线：

注 12. *1. 一般地，判断是哪一种二次曲线，只要算一算行列式就可以；*

2. 求长轴和短轴只需要算特征值就可以，特别注意区分长轴和半长轴；

3. 如果还要你求长轴短轴方程，那就一定要把系数矩阵做正交合同，即用正交矩阵给它做合同变换.

5 线性变换理论

你需要掌握的是给定一个线性变换，写出它的一个变换矩阵：

例 10. *给定 2 阶矩阵*
$$A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix},$$

我们可以导出 2 阶矩阵构成的线性空间上的线性变换

$$\sigma(X) = AX,$$

写出 σ 的变换矩阵.

A 用标准型来看对角化

A.1 一个重要的不能对角化的例子

例 11. 形如 $\begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$ 的矩阵不可对角化，这个矩阵的阶数大于 1.

证明留作练习，考虑它的特征多项式就可以了. 这样的矩阵叫做约当块. 一个重要的定理是说，任意复矩阵都可以相似为一个约当矩阵，即形如

$$\begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & J_{n_k}(\lambda_k) \end{pmatrix}$$

的矩阵，这里

$$J_{n_t}(\lambda_t) = \begin{pmatrix} \lambda_t & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_t \end{pmatrix}_{n_t \times n_t}$$

是个约当块，n_t 表示这个约定块的阶数，λ_t 是对角线上的元素值. 注意，不同约当块的 λ_t 是可以相同的.

例 12. *1.* $\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{pmatrix}$ 是约当矩阵，它由 4 个一阶的约当块构成；

2. $\begin{pmatrix} 2 & & & \\ & 2 & 1 & \\ & & 2 & \\ & & & 4 \end{pmatrix}$ 是约当矩阵，它由一个特征值为 2 的一阶约当块，一个特征值为 2 的二阶约当块，和一个特征值为 4 的一阶约当块构成；

3. $\begin{pmatrix} 2 & & & \\ & 2 & 1 & \\ & & 4 & 1 \\ & & & 4 \end{pmatrix}$ 不是约当块，同一个约当块只能有一个特征值，而不同约定块之间不能有粘连的数.

一个矩阵可能有不同的约当标准型，但是它们之间的差别仅仅是调换了对角线上分块矩阵的次序——这样的两个矩阵当然是相似的.

例 13.

$$\begin{pmatrix} 1 & & \\ & 2 & 1 \\ & & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 & \\ & 2 & \\ & & 1 \end{pmatrix}$$

是相似的，它们也是同一个矩阵的约当标准型.

A.2 任意复矩阵的对角化结果

首先是一个简单却重要的结果：

命题 1. 分块对角矩阵 $A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix}$ 可对角化当且仅当每个 A_i 可对角化.

于是我们考虑复矩阵可对角化的问题时，可以先考虑这个复矩阵的约当标准型，再考虑这个标准型的每一个约定块，只要出现阶数大于 1 的约定块，就可以认定它不可以对角化. 这衍生了出一些等价条件，线代 1 中学过的主要是两条：

- n 阶复矩阵可对角化等价于有 n 个线性无关的特征向量；
- n 阶复矩阵可对角化等价于某个特征值的几何重数小于代数重数.

A.3 特征向量的等价条件.

你可以很容易构造出可对角化矩阵的所有特征向量，有 n 个线性无关的特征向量就一定可以对角化也很简单——该矩阵对应的线性变换在这组特征向量下的基底就是对角矩阵. 而线性变换在不同基底下的矩阵是相似的，因为变换基底等价于矩阵的相似. 可能还有同学不知道我在说什么，但等上完这门课后你就会深刻地理解这里每个字的含义.

我们也可以从约当标准型的角度理解，每一个约当块都只有一个特征向量，对于阶数大于 1 的约当块来说，它就不可能有和阶数一样多的线性无关的特征向量. 把这个结果延伸到约当标准型需要对分块对角矩阵有更深刻的认识，这里暂先略去.

如果你仍然想了解更多的话，可以去算一算 $\begin{pmatrix} 1 & & \\ & 2 & 1 \\ & & 1 \end{pmatrix}$ 的特征向量，观察它是如何受到阶数大于一的约当块影响的.

A.4 几何重数与代数重数

所谓代数重数，是指特征值在特征多项式中作为根的重数，用约当标准型的话说，它是这一特征值对应的约当块的阶数之和.

例 14. *1.* $\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{pmatrix}$ 的代数重数是 4;

2. $\begin{pmatrix} 2 & & & \\ & 2 & 1 & \\ & & 2 & \\ & & & 4 \end{pmatrix}$ 的代数重数是 3.

而矩阵 A 的特征值 λ 的几何重数定义为 $(\lambda I - A)$ 解空间维数，这个看起来很复杂，其实用约当标准型的话来说，几何重数就是 λ 对应的约当块个数——因为只有 λ 对应的约当块才对 $(\lambda I - A)$ 的解空间维数有贡献，并且一个约当块的贡献只有 1.

例 15. *1.* $\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{pmatrix}$ 的几何重数是 4;

2. $\begin{pmatrix} 2 & & & \\ & 2 & 1 & \\ & & 2 & \\ & & & 4 \end{pmatrix}$ 的几何重数是 2.

于是我们可以发现，某个特征值的几何重数小于代数重数事实上等价于有阶数大于 1 的约定块，因此我们可以使用这个方法去判断能否对角化.

B 线性映射视角下的矩阵问题

B.1 从线性映射出发得到矩阵

我们先从一个最简单的问题开始,给你一个线性变换$\mathcal{A}$ 和一组基 $e_1, \ldots, e_n$，你该如何写出它的矩阵？

这当然是个非常平凡的问题，但我相信现在仍然有同学做不出来，所以我仍然想多花点时间把这个再复习一遍. 设

$$\mathcal{A}e_i = a_{1i}e_i + \cdots + a_{ni}e_n,$$

把它写成

$$\mathcal{A}e_i = (e_1, e_2, \cdots, e_n)\begin{pmatrix} a_{1i} \\ a_{2i} \\ \cdots \\ a_{ni} \end{pmatrix},$$

再将所有的 e_i 拼一起，就有

$$\mathcal{A}(e_1, e_2, \cdots, e_n) = (\mathcal{A}e_1, \cdots, \mathcal{A}e_n) = (e_1, e_2, \cdots, e_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

一般我们令 $A=(a_{ij})$，于是上式可简记为

$$\mathcal{A}(e_1,e_2,\cdots,e_n)=(e_1,e_2,\cdots,e_n)A.$$

大部分学过线代 1 的同学，你可能对这个式子仍然有些疑问，比如为什么写成

$$\mathcal{A}e_i=(e_1,e_2,\cdots,e_n)\begin{pmatrix}a_{1i}\\a_{2i}\\\cdots\\a_{ni}\end{pmatrix},$$

而不是

$$\mathcal{A}e_i=(a_{1i},a_{2i},\cdots,a_{ni})\begin{pmatrix}e_1\\e_2\\\cdots\\e_n\end{pmatrix},$$

假如是这样，我们就有

$$\mathcal{A}(e_1,e_2,\cdots,e_n)=A(e_1,e_2,\cdots,e_n),$$

初看没啥，你甚至觉得更自然更协调，但是当你考虑到两个线性变换复合的时候，你就发现

$$\mathcal{A}_1\mathcal{A}_2(e_1,e_2,\cdots,e_n)=\mathcal{A}_1A_2(e_1,e_2,\cdots,e_n),$$

欸，$\mathcal{A}_1A_2(e_1,e_2,\cdots,e_n)$ 是什么，我只记得线性变换怎么作用到向量上，现在它和矩阵撞上了，我该如何协调它们呢？当你终于想起来了线性映射的线性性——这一条决定矩阵可以从线性映射中“拿出去”，就有

$$\mathcal{A}_1\mathcal{A}_2(e_1,e_2,\cdots,e_n)=A_2\mathcal{A}_1(e_1,e_2,\cdots,e_n)=A_2A_1(e_1,e_2,\cdots,e_n),$$

矩阵与线性变换反序了！这是我们不希望看见的事情. 这种记号还有一些其它的好处，我们可以在学习和使用的使用慢慢体会.

顺便说一句,基于这种记号,为了让 e_i 为数组向量的时候,$(e_1,e_2,\cdots,e_n)$ 可以成为矩阵，所以我们一般令其为列向量，这也正是数学家们更喜欢列向量的原因.

例 16. 作为一个例子，我们来解释一个事实，为什么复数 $a+b\sqrt{-1}$ 可以等同于实矩阵 $\begin{pmatrix}a&-b\\b&a\end{pmatrix}$. 注意 $\mathbb{C}$ 是 $\mathbb{R}$ 上以 $1,\sqrt{-1}$ 为基底的二维线性空间，

我们把乘上 $a+b\sqrt{-1}$ 理解成一个线性变换[3]，于是有

$$a+b\sqrt{-1}(1,\sqrt{-1})=(a+b\sqrt{-1},a\sqrt{-1}-b)=(1,\sqrt{-1})\begin{pmatrix}a & -b\\ b & a\end{pmatrix}.$$

这种把某些数学对象理解为某个线性空间上的线性变换的手段叫做表示，一些物理系的同学可能会对它更感兴趣[4].

现在我们来回顾下怎么用这种记号去理解基变换. 仿造前面的，由于新的基底 e_i' 在旧的基底下面有线性表示

$$e_i'=p_{1i}e_1+\cdots+p_{ni}e_n,$$

因此

$$e_i'=(e_1,\cdots,e_n)\begin{pmatrix}p_{1i}\\ p_{2i}\\ \cdots\\ p_{ni}\end{pmatrix},$$

再令 $P=(p_{ij})$，同样有

$$(e_1',\cdots,e_n')=(e_1,\cdots,e_n)P.$$

与前面不同的是，前面的 A 可以不可逆，但是这里的 P 一定可逆. 最后我们再来说下线性映射，先请大家回忆一下，线性映射和线性变换有什么区别？简单来说，线性映射是从 V_1 到 V_2 的箭头，它的源 (source) 和靶 (target) 可以是两个不同的线性空间. 而线性变换一定是一个相同的线性空间上的，是 V 到自己的箭头.

我们来试着写下线性映射的矩阵，通过这个你也可以看看线性映射与线性变换的区别.

给定 V_1 的一组基底 $(e_1,\cdots,e_n)$，V_2 的一组基底 $(e_1',\cdots,e_m')$ 与线性映射 $\mathcal{A}:V_1\to V_2$. 通过前面的那么话，相应你已经明白构造矩阵的精髓在于把基底的像用新的基底表示出来. 我们把 $\mathcal{A}e_i$ 用 V_2 的基底表示出来，有

$$\mathcal{A}e_i=a_{1i}e_1'+\cdots+a_{mi}e_m',$$

[3]可以严格证明这一事实.

[4]所以你们更要学好这门课.

于是我们可以得到

$$\mathcal{A}e_i = (e'_1, \cdots, e'_m)\begin{pmatrix} a_{1i} \\ a_{2i} \\ \cdots \\ a_{mi} \end{pmatrix},$$

把它们拼成一个矩阵，就可以得到

$$\mathcal{A}(e_1, e_2, \cdots, e_n) = (\mathcal{A}e_1, \cdots, \mathcal{A}e_n) = (e'_1, e'_2, \cdots, e'_m)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

关注这里 m 与 n，这说明一般的 $m \times n$ 的非方阵本质上是线性映射的矩阵.

B.2 矩阵运算

我们先来解释一件事情，矩阵的乘法是怎么构造出来的？为什么我们不把矩阵的乘法定义为对应位置的元素之积？为什么我们要选择这么别扭的乘法定义？

给定两个线性映射 $\mathcal{A}_1 : V_1 \to V_2$ 与 $\mathcal{A}_2 : V_2 \to V_3$，它们对应矩阵分别是 $A_1 = (a_{ij})_{m\times n}$ 与 $A_2 = (b_{jk})_{t\times m}$，并且 V_1 有基底 $e_i (1 \le i \le n)$，V_2 有基底 $e'_j (1 \le j \le m)$，V_3 有基底 $e"_k (1 \le k \le t)$，于是

$$\mathcal{A}_1 e_i = \sum_{j=1}^{m} a_{ji} e'_j = a_{1i} e'_1 + \cdots + a_{mi} e'_m$$

$$\mathcal{A}_2 e'_j = \sum_{k=1}^{t} a_{kj} e"_k = a_{1j} e"_1 + \cdots + a_{tj} e"_t$$

因此

$$\mathcal{A}_2\mathcal{A}_1 e_i = \mathcal{A}_2(\sum_{j=1}^{m} a_{ji} e'_j) = \sum_{j=1}^{m} a_{ji}\mathcal{A}_2 e'_j = \sum_{j=1}^{m} a_{ji} \sum_{k=1}^{t} a_{kj} e"_k = \sum_{k=1}^{t}(\sum_{j=1}^{m} a_{ji} a_{kj}) e"_k,$$

令一方面

$$\mathcal{A}_2\mathcal{A}_1(e_1, \cdots, e_n) = \mathcal{A}_2(e'_1, \cdots, e'_m)A_1 = (e"_1, \cdots, e"_t)A_2A_1,$$

回忆前面构造矩阵的过程，矩阵第 i 行第 j 列的元素 a_{ij} 本质上是 $\mathcal{A}e_j$ 在 e'_i 前的系数，于是我们如果把 A_2A_1 看成一个新矩阵 $(c_{ki})_{t\times n}$ 的话，那么一定有

$$c_{ki}=\sum_{j=1}^{m}a_{kj}a_{ji}.$$

习题 1. 这个练习仅为那些对这些推导感觉晕头转向的同学准备，你可以取 $m=n=t=2$。然后把前面的过程重新写一遍.

习题 2. 模仿前面的过程，指出矩阵加法的几何含义，特别思考加法为什么一定是同型矩阵相加.

给定一组基之后，我们可以把线性映射的运算转移到矩阵上. 比如

$$(\mathcal{A}_1+\mathcal{A}_2)(e_1,e_2,\cdots,e_n)=(e'_1,e'_2,\cdots,e'_m)(A_1+A_2)$$
$$\mathcal{A}_2\mathcal{A}_1(e_1,\cdots,e_n)=(e"_1,\cdots,e"_t)A_2A_1.$$

并且注意到，线性映射和矩阵是一一对应的. 这说明[5]给定基底后，线性映射构成的线性空间与矩阵构成的线性空间是线性同构[6]. 我们这里并不打算严格地去证明它，否则还要讲清楚它们为什么能构成线性空间等等. 列出这个事实仅是为了再次强调一个观念，线性映射和矩阵是一体两面的两个数学对象. 再举一个经典的例子，矩阵的相似相信大家印象深刻，线代 1 最难的对角化问题就是用相似的语言给出来的. 其实相似的矩阵本质上都是某个线性变换[7]在不同基下的矩阵.

给定一个线性变换 $\mathcal{A}$ 与基底变换矩阵 P 使得

$$(e'_1,\cdots,e'_n)=(e_1,\cdots,e_n)P,$$

设

$$\mathcal{A}(e_1,\cdots,e_n)=(e_1,\cdots,e_n)A,\qquad \mathcal{A}(e'_1,\cdots,e'_n)=(e'_1,\cdots,e'_n)A',$$

即 $\mathcal{A}$ 在新基底与旧基底下的矩阵分别为 A 与 A'，把基变换公式代进去，有

$$\mathcal{A}(e_1,\cdots,e_n)P=(e_1,\cdots,e_n)PA',$$

[5]开始套娃，但事实上正是因为能进行这种套娃，代数学才有了美感与生命力.

[6]这是一种抽象废话——它仅是对已有结论的抽象概括，本质上并没有什么新的东西. 但很多数学学家都偏爱抽象废话，如果你们以后会阅读一些代数学或者其它数学的书籍，一定要有把抽象废话转化为例子和直观的能力.

[7]注意，我们只对方阵考虑相似，所以它的几何背景是线性变换.

两边右乘上 P^{-1}，有

$$\mathcal{A}(e_1, \cdots, e_n) = (e_1, \cdots, e_n)PA'P^{-1},$$

这里左边其实就是 $(e_1, \cdots, e_n)A$，利用 $(e_1, \cdots, e_n)$ 的线性无关性，我们可以得到 $A = PA'P^{-1}$.

习题 3. *如果你对前面如何利用线性无关性得到两个矩阵相等仍然迷惑不解的话，不妨来做做这道习题. 给定基底 $(e_1, \cdots, e_m)$ 与矩阵 $A = (a_{ij})_{m\times n}$，证明：*

1. $(e_1, \cdots, e_m)A = (\sum_{i=1}^m a_{i1}e_i, \cdots, \sum_{i=1}^m a_{in}e_i)$；
2. 如果 $(e_1, \cdots, e_m)A_1 = (e_1, \cdots, e_m)A_2$,那么 $(e_1, \cdots, e_m)(A_1 - A_2) = 0$；
3. 如果 $(e_1, \cdots, e_m)A = 0$，那么 A 的每个元素都是 0，从而 A 是零矩阵.

习题 4. *设线性变换 $\mathcal{A}$ 在基底 $(e_1, \cdots, e_n)$ 下的矩阵为 A，P 是一个可逆矩阵，并且 $A' = PAP^{-1}$，构造一组基底，使得 $\mathcal{A}$ 在新基底下的矩阵是 A'.*

注 13. *现在你应该可以明白很多事情了，比如为什么相似的矩阵有着相同的特征值、特征多项式、最小多项式这些东西？因为它们本质上都是对线性变换定义的，矩阵能有这些定义不过是因为它与线性变换的一体两面罢了.*

相比于相似，你可能对等价这一概念会陌生一些：

定义 2. *我们称矩阵 A_1 与 A_2 等价，是说存在可逆矩阵 P 与 Q，使得 $A_1 = PA_2Q$.*

显然这里 A_1 与 A_2 不必是方阵，它们只要是同型的就可以[8]，于是我们考虑线性映射能否作为它的几何背景.

给定线性映射 $\mathcal{A} : V_1 \to V_2$，这里 V_1 有基底 $(\alpha_1, \cdots, \alpha_n)$，$V_2$ 有基底 $(\beta_1, \cdots, \beta_m)$，并且

$$\mathcal{A}(e_1, \cdots, e_n) = (\beta_1, \cdots, \beta_m)A_1,$$

我们把 $A_1 = PA_2Q$ 代入，则有

$$\mathcal{A}(\alpha_1, \cdots, \alpha_n) = (\beta_1, \cdots, \beta_m)PA_2Q,$$

[8] 事实上，同型矩阵只要有相同的秩就是等价的.

继续化简这个式子，有

$$\mathcal{A}(\alpha_1,\cdots,\alpha_n)Q^{-1}=(\beta_1,\cdots,\beta_m)PA_2,$$

我们令 $(\alpha_1',\cdots,\alpha_n')=(\alpha_1,\cdots,\alpha_n)Q{*}{-1}$,再令 $(\beta_1',\cdots,\beta_n')=(\beta_1,\cdots,\beta_m)P$，于是

$$\mathcal{A}(\alpha_1',\cdots,\alpha_n')=(\beta_1',\cdots,\beta_n')A_2.$$

你发现等价也是一种基变换，这点你并不意外，因为你早已熟知相似的矩阵也是等价的.

B.3 矩阵证明与映射证明

时间原因，我们仅举一个经典例子，你可以回忆线代 1 的很多结果，并尝试使用线性空间的方法证明它们：

例 17. 设 A,B 为 n 阶矩阵，并且 $AB=BA$，证明：

$$rank(A)+rank(B)\geq rank(A+B)+rank(AB).$$

先把它翻译成线性变换的语言，注意矩阵 A 的秩其实就是对应的线性变换 $\mathcal{A}$ 的维数，于是我们只要在 $\mathcal{AB}=\mathcal{BA}$ 的前提下，证明

$$dim(Im(\mathcal{A}))+dim(Im(\mathcal{B}))\geq dim(Im(\mathcal{A}+\mathcal{B}))+dim(Im(\mathcal{AB})).$$

法一. 注意

$$ker(\mathcal{A})\subset ker(\mathcal{BA}),\qquad ker(\mathcal{B})\subset ker(\mathcal{AB}),$$

于是

$$ker(\mathcal{A})+ker(\mathcal{B})\subset ker(\mathcal{AB})=ker(\mathcal{BA}).$$

并且还有

$$ker(\mathcal{A})\cap ker(\mathcal{B})\subset ker(\mathcal{A}+\mathcal{B}).$$

这得出

$$\begin{aligned}dim(ker(\mathcal{A}))+dim(ker(\mathcal{B}))&=dim(ker(\mathcal{A})+ker(\mathcal{B}))+dim(\mathcal{A}\cap\mathcal{B})\\&\leq dim(ker(\mathcal{AB}))+dim(ker(\mathcal{A}+\mathcal{B})),\end{aligned}$$

利用核与像的维数和等于全空间维数 n，于是有

$$n-dim(Im(\mathcal{A}))+n-dim(Im(\mathcal{B}))\leq n-dim(Im(\mathcal{AB}))+n-dim(Im(\mathcal{A}+\mathcal{B})),$$

化简可得.

看吧，这就是线性空间的解法，多么自然且优美，解决看似困难的问题仍然举重若轻. 反观线代 1 中纯粹矩阵论的解法，就流于技巧化了：

法二. 利用分块矩阵乘法，有

$$\begin{pmatrix} I & I \\ 0 & I \end{pmatrix}\begin{pmatrix} A & O \\ O & B \end{pmatrix}\begin{pmatrix} I & O \\ I & I \end{pmatrix} = \begin{pmatrix} A+B & B \\ B & B \end{pmatrix},$$

再利用 $AB = BA$，有

$$\begin{pmatrix} I & O \\ B & -A-B \end{pmatrix}\begin{pmatrix} A+B & B \\ B & B \end{pmatrix} = \begin{pmatrix} A+B & B \\ O & -AB \end{pmatrix},$$

这说明

$$rank(A)+rank(B) = rank\begin{pmatrix} A+B & B \\ B & B \end{pmatrix} \geq \begin{pmatrix} A+B & B \\ O & -AB \end{pmatrix} \geq rank(A+B)+rank(AB).$$