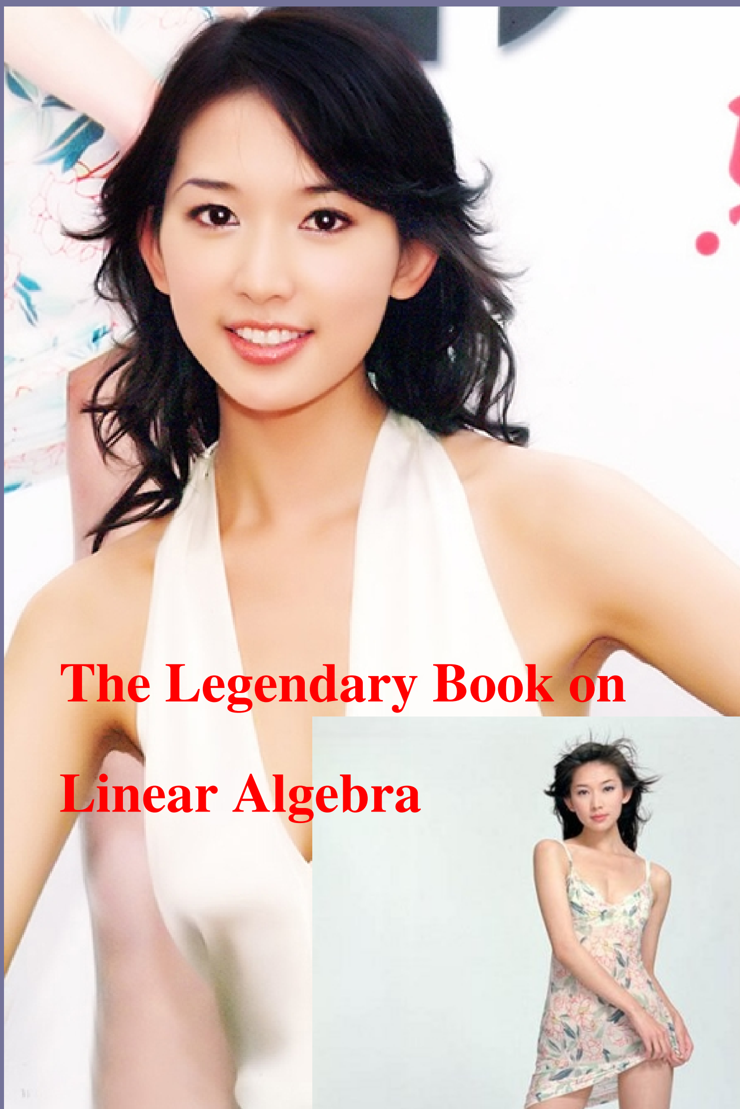
高等代数葵花宝典



The Legendary Book on Linear Algebra

目 录

第零章　番外话 1

第一章　将打洞进行到底 2

第二章　Jordan 标准形总结 7

第三章　秩不等式 12

第四章　交结数：刻画相似程度的不变量 16

第五章　同时上三角化 19

第六章　覆盖定理 23

第七章　有理标准形和交换的矩阵 25

第八章　解题的艺术 30

0

番外话

先说一件很囧的事。两年前我给北京大学化学学院一年级的学弟学妹们上高数的习题课。开学第一次课来了三十个人，到期末的最后一次课只剩下十三个人。虽说习题课不管讲的好坏都拿那份钱，学生也不会拿鸡蛋西红柿拍你，但是看着来上课的人越来越少确实对自尊是一种打击。特别让我印象深刻的是一个相貌气质都很不错的MM（那十三个人之一），她每次课都在下面很认真的听，很安静，整个学期她只站起来问了三次问题，但是每一次都把我问倒了。很显然这是对我的进一步打击。我很无奈的承认自己不配拿那份津贴，就转行做了本学院高等代数课的助教。这次给我打击的是另一个很清纯的 MM。有很多次我在黑板上出了题目，然后微笑着、踱着步子显示高深莫测的时候，她都举手表示已经做出了答案。接下来我只能用凝固的微笑和景仰的目光看着她在全班面前用柔柔的声音解释如何如何。不过总的来说，我还是成功的 Hold 住了局面，当时一个学年下来到课人数无明显下滑。

习题课上多了，自己也有一些体会。讲课跟做题是不一样的，你必须脑子里时刻清楚自己在讲什么，接下来要讲什么，然后把它们用平缓的节奏一遍讲正确。你讲的语气速度快了，或者思维有了跳跃，学生一下跟不上，那么你后面的内容他们听起来都很茫然。当我一时不知道说什么好的时候，我会面色如常地擦擦黑板，换换粉笔，整理一下自己的思路，绝不轻易开口。因为如果你不小心说错了话，那比没说要糟糕一百倍：接下来你要用十句话来挽救你的错误，学生很可能就被绕晕了。即使是"嗯"、"啊"、"那么"这些口头禅，也会暴露你的思路的紊乱。高深莫测永远是 Hold 局面的不二法宝。我曾经开玩笑地给学生说，我讲课有一个优点，就是从来没有口头禅。结果大家都笑了。我不解，然后大家异口同声的告诉我：老师，你讲课有一个口头禅，就是"很显然"（囧）。希望我在这个文档里没有再犯这个错误 :P。

本文档脱胎于以前的同名文档，经过多次修改以后与最初的版本相比已经面目全非。但是变薄变精炼的趋势一直没有改变。那些武侠小说中出现的秘笈宝典，几乎无一例外都是"薄薄的一本小册子"，因为浓缩的才是精华。本文档也照此看齐，不求全，但求精致，通过几个专题来体现高等代数的方法和想法。还是那句话，与其炖上一锅大杂烩，倒不如几样精致的小菜来得有滋味。至于纯粹为难而难，或者为收录而收录的内容，就不在考虑之列了。文档薄一点，也是为了激发大家速成的欲望。

本文档是本人心血之作，也算经过了教学的实践检验，因此我相信质量不会太糟，但是错误恐怕仍然难以避免。欢迎大家来信指教：xidalapuda@126.com

1

将打洞进行到底

之所以把这一章作为整本书的开始，是因为打洞是矩阵里面最基本最重要的技巧，江湖上出来混的没有不知道的，所以怎样强调它的重要性也不过分。下面这个例子就很好地说明了什么是打洞。

定理 1.1. 设

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

是一个方阵，其中 A 是可逆的子方阵，那么

$$|M|=|A|\cdot|D-CA^{-1}B|.$$

结论不难记，从 D 出发顺时针走一圈就可以了。

证明. 思路就是利用 A 的可逆性来打洞，干掉 B,C 之一：

$$\begin{pmatrix}A & B\\ C & D\end{pmatrix}\xrightarrow{\text{第一行左乘以 } -CA^{-1} \text{ 加到第二行}}\begin{pmatrix}A & B\\ 0 & D-CA^{-1}B\end{pmatrix},$$

也就是

$$\begin{pmatrix}I_n & 0\\ -CA^{-1} & I_m\end{pmatrix}\begin{pmatrix}A & B\\ C & D\end{pmatrix}=\begin{pmatrix}A & B\\ 0 & D-CA^{-1}B\end{pmatrix},$$

两边求行列式即可。 □

类似地，D 可逆的时候结论变成 $|M|=|D|\cdot|A-BD^{-1}C|$（从 A 出发顺时针走一圈）。打洞说白了就是一个降阶的过程。

注意到如果把上式写成

$$\begin{pmatrix}A & B\\ C & D\end{pmatrix}=\begin{pmatrix}I_n & 0\\ -CA^{-1} & I_m\end{pmatrix}^{-1}\begin{pmatrix}A & B\\ 0 & D-CA^{-1}B\end{pmatrix},$$

这就很像一个分块的 LU 分解。其实真正的 LU 分解和这个是一回事，这里就不具体写了。

如果把打洞的过程倒过来用，就是提升：

定理 1.2. 设 A 是 $n\times m$ 矩阵，B 是 $m\times n$ 矩阵，则 AB 和 BA 的特征多项式只差一个因子 λ^{n-m}，即

$$\lambda^m|\lambda I_n-AB|=\lambda^n|\lambda I_m-BA|.$$

只需要对 $\lambda \neq 0$ 证明即可。我们先证明 $\lambda = 1$ 的时候结论成立，也就是 $|I_m - AB| = |I_n - BA|$ 成立。这只要在矩阵

$$\begin{pmatrix} I_m & B \\ A & I_n \end{pmatrix}$$

中分别用 I_m 和 I_n 各打一次洞就可得证：

$$\begin{vmatrix} I_m & B \\ A & I_n \end{vmatrix} = |I_n| \cdot |I_m - BA| = |I_m| \cdot |I_n - AB|.$$

对于一般的 $\lambda \neq 0$，只要在等式 $|I_m - AB| = |I_n - BA|$ 中用 A/λ 替换 A 即可。

1.1 对称矩阵的打洞

打洞有很多重要的应用，特别是当 M 是对称矩阵的时候，如果你用 A 打两次洞干掉 B 和 B' 就会发现这恰好是一个合同变换：

$$\begin{pmatrix} I_n & 0 \\ -A^{-1}B & I_m \end{pmatrix}' \begin{pmatrix} A & B \\ B' & D \end{pmatrix} \begin{pmatrix} I_n & -A^{-1}B \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & D - B'A^{-1}B \end{pmatrix}.$$

特别强调的是，对称矩阵的打洞有特别重要的意义：由于 M 可以看作一个“内积”的度量矩阵，所以两边打洞实际上就是在这个“内积”下做 Schmidt 正交化，化二次型为标准形的配方法和矩阵法都源自于此。这里简要描述一下矩阵法，详细的叙述请查阅教科书。

定理 1.3 (化二次型为标准形的算法). 设 $A = (a_{ij})$ 是一个 n 阶对称矩阵，现在要把它合同为对角形。

- 如果 $a_{11} \neq 0$，那就用 a_{11} 两次打洞合同掉第一行和第一列的其它元素，把 A 变成
$$\begin{pmatrix} a_{11} & 0 \\ 0 & * \end{pmatrix},$$
然后考虑右下角的 $n-1$ 阶的矩阵 $*$。

- 如果 $a_{11} = 0$ 但是某个 $a_{ii} \neq 0$，那就交换第 i 行和第 1 行，交换第 i 列和第 1 列，把 a_{ii} 变到 a_{11} 的位置上来，然后返回上一步。

- 如果 A 的对角线上都是 0，但是某个 a_{ij} 不是 0，那就把第 j 行加到第 i 行，第 j 列加到第 i 列，这样 a_{ii} 的位置上就出现了 $2a_{ij}$，然后返回上一步。

这样经过有限步以后就可以把 A 变成对角形。

这个算法说白了就是一句话：制造非零的对角元来干掉非对角元，其实就是不断地做 Schmidt 正交化。

正定矩阵是最容易化为标准形的对称矩阵，因为正定矩阵的对角元总不是 0（想一想，为什么？），所以只需要第一个步骤就可以化为标准形。半正定矩阵的打洞也很简单，虽然对角元可能出现 0，但是我们有下面的引理：

引理 1.4. 如果半正定矩阵 A 的某个对角元是 0，那么该对角元所在的行和列所有元素都是 0。

证明. 由于 A 半正定，所以有平方根分解 $A = P'P$。记 $P = (v_1, v_2, \dots, v_n)$，则 $a_{ij} = (v_i, v_j)$，这里的 (v_i, v_j) 表示 v_i 和 v_j 的通常的欧式内积。$a_{ii} = 0$ 说明 $v_i = 0$，从而第 i 行第 i 列都是 0。 □

可见半正定矩阵化为标准形本质上也只需要步骤 1，只不过对角线上遇到 0 的时候不用打洞，自动跳过去继续考虑右下角的矩阵。

接下来是引理 1.4 的两个应用：

定理 1.5. 设 A 是一个实对称矩阵，$\lambda_{\min}$ 和 $\lambda_{\max}$ 为 A 的最小和最大的实特征值，a_{ii} 是 A 的任一对角元，则有

$$\lambda_{\min} \leq a_{ii} \leq \lambda_{\max},$$

而且两个不等号只要有一个成立则 a_{ii} 所在的行和列的其它元素就必然都是 0。

证明. 只要对 $A + \lambda_{\min} I$ 和 $\lambda_{\max} I - A$ 这两个半正定矩阵应用引理 1.4 即可。 □

定理 1.6 (两半正定矩阵同时合同于对角形). 设 A, B 是两个 n 阶半正定矩阵，则存在可逆矩阵 T 使得 $T'AT, T'BT$ 都是对角矩阵。

证明. 首先做合同变换把 A 化成标准形

$$A \sim \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

这时 B 仍然是半正定的（虽然 B 也发生了变化），所以不妨从一开始就假设 A 就是如上的标准形，并设

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \qquad B_{12} = B_{21}',$$

我们要在保持 A 的形状的前提下把 B 化成标准形。

设正交矩阵 Q 使得

$$Q'B_{22}Q = \begin{pmatrix} I_s & 0 \\ 0 & 0 \end{pmatrix},$$

那么用矩阵

$$\begin{pmatrix} I_r & 0 \\ 0 & Q \end{pmatrix}$$

作合同变换保持 A 不变，把 B 化为形如

$$\widetilde{B} = \begin{pmatrix} B_{11} & * & 0 \\ * & I_s & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

的矩阵。注意这里已经利用了引理 1.4 的结论，由于 $\widetilde{B}$ 的最后一个对角元是零矩阵，所以它的最后一行和最后一列中的矩阵都是 0。这个时候再用 I_s 打洞消去 "$*$" 的部分，这还是一个不影响 A 的合同变换，这就把 A, B 同时变成了准对角形，最后再用一次正交变换就可以了。 □

1.2 正定矩阵

正定矩阵的另一个名字是内积的度量矩阵，永远不要忘记这一点。正定矩阵几乎所有结论都有对应的几何解释，所以只要你搞清楚这些结论的几何意义，正定矩阵其实就是一个很简单的东东。

设 $v_1, v_2, \ldots, v_n$ 是 $\mathbb{R}^n$ 的一组基，那么矩阵

$$A = \begin{pmatrix} (v_1, v_1) & (v_1, v_2) & \cdots & (v_1, v_n) \\ (v_2, v_1) & (v_2, v_2) & \cdots & (v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ (v_n, v_1) & (v_n, v_2) & \cdots & (v_n, v_n) \end{pmatrix}$$

就是一个正定矩阵。反过来，每一个正定矩阵都有如上的表示形式。很显然，A 刻画了向量组 $v_1 \sim v_n$ 的长度以及它们之间的相互夹角，所以不难想象 $v_1 \sim v_n$ 的一些几何性质可以用 A 的代数性质来描述。反过来，如果有人问你正定矩阵的代数性质，你也要立刻想到它对应的几何解释。

举几个例子：

- 正定矩阵的对角元都不是零。这是显然的，因为 a_{ii} 代表 v_i 的长度的平方，当然不能是零。
- 正定矩阵中最大的元素必然出现在对角线上。这是因为内积满足 Schwatz 不等式 $(v_i, v_j)^2 \leq (v_i, v_i)(v_j, v_j)$，即 $a_{ij}^2 \leq a_{ii}a_{jj}$，从而 $a_{ij} \leq \max\{a_{ii}, a_{jj}\}$。
- 正定矩阵的行列式的值等于 $v_1, v_2, \ldots, v_n$ 张成的平行多面体的体积的平方。正定矩阵的主子式都大于零，这是因为主子式 $A_{i_1 i_2 \cdots i_m}$ 的值是 $v_{i_1}, v_{i_2}, \ldots, v_{i_m}$ 张成的平行多面体的体积的平方，所以大于零。

例 1.7. 设 A 是 n 阶正定矩阵，求证 $|A| \leq a_{11}a_{22}\cdots a_{nn}$，等号成立当且仅当 A 是对角矩阵。

这个结论的几何解释就是：平行多面体的体积不大于各个棱长的乘积，当且仅当各棱垂直的时候等号成立。

证明. 用归纳法，假设 $n-1$ 的时候结论成立。设

$$A = \begin{pmatrix} A_{n-1} & \alpha \\ \alpha' & a_{nn} \end{pmatrix},$$

则 $|A| = |A_{n-1}| \cdot |a_{nn} - \alpha' A_{n-1}^{-1}\alpha|$。注意 $0 < a_{nn} - \alpha' A_{n-1}^{-1}\alpha \leq a_{nn}$，所以使用归纳假设即可。等号成立的条件也不难证明。 □

实际上 $a_{nn} - \alpha' A_{n-1}^{-1}\alpha$ 这个量也是有它的几何解释的。我们来这样分析：记 $v_1 \sim v_{n-1}$ 张成的底面为 $\mathbb{P}$，v_n 可以分解为垂直于 $\mathbb{P}$ 的分量和属于 $\mathbb{P}$ 的分量的和：

$$v_n = v_n^{\shortparallel} + v_n^{\perp}, \qquad v_n^{\shortparallel} \in \mathbb{P}，v_n^{\perp} \perp \mathbb{P}.$$

那么设

$$v_n^{\shortparallel} = v_n - v_n^{\perp} = x_1 v_1 + \cdots + x_{n-1}v_{n-1},$$

两边依次用 $v_1, \ldots, v_{n-1}$ 作内积，我们得到这样一个方程组：

$$\begin{aligned}
(v_1, v_n) &= x_1(v_1, v_1) + \cdots + x_{n-1}(v_1, v_{n-1}), \\
(v_2, v_n) &= x_1(v_2, v_1) + \cdots + x_{n-1}(v_2, v_{n-1}), \\
&\cdots\cdots\cdots \\
(v_{n-1}, v_n) &= x_1(v_{n-1}, v_1) + \cdots + x_{n-1}(v_{n-1}, v_{n-1}).
\end{aligned}$$

采用上面例题中的记号，这个方程组就是 $\alpha = A_{n-1}X$，所以

$$\begin{aligned}
v_n^{\shortparallel} &= (v_1, v_2, \ldots, v_{n-1})X = (v_1, v_2, \ldots, v_{n-1})A_{n-1}^{-1}\alpha \\
&\Rightarrow \|v_n^{\shortparallel}\|^2 = \Big((v_1, v_2, \ldots, v_{n-1})A_{n-1}^{-1}\alpha\Big)'\Big((v_1, v_2, \ldots, v_{n-1})A_{n-1}^{-1}\alpha\Big) \\
&= \alpha' A_{n-1}^{-1}\alpha.
\end{aligned}$$

根据勾股定理，$\|v_n\|^2 = \|v_n^{\shortparallel}\|^2 + \|v_n^{\perp}\|^2$，我们就得到

$$\|v_n^{\perp}\|^2 = a_{nn} - \alpha' A_{n-1}^{-1}\alpha.$$

你看到了什么？$a_{nn} - \alpha' A_{n-1}^{-1}\alpha$ 是 v_n 到 $\mathbb{P}$ 的距离的平方！那它当然必须大于 0，同时小于等于 v_n 的长度的平方 a_{nn}。

思考题 1.8. 书上有这样一个定理：对称矩阵 A 是正定的当且仅当 A 的顺序主子式都大于 0。看看这个是怎样打洞的？和 LU 分解定理比较一下，它们是不是很像？

思考题 1.9. 设 A 是一个元素都是整数的反对称矩阵，求证 $|A|$ 是完全平方数。

2

Jordan 标准形总结

这一章主要介绍 Jordan 形的两个运算性质和两个代数性质。两个运算性质分别是计算 Jordan 块的多项式和与 Jordan 块交换的矩阵；两个代数性质是 Jordan 块的不可分解性和分裂性质。熟悉 Jordan 块的运算很重要，等你学到常微分方程的时候就会体会到这一点。

2.1 Jordan 标准形定理的证明

我们要说的这个证明在思想上没有什么先进之处，只是把老想法用新语言说了一遍，但是这的确是最简单的说法!

我们先说一个简单的引理：

引理 2.1. 设 A 是一个线性变换，如果向量 v 满足 $A^k v \neq 0$ 但是 $A^{k+1}v = 0\,(k \in \mathbb{N}^+)$，那么 $v, Av, \ldots, A^k v$ 线性无关。

这个引理证明很简单，留给大家完成。

定理 2.2. 设 A 是 V 上的幂零线性变换，则存在 V 的一组基使得 A 在这组基下的矩阵是一些 Jordan 块的和。

证明. 对 V 的维数 n 归纳。$n=1$ 时显然，设 $\dim V < n$ 时结论成立，考虑 $\dim V = n$ 的情形。这时 A 的像空间 $A(V)$ 是 V 的 $A-$ 不变子空间且 $\dim A(V) < \dim V$，所以根据归纳假设存在 $A(V)$ 中的一组基

$$\{v_1, Av_1, \ldots, A^{a_1-1}v_1\}，\{v_2, Av_2, \ldots, A^{a_2-1}v_2\}，\ \cdots，\ \{v_m, Av_m, \ldots, A^{a_m-1}v_m\}$$

使得 A 在 $A(V)$ 上的限制在这组基下为 Jordan 标准型。其中 $A^{a_1}v_1 = A^{a_2}v_2 = \cdots = A^{a_m}v_m = 0$。显然 $A^{a_1-1}v_1, \ldots, A^{a_m-1}v_m$ 都属于 $\operatorname{Ker} A$。

下面把 $A^{a_1-1}v_1, \ldots, A^{a_m-1}v_m$ 扩充为 $\operatorname{Ker} A$ 的一组基，比如说扩充为

$$A^{a_1-1}v_1，\ \cdots，\ A^{a_m-1}v_m，\ w_1，\ \cdots，\ w_r，$$

并选取 $u_i \in V$ 使得 $Au_i = v_i$。我们断言向量组

$$\{u_1, Au_1, \ldots, A^{a_1}u_1\}，\{u_2, Au_2, \ldots, A^{a_2}u_2\}，\ \cdots，\ \{u_m, Au_m, \ldots, A^{a_m}u_m\}，\ \{w_1, \ldots, w_r\}$$

构成 V 的一组基。如果这一断言成立，那么 A 在这组基下显然就是 Jordan 标准型。

注意现在 $A^{a_1}u_1, \ldots, A^{a_m}u_m, w_1, \ldots, w_r$ 构成 $\operatorname{Ker} A$ 的一组基。

这组向量的线性无关性很好证，假设这些向量的某个线性组合 L 等于 0，两边用 A 作用以后 $A^{a_1}u_1, \ldots, A^{a_m}u_m, w_1, \ldots, w_r$ 这些项被消掉，剩下的是一个只含有 $v_1, \ldots, A^{a_1-1}v_1, \ldots, v_m, \ldots, A^{a_m-1}v_m$ 的线性组合为 0 的等式，所以它们前面的系数都是 0，即 $u_1, \ldots, A^{a_1-1}u_1, \ldots, u_m, \ldots, A^{a_m-1}u_m$ 这些项在 L 中实际上不出现，从而 L 只包含 $A^{a_1}u_1, \ldots, A^{a_m}u_m, w_1, \ldots, w_r$ 这些项。但是这些项是 $\operatorname{Ker} A$ 的一组基，所以它们前面的系数也都是 0。

有了线性无关，要证明这组向量是一组基，只要再算算维数即可。这组向量一共有 $a_1 + \cdots + a_m + r + m$ 个。另一方面，$\dim A(V) = a_1 + \cdots + a_m$，$\dim \operatorname{Ker} A = m + r$。注意由同态基本定理 $\dim V = \dim A(V) + \dim \operatorname{Ker} A$，所以这些向量的个数等于 V 的维数，从而它们构成 V 的一组基。 □

2.2 Jordan 块的运算特点

学到现在，你应该知道这个小规律（不知道的话用力将头撞墙三下 :P）当 J 是一个幂零的 Jordan 块的时候，

$$J = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{n\times n},$$

那么 J^2 就是把 J 中的 1 向右上方平移一层，J^k 就平移 $k-1$ 层，J^n 就变成零矩阵了。用这个规律我们可以很快算出一般的 Jordan 块

$$J = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}_{n\times n}$$

的多项式来：对于给定的 m 次多项式 $f(x)$，在 λ 点作 f 的 Taylor 展开：

$$f(x) = a_0 + a_1(x-\lambda) + \cdots + a_m(x-\lambda)^m,$$

那么

$$f(J) = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ & a_0 & \ddots & \vdots \\ & & \ddots & a_1 \\ & & & a_0 \end{pmatrix}.$$

可见 Jordan 块的多项式是很好算的，其形状是一个上三角的分层矩阵。

2.3 与 Jordan 块交换的矩阵

反过来我们可以证明与一个 Jordan 块 J 交换的矩阵必然可以表示为 J 的多项式。

定理 2.3. 设

$$J = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}_{n\times n},$$

且矩阵 A 满足 $AJ = JA$，则 A 必然形如

$$A = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ & a_0 & \ddots & \vdots \\ & & \ddots & a_1 \\ & & & a_0 \end{pmatrix},$$

即 A 可以表示为 J 的多项式。

证明. 设 $\varepsilon, J\varepsilon, \ldots, J^{n-1}\varepsilon$ 构成空间的一组基，则 $A\varepsilon$ 可以表示为它们的线性组合

$$A\varepsilon = a_0\varepsilon + a_1 J\varepsilon + \cdots + a_{n-1}J^{n-1}\varepsilon = f(J)\varepsilon.$$

其中 $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$。不难验证对任何的 i，

$$A(J^i\varepsilon) = J^i(A\varepsilon) = J^i(f(J)\varepsilon) = f(J)(J^i\varepsilon).$$

既然在一组基上有 $A = f(J)$ 成立，那么自然在全空间上也成立。 □

2.4 Jordan 块的不可分解性

设 J 是有限维向量空间 V 上的线性变换，且 J 在一组基下的矩阵是 Jordan 形：

$$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)J = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)\begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}.$$

我们有如下的结论：

定理 2.4. 不存在 V 的一对 $J-$ 不变的真子空间 U, W 使得 $V = U \oplus W$。

证明. 只要证明 $\lambda=0$ 的情形即可（否则就用 $J-\lambda I_n$ 代替 J）。这里的关键在于 J 仅有一个线性无关的特征向量：由于 $\mathrm{r}(J)=n-1$，所以齐次线性方程组 $JX=0$ 的解空间是一维的，这就等价于说 J 关于特征值 0 的特征子空间是一维的。如果存在满足条件的一对 U,W，那么 J 在 U 和 W 上就会各自有一个特征向量，这就导致了矛盾。 □

这个结论的意义是什么呢？说的就是 Jordan 块具有不可分解性：Jordan 块矩阵对应的线性变换不能再分解为两个更小的线性变换的直和。就像整数可以唯一分解为素数的乘积，多项式在复数域上可以唯一分解为一次因式的乘积一样，线性变换在复数域上可以唯一分解为 Jordan 块的和，这些组成单元不能够再进行分解。

如果说 Jordan 块的不可分解性是一个极端的话，那么对角矩阵的“完全可约性”则是另一个极端：

思考题 2.5. 设 A 是复数域上有限维向量空间 V 上的线性变换，则 A 可对角化的充要条件是对任何 $A-$ 不变子空间 U，都有 $A-$ 不变子空间 W 使得 $V=U\oplus W$ 成立。

然而 Jordan 块的幂却是可以分解的，而且分解的很规则，这就是下面要讨论的。

2.5 Jordan 块的分裂

我们来算 J^k 的标准形。分情况讨论，先看 $\lambda=0$ 的情形：

这个时候 J 是一个移位算子：

$$J:\quad \varepsilon_n\to\varepsilon_{n-1}\to\cdots\to\varepsilon_1\to 0,$$

整个轨道只有一条。但是 J^k 则是 k 步 k 步地跳：

$$J^k:\quad \begin{cases}\varepsilon_n\to\varepsilon_{n-2k}\to\cdots\to 0,\\ \varepsilon_{n-1}\to\varepsilon_{n-1-k}\to\cdots\to 0,\\ \cdots\\ \varepsilon_{n-k+1}\to\varepsilon_{n-2k+1}\to\cdots\to 0.\end{cases}$$

所以 J^k 有 k 条轨道，每个轨道都是一个 Jordan 块，即 J^K 的标准形中有 k 个 Jordan 块。设 $n=qk+r$，这里 $0\le r<k$，则这 k 个 Jordan 块中有 r 个是 $q+1$ 阶的，另外 $k-r$ 个是 q 阶的。

举个例子，一个 8 阶的 0 特征值 Jordan 块 J，J^3 的 Jordan 标准形是什么样子的？这个时候 J^8 有 3 个轨道 $\{\varepsilon_8,\varepsilon_5,\varepsilon_2\},\{\varepsilon_7,\varepsilon_4,\varepsilon_1\},\{\varepsilon_6,\varepsilon_3\}$，所以 J^7 的 Jordan 标准形是 2 个 3 阶的 Jordan 块和 1 个 2 阶的 Jordan 块的和。

再来看 $\lambda\neq 0$ 的情形。

$$J^k=\begin{pmatrix}\lambda^k & k\lambda^{k-1} & \ddots & \\ & \lambda^k & \ddots & \ddots\\ & & \ddots & k\lambda^{k-1}\\ & & & \lambda^k\end{pmatrix},$$

所以只需要分析 $B = J^k - \lambda^k I$ 的标准形。我们有如下的结论：

定理 2.6. 设

$$A = \begin{pmatrix} 0 & a_{12} & \cdots & \cdots & a_{nn} \\ & 0 & a_{23} & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & a_{n-1,n} \\ & & & & 0 \end{pmatrix},$$

而且 $a_{12}a_{23}\cdots a_{n-1,n} \neq 0$，则 A 的 Jordan 标准形就是一个 n 阶的 Jordan 块。

证明. 由于 B^n 是零线性变换，但 $B^{n-1}\varepsilon_n \neq 0$，所以 $\varepsilon_n, B\varepsilon_n, \ldots, B^{n-1}\varepsilon_n$ 构成空间的一组基，在这组基下 B 的矩阵就是一个 n 阶的 Jordan 块。 □

总结一下：零特征值的 Jordan 块的高次幂一定会分裂，而且是尽可能均匀的分裂；非零特征值的 Jordan 块的任意次幂都不会分裂。

一个不可约的代数结构，在某种限制或者扩张的意义下却能均匀的“碎裂”，这是代数学中一个常见而重要的现象。比如设 f 是一个有理数域 $\mathbb{Q}$ 上的不可约多项式，F 是 Q 的一个正规扩域，则如果 f 在 F 上是可约的，那么 f 必然分解成一些次数相同的多项式的乘积：

$$f = f_1 f_2 \cdots f_r，\ \deg f_1 = \cdots = \deg f_r.$$

类似的还有代数数论中素理想的分解，群表示论中不可约表示（在诱导和限制下）的分解，代数几何中不可约代数簇的分解等等。

最后我想多说一点废话。Jordan 标准形的理论告诉我们复数域上的线性变换本质上只有两种，那就是数乘变换和移位变换：

$$A\varepsilon_1 = \varepsilon_2，\ A\varepsilon_2 = \varepsilon_3，\ \cdots，\ A\varepsilon_{n-1} = \varepsilon_n，\ A\varepsilon_n = 0.$$

复数域上的任何线性变换都可以分解为数乘变换和移位变换的组合，而数乘变换是很平凡的变换，所以可以说移位运算是最重要的线性变换。Halmos 说过在泛函分析里面最重要的泛函仍然是移位运算（至少你要想想紧算子，Toeplitz 算子，Hankel 算子）。

不说泛函了，我懂的不多，再说就露馅了 :P。看来线性变换和群在集合上的作用是很相似的，就是一些不动点（相当于特征向量）和非平凡的轨道（Jordan 块）的组合，把作用限制在子群上（相当于子模 A^k）轨道还可以继续分解。有些矩阵不能对角化的原因就在于它们有非平凡的轨道，或者说，组成成分里含有移位算子。

3 秩不等式

秩不等式问题有比较模式化的解决方法，就是利用分块矩阵的初等变换，其原理就是下面这个引理：

引理 3.1. 设

$$M = \begin{pmatrix} A & 0 \\ C & B \end{pmatrix},$$

其中 A, B 都是方阵，那么 $\mathrm{r}(M) \geq \mathrm{r}(A) + \mathrm{r}(B)$。

引理的证明留给大家完成（真的很简单）。

在证明有关秩的不等式的时候，基本思路就是从准对角矩阵出发，用初等变换化成三角形的分块矩阵，来应用上面这个引理。

下面是秩不等式里面最基本的结论：

定理 3.2 (Frobenius 不等式)**.** 设 A 是数域 F 上的 $m \times n$ 矩阵，B 是 $n \times k$ 矩阵，C 是 $k \times s$ 矩阵，则

$$\mathrm{r}(AB) + \mathrm{r}(BC) \leq \mathrm{r}(ABC) + \mathrm{r}(B).$$

证明.

$$\begin{pmatrix} ABC & 0 \\ 0 & B \end{pmatrix} \to \begin{pmatrix} ABC & AB \\ 0 & B \end{pmatrix} \to \begin{pmatrix} 0 & AB \\ -BC & B \end{pmatrix}.$$

比较两端的矩阵的秩即可。 □

例 3.3 (北京大学2005)**.** 设 A 是数域 F 上 n 维向量空间上的线性变换，求证

$$A^3 = I \Leftrightarrow \mathrm{r}(I - A) + \mathrm{r}(I + A + A^2) = n.$$

证明. 注意到存在 $u(x), v(x)$ 使得 $u(x)(1-x) + v(x)(1 + x^2 + x) = 1$。考虑大的 $2n$ 阶的方阵

$$\begin{aligned} &\begin{pmatrix} I - A & 0 \\ 0 & I + A + A^2 \end{pmatrix} \xrightarrow{\text{第一列乘以 } u(A) \text{ 加到第二列}} \begin{pmatrix} I - A & u(A)(I - A) \\ 0 & I + A + A^2 \end{pmatrix} \\ &\xrightarrow{\text{第二行乘以 } v(A) \text{ 加到第一列}} \begin{pmatrix} I - A & I \\ 0 & I + A + A^2 \end{pmatrix} \\ &\xrightarrow{\text{用 } I \text{ 打洞干掉两边的矩阵}} \begin{pmatrix} 0 & I \\ A^3 - I & 0 \end{pmatrix}. \end{aligned}$$

所以这个矩阵的秩是 n 当且仅当 $A^3 - I = 0$，这就得到了证明。 □

例 3.4. 设 A, B 是数域 F 上的 n 阶方阵且满足 $AB = BA = 0, \mathrm{r}(A) = \mathrm{r}(A^2)$，求证

$$\mathrm{r}(A+B) = \mathrm{r}(A) + \mathrm{r}(B).$$

证明. 只要证明 $\mathrm{r}(A+B) \geq \mathrm{r}(A) + \mathrm{r}(B)$ 即可。用老办法，从

$$\begin{pmatrix} A+B & 0 \\ 0 & 0 \end{pmatrix}$$

出发，通过初等变换化为形如

$$\begin{pmatrix} B & * \\ 0 & A \end{pmatrix}$$

的矩阵，就得到了结论。在变换前先注意两点：

(1) 由于 A^2 的列向量都是 A 的列向量的线性组合，所以 $\mathrm{r}(A) = \mathrm{r}(A^2)$ 说明 A 的列向量组和 A^2 的列向量组是等价的，从而 A 的列向量组可以被 A^2 的列向量组线性表示。设 α_i 是 A 的第 i 个列向量，那么线性方程组 $A^2X = \alpha_i$ 有解 X_i，令 $n \times n$ 矩阵 $P = \{X_1, X_2, \ldots, X_n\}$，那么 $A^2P = A$。

(2) 当 $\mathrm{r}(A) = \mathrm{r}(A^2)$ 时有 $\mathrm{r}(A) = \mathrm{r}(A^2) = \mathrm{r}(A^3) = \cdots$

这个在 Frobenius 不等式

$$\mathrm{r}(B) + \mathrm{r}(ABC) \geq \mathrm{r}(AB) + \mathrm{r}(BC)$$

中令 $B = C = A$ 即可。

下面进行变换：

$$\begin{aligned} \begin{pmatrix} A+B & 0 \\ 0 & 0 \end{pmatrix} &\xrightarrow{\text{第一行左乘以 } A \text{ 加到第二行}} \begin{pmatrix} A+B & 0 \\ A^2 & 0 \end{pmatrix} \\ &\xrightarrow{\text{第一列右乘以 } A \text{ 加到第二列}} \begin{pmatrix} A+B & A^2 \\ A^2 & A^3 \end{pmatrix} \\ &\xrightarrow{\text{第二列右乘以 } -P \text{ 加到第一列}} \begin{pmatrix} B & A^2 \\ 0 & A^3 \end{pmatrix}. \end{aligned}$$

从而 $\mathrm{r}(A+B) \geq \mathrm{r}(B) + \mathrm{r}(A^3) = \mathrm{r}(B) + \mathrm{r}(A)$，得证。

下面再给一种解法：

可以假设 A 是 Jordan 标准形，那么 $\mathrm{r}(A) = \mathrm{r}(A^2)$ 说明 A 的 0 特征值对应的 Jordan 块都是一阶的，从而 A 形如

$$\begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix},$$

这里 A_1 是可逆的。$AB = BA = 0$ 说明 B 形如

$$\begin{pmatrix} 0 & 0 \\ 0 & B_1 \end{pmatrix},$$

这样就得到了证明。 $\square$

定理 3.5. 设 $A_1, A_2, \ldots, A_m$ 为 n 阶方阵，适合条件 $A_1 + \cdots + A_m = I$。试证明下面三个条件等价：

(a). $A_1, A_2, \ldots, A_m$ 都是幂等矩阵；

(b). $\mathrm{r}(A_1)+\mathrm{r}(A_2)+\cdots+\mathrm{r}(A_m) = n$；

(c). 当 $1 \le i < j \le n$ 时有 $A_iA_j = A_jA_i = 0$。

证明. $(a) \Rightarrow (b)$：用熟知的结论，幂等矩阵的秩等于它的迹，可得

$$\sum_{i=1}^{m} \mathrm{r}(A_i) = \sum_{i=1}^{m} \mathrm{tr}(A_i) = \mathrm{tr}(I_n) = n.$$

$(b) \Rightarrow (c)$：考虑大的准对角方阵

$$\begin{pmatrix} 0 & & & \\ & A_1 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix} \xrightarrow{\text{都加到第一行}} \begin{pmatrix} 0 & A_1 & \cdots & A_m \\ & A_1 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix} \xrightarrow{\text{都加到第一列}} \begin{pmatrix} I_n & A_1 & \cdots & A_m \\ A_1 & A_1 & & \\ \vdots & & \ddots & \\ A_m & & & A_m \end{pmatrix}$$

$$\xrightarrow{\text{打洞干掉第一行和第一列}} \begin{pmatrix} I_n & 0 & \cdots & 0 \\ 0 & A_1 - A_1^2 & \cdots & -A_1A_m \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -A_mA_1 & \cdots & A_m - A_m^2 \end{pmatrix}.$$

所以 $\mathrm{r}(A_1) + \mathrm{r}(A_2) + \cdots + \mathrm{r}(A_m) = n$ 当且仅当右下角的大的矩阵为 0，即 A_i 都是幂等矩阵而且互相正交。

$(c) \Rightarrow (a)$：很简单，省略。 □

定理 3.5 是数理统计中著名的 Cochran 定理的矩阵版本：

定理 3.6 (Cochran **定理**)**.** 设 $X_1, X_2, \ldots, X_n$ 是相互独立且服从正态分布 $N(0, 1)$ 的随机变量，若

$$Q_1 + Q_2 + \cdots + Q_k = \sum_{i=1}^{n} X_i^2,$$

其中 Q_i 是秩为 n_i 的关于 $X_1, X_2, \ldots, X_n$ 的二次型，则下面三个条件等价：

1. $Q_1, Q_2, \ldots, Q_k$ 相互独立；
2. $Q_i \sim \chi^2(n_i)$；
3. $\sum\limits_{i=1}^{k} n_i = n$。

例 3.7 (北京大学2007). A, B 是两个 n 阶矩阵满足 $AB = BA$。求证

$$\mathrm{r}(A) + \mathrm{r}(B) \geq \mathrm{r}(A + B) + \mathrm{r}(AB).$$

证明. 设 X 是齐次线性方程组 $AX = 0$ 的解空间，Y 是齐次线性方程组 $BX = 0$ 的解空间，Z 是齐次线性方程组 $ABX = BAX = 0$ 的解空间，W 是齐次线性方程组 $(A + B)X = 0$ 的解空间，那么我们有 $X \subset Z, Y \subset Z$，从而 $X + Y \subset Z$，而且 $X \cap Y \subset W$。由维数公式，

$$\dim X + \dim Y = \dim X \cap Y + \dim(X + Y) \leq \dim W + \dim Z.$$

从而

$$n - \mathrm{r}(A) + n - \mathrm{r}(B) \leq n - \mathrm{r}(A + B) + n - \mathrm{r}(AB),$$

即 $\mathrm{r}(A) + \mathrm{r}(B) \geq \mathrm{r}(A + B) + \mathrm{r}(AB)$。 □

记得以前在论坛上有个网友发了一个帖子，说他写了一篇论文，证明了这样一个结论：设 A, B 都是 n 阶方阵且 $AB = BA$，那么有如下的不等式成立：

$$\mathrm{r}(A^2) + \mathrm{r}(B^2) \geq 2\mathrm{r}(AB).$$

他说这是均值不等式在矩阵里的推广。请大家思考一下这个问题，看看结论成立不?

4

交结数：刻画相似程度的不变量

以下都固定 A, B 分别是数域 F 上的 n 阶和 m 阶方阵，K 是 F 的扩域。这一章要讨论的是矩阵方程 $AX = XB$ 的解空间

$$U_F = \left\{ X \in \mathrm{Mat}_{n\times m}(F) \;\middle|\; AX = XB \right\}$$

的维数，这个维数又叫做 A 与 B 的交结数，用 $i(A, B)$ 来表示。下面就会看到，交结数是一个刻画 A 和 B 相似程度的不变量，它包含了很丰富的信息。

首先为什么说交结数是个不变量呢？这里的不变有两个含义，第一是交结数是一个不依赖于域的量。假设把 A, B 看成更大的域 K 上的矩阵，考虑

$$U_K = \left\{ Y \in \mathrm{Mat}_{n\times m}(K) \;\middle|\; AY = YB \right\},$$

那么我们有

$$\dim U_F = \dim U_K = i(A, B).$$

这个道理很简单，解矩阵方程就是解齐次线性方程组，大家求基础解系的过程是一样的，在 F 内就能完成。只不过 U_F 是基础解系的 $F-$ 线性组合，U_K 是基础解系的 $K-$ 线性组合。

交结数是不变量的第二个含义是在相似变换下保持不变。即若矩阵 C 相似于 A，D 相似于 B，则 $i(A, B) = i(C, D)$。这一点的验证留给大家完成。

由上面的讨论可以得到

定理 4.1. [矩阵的相似性和域无关] 如果 A, B 在 K 上相似，则它们在 F 上也相似。

证明. A, B 在 K 上相似说明矩阵方程 $AX = XB$ 在 K 上有一个可逆解，现在要证明它在 F 上也有一个可逆解。设 $X_1, X_2, \ldots, X_s$ 是 U_F 的一组基，那么它们也构成 U_K 的一组基。$AX = XB$ 在 K 上有一个可逆解说明存在一组 $a_i \in K$ 使得 $a_1X_1 + \cdots + a_sX_s$ 可逆。我们要证明一定有一组 $b_j \in F$ 使得 $b_1X_1 + \cdots + b_sX_s$ 可逆。自然我们要考察 s 个变元 $t_1, t_2, \ldots, t_s$ 的多元多项式

$$f(t_1, \cdots, t_s) = \det |t_1X_1 + t_2X_2 + \cdots + t_sX_s|\,.$$

f 是一个数域 F 上的多元多项式（当然也是 K 上的），$f(0, \cdots, 0) = 0$ 和 $f(a_1, \cdots, a_s) \neq 0$ 说明 f 不是常数多项式。由于数域有无穷多个元素，所以必定存在 F 中的 $b_1, b_2, \ldots, b_s$ 使得 $f(b_1, \cdots, b_s) \neq 0$（想一想为什么？）。这就说明了 A, B 在 F 上也是相似的。 □

你也许要说，这样做真麻烦，用不变因子直接说明不就行了吗？还用不着数域的假设。可是你再看看下面这个问题：

定理 4.2. 设 $A_1, A_2, \ldots, A_m, B_1, B_2, \ldots, B_m$ 是 F 上的 n 阶方阵，如果存在 K 上的可逆矩阵 S 使得 $S^{-1}A_iS = B_i\,(i = 1, 2, \ldots, m)$，那么必存在 F 上的可逆矩阵 T 使得 $T^{-1}A_iT = B_i\,(i = 1, 2, \ldots, m)$。

这个定理来自群表示论，意义就是两个表示在 K 上等价就在 F 上等价。显然这个时候不变因子的办法就不好使了，因为没法说明 T 是“公共”的。但是只要 F 有无限多个元素，定理 4.1 的办法就能用，所以它还是很有应用前景的。

证明. 沿用前面的记号，考虑向量空间

$$U_F = \left\{ T \in \mathrm{Mat}_n(F) \;\middle|\; A_iT = TB_i，\ i = 1, 2, \ldots, m \right\},$$

$$U_K = \left\{ S \in \mathrm{Mat}_n(K) \;\middle|\; A_iS = SB_i，\ i = 1, 2, \ldots, m \right\}.$$

其实就是把 m 个线性方程组联立起来。所以和前面同样的道理，U_F 和 U_K 的维数是一样的。设 $T_1, T_2, \ldots, T_s$ 是 U_F 和 U_K 共同的一组基，考察多元多项式

$$f(t_1, \cdots, t_s) = \det|\, t_1T_1 + \cdots + t_sT_s \,|\,.$$

S 的存在说明 f 不是零多项式，基域 F 有无限多个元素就保证了存在 $(a_1, a_2, \ldots, a_s) \neq (0, \ldots, 0)$ 使得 $f(a_1, \cdots, a_s) \neq 0$，所以 $T = a_1T_1 + \cdots + a_sT_s$ 满足要求。 □

其实不管 F 是什么域这个定理都是成立的，我们已经证明的是 F 是无限域的情形。至于 F 是有限域的情形，要用到比较深的代数知识，文档里就不写了。

接下来是个很有用的结论：

定理 4.3. 若 A, B 没有共同的特征值，则矩阵方程 $AX = XB$ 只有零解。

证明. 设矩阵 X 满足 $AX = XB$，则不难验证对任何正整数 k 都有 $A^kX = XB^k$，从而对任何多项式 $f(x)$ 都有 $f(A)X = Xf(B)$ 成立。特别令 $f(x)$ 是 A 的特征多项式，由 Hamilton-Cayley 定理有 $f(A) = 0$，所以 $Xf(B) = 0$。但是 $f(B)$ 是可逆矩阵（想一想，为什么？），所以 $X = 0$，定理得证。 □

这个结论有个很好的应用，就是当一个矩阵 A 形如

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

（这里 A_1 和 A_2 没有共同的特征值）的时候，与 A 交换的矩阵 B 也必然形如

$$\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}.$$

反复地应用这个结论可以推广到对角线上有 n 个矩阵的情形。换言之，当把 A 按特征值分块的时候与 A 交换的矩阵 B 也就同时被分块了，这就起到了简化问题的作用。例题 3.4 第二个解法就是用的这一招，而且这一招还会反复出现，所以一定要注意。

最后来给出交结数的确切值：

定理 4.4. 设 $f_1(x), \ldots, f_p(x)$ 为 A 的全部不变因子，$g_1(x), \ldots, g_q(x)$ 为 B 的全部不变因子，则

$$i(A, B) = \sum_{i=1}^{p}\sum_{j=1}^{q} \deg\gcd\Big(f_i(x), g_j(x)\Big).$$

证明大意：由于交结数是相似变换下的不变量，所以可以在 A, B 都是 Jordan 标准型的前提下进行计算。设

$$A = \begin{pmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_r}(\lambda_r) \end{pmatrix}, \quad B = \begin{pmatrix} L_{m_1}(\mu_1) & & & \\ & L_{m_2}(\mu_2) & & \\ & & \ddots & \\ & & & L_{m_s}(\mu_s) \end{pmatrix}.$$

相应地把 X 分为 rs 个子块 X_{ij}，X_{ij} 的阶是 $k_i \times m_j$，那么 $AX = XB$ 就变成了

$$J_iX_{ij} = X_{ij}L_j.\ \ i = 1, 2, \ldots, r，\ j = 1, 2, \ldots, s.$$

整个矩阵方程 $AX = XB$ 的解空间的维数就是各个方程 $J_iX_{ij} = X_{ij}L_j$ 的解空间维数的和。不难算出矩阵方程 $J_iX_{ij} = X_{ij}L_j$ 的解空间的维数是

$$\gcd\Big((x-\lambda_i)^{k_i}, (x-\mu_j)^{m_j}\Big)$$

的次数，所以矩阵方程 $AX = XB$ 的解空间的维数就是两两求出 A 和 B 的初等因子的最大公因式，然后把次数加起来。进一步这个结果还等于两两求出不变因子的最大公因式，然后把次数相加，这就证明了定理。

可以看到，A 和 B 的“相似”程度越高，交结数的值就越大，可以说交结数刻画了 A 和 B 之间的相似程度。用模的语言可以更好的解释清楚：交结数就是从 $B-$ 模 W 到 $A-$ 模 V 的模同态组成的向量空间的维数。至于定理 4.3，可以解释为当矩阵方程有非零解的时候，由模同态基本定理，有模同构

$$W/\operatorname{Ker} X \cong \operatorname{Im} X.$$

即 B 在某个商空间上的诱导变换和 A 在某个子空间上的限制是相似的，从而 A, B 有共同的特征值。

思考题 4.5. 已知方阵 A, B 的特征值都是正数且满足 $A^2 = B^2$，求证 $A = B$。

5

同时上三角化

在叙述这一章内容之前，我们要假定大家已经熟悉下面的结论：

- 设 $\mathfrak{F}$ 是复数域上向量空间 V 上的一族线性变换且 $\mathfrak{F}$ 中的元素两两可交换，则存在 V 中的一组基使得 $\mathfrak{F}$ 中所有元素在这组基下的矩阵都是上三角形。
- 仍设 $\mathfrak{F}$ 是复数域上向量空间 V 上的一族两两可交换的线性变换。如果 $\mathfrak{F}$ 中的元素都是可对角化的，则存在 V 中的一组基使得 $\mathfrak{F}$ 中所有变换在这组基下的矩阵都是对角矩阵。

由于这是两个很常见的命题，它们的证明在大多数参考书上都可以找到，所以我在这里就偷偷懒，不再写出证明了。但是你一定要把它们当做很重要的结论来记，既要熟悉证明，又要牢记结论，因为线性变换的同时上三角化（或者对角化）是处理多个线性变换最重要的办法，数学中随处可见它们的应用。需要提醒大家的是，同时上三角化的关键是找公共的特征向量。

举两个简单的例子：

例 5.1. 设 A 是酉空间上的正规变换，A^* 为其共轭变换，求证 A^* 可以表示为 A 的多项式。

我第一次看到这个问题也不会做。后来会做了，还是觉得不简单。现在看看，自己当时真的很傻很天真。

证明. 正规变换的条件就是 $AA^* = A^*A$。这是两个交换的可对角化的线性变换，因此可以同时对角化，因此，拉格朗日插值即可。 □

例 5.2. 设 A 是数域 F 上的 n 阶矩阵，考虑 $\mathrm{Mat}_n(F)$ 上的线性变换 $T(X) = AX - XA$。求证：如果 A 是幂零的，则 T 也是幂零的；如果 A 是可对角化的，则 T 也是可对角化的。

证明. 考虑这样两个线性变换 $l_A(X) = AX$ 和 $r_A(X) = XA$，则 $T = l_A - r_A$ 而且 l_A 和 r_A 是可交换的。

1. 当 A 是幂零线性变换时，$(l_A)^n(A) = A^nX = 0$，所以 l_A 是 $\mathrm{Mat}_n(F)$ 上的幂零线性变换，同理 r_A 也是，它俩又是可以交换的，从而它们的差也是幂零的。

2. 如果 A 是可对角化的，则 l_A 也是可对角化的。这是因为假设 A 的 n 个线性无关的特征向量是 $\alpha_1,\dots,\alpha_n$，那么每一个特征向量 α_i 都对应于 l_A 的 n 个特征向量

$$(\alpha_i,0,\dots,0),\ (0,\alpha_i,\dots,0),\ (0,0,\dots,\alpha_i).$$

同理 r_A 也是可对角化的，再结合 l_A 与 r_A 可交换就得到 l_A 和 r_A 可以同时对角化，从而它们的差 $T=l_A-r_A$ 可对角化。

□

下面是两个在没有可交换的条件下同时上三角化的情形：

定理 5.3. 设 A,B 是两个 n 阶复矩阵，且存在复数 a,b 使得 $AB-BA=aA+bB$，求证 A,B 可以同时上三角化。

证明. 如果 a,b 都是 0，那么交换的矩阵可以同时上三角化，结论成立。所以不妨设 $a\neq 0$。以 B/a 代替 B 我们还可以假设 $a=1$。记 $C=AB-BA=A+bB$，则 $CB-BC=C$。我们来证明 B 和 C 可以同时上三角化，那么 A 也就被上三角化了。首先证明 B 和 C 必有公共的特征向量。假设 B 和 C 没有公共的特征向量，那么设 $Bx=\lambda x$ 是 B 的特征向量，则

$$CBx-BCx=Cx,\ (B-\lambda+1)Cx=0.$$

这里 $Cx\neq 0$，否则 x 就是 B 和 C 的公共特征向量了，所以 B 有特征值 $\lambda-1$。同理 B 也应该有特征值 $\lambda-2$。如此论证下去，B 会有无穷多个特征值，这不可能，所以 B 和 C 必有公共的特征向量，记作 v。

把 v 开拓为整个空间的一组基，B,C 的矩阵形如

$$B=\begin{pmatrix}\lambda & *\\ 0 & B_1\end{pmatrix},\ C=\begin{pmatrix}\mu & *\\ 0 & C_1\end{pmatrix},$$

那么 $C_1B_1-B_1C_1=C_1$，再用归纳法即可。 □

定理 5.4. 当 $\mathrm{r}(AB-BA)\leq 1$时 A,B 可以同时上三角化。

证明. 首先我们可以假设 A 不是可逆的，否则的话就以 $A-\lambda I$ 代替 A。这个假定的目的就是保证 $\operatorname{Ker} A$ 和 $\operatorname{Im} A$ 都是非平凡的 $A-$ 不变子空间。我们要论证 $\operatorname{Ker} A$ 和 $\operatorname{Im} A$ 中至少有一个是 B 的不变子空间，这样就可以降低空间的维数，便于应用归纳法。记齐次线性方程组 $(AB-BA)X=0$ 的解空间为 X，分两种情况讨论：

1. $\operatorname{Ker} A\subset X$。设$x\in\operatorname{Ker} A$，立刻有 $ABx=BAx=0$，从而 $\operatorname{Ker} A$ 也是 B 的不变子空间。

2. $\operatorname{Ker} A \not\subseteq X$。则存在向量 $x \in \operatorname{Ker} A$ 但是 $y = (AB-BA)x \neq 0$。注意到 $\mathrm{r}(AB-BA) = 1$，所以 $\operatorname{Im}(AB-BA) = \{y\}$。但是
$$y = (AB-BA)x = ABx - BAx \in \operatorname{Im} A,$$
这就说明 $\operatorname{Im}(AB-BA) \subset \operatorname{Im} A$，那么对任何 $u = Av \in \operatorname{Im} A$ 我们有
$$BAv = ABv - (AB-BA)v \in \operatorname{Im} A,$$
从而 $\operatorname{Im} A$ 确实是 B 的不变子空间。

剩下的就可以对空间的维数进行归纳来证明结论了。不妨设 $\operatorname{Im} A$ 是 A, B 共同的不变子空间，那么取 $\operatorname{Im} A$ 的一组基然后开拓，A 和 B 的矩阵就形如
$$A = \begin{pmatrix} A_1 & A_3 \\ 0 & A_2 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & B_3 \\ 0 & B_2 \end{pmatrix}.$$
显然
$$\mathrm{r}(AB-BA) \geq \mathrm{r}(A_1B_1 - B_1A_1) + \mathrm{r}(A_2B_2 - B_2A_2),$$
所以由归纳假设 A_1 和 B_1 可以同时上三角化，A_2 和 B_2 可以同时上三角化，这就完成了证明。 □

最后来看一个应用：

例 5.5 (交换矩阵空间最大维数). 设 M 是 $\mathrm{Mat}_n(\mathbb{C})$ 的一个子空间，如果 M 中的矩阵两两可以交换，求证 M 的维数最大是 $[\frac{n^2}{4}]+1$。

证明. 对 n 归纳，$n=1$ 结论显然成立，设小于 n 的时候结论成立，来看 n 的情形：

由于 M 中的矩阵两两可以交换，所以它们可以同时上三角化，所以无妨假设 M 中的每一个矩阵都是上三角矩阵。对每个 $A \in M$，我们截取 A 左上角的 $n-1$ 阶子矩阵，把这个矩阵记作 $f(A)$；同时截取 A 的右下角的 $n-1$ 阶子矩阵，把它记作 $g(A)$。那么所有的 $f(A)$ 之间两两可以交换，所有的 $g(A)$ 两两之间可以交换。由于 f 和 g 都可以看做是 M 到 $\mathrm{Mat}_{n-1}(\mathbb{C})$ 的交换子空间的线性映射，所以由归纳假设
$$\dim f(M) \leq [\frac{(n-1)^2}{4}], \quad \dim g(M) \leq [\frac{(n-1)^2}{4}].$$

不难看出 $\operatorname{Ker} f$ 中的元素形如 $\begin{pmatrix} 0_{n\times n-1} & \alpha \end{pmatrix}$，$\operatorname{Ker} g$ 中的元素形如 $\begin{pmatrix} \beta' \\ 0_{n-1\times n} \end{pmatrix}$。这里的 α 和 β 都是 $n\times 1$ 向量。两者交换意味着 $\beta'\alpha = 0$，即 $\alpha \perp \beta$，所以 $\dim \operatorname{Ker} f + \dim \operatorname{Ker} g \leq n$，从而
$$\begin{aligned}
\dim M = \dim \operatorname{Ker} f + \dim f &= \dim \operatorname{Ker} g + \dim g \\
&\leq \frac{\dim \operatorname{Ker} f + \dim \operatorname{Ker} g}{2} + [\frac{(n-1)^2}{4}] + 1 \\
&\leq \frac{n}{2} + [\frac{(n-1)^2}{4}] + 1 \\
&\leq [\frac{n^2}{4}] + 1.
\end{aligned}$$

最后给出构造的例子：

$$\begin{pmatrix} \lambda I_m & N \\ 0 & \lambda I_m \end{pmatrix} \ (n=2m), \quad \begin{pmatrix} \lambda I_m & N \\ 0 & \lambda I_{m+1} \end{pmatrix} \ (n=2m+1).$$

□

当然你也可以问一些类似的问题，比如设 M 是 $\mathrm{Mat}_n(\mathbb{C})$ 的一个子空间，如果 M 中的矩阵都不可逆，那么 M 的维数最大是多少？答案不难猜到，是 $n(n-1)$，但是解决起来很繁琐。

最著名的是这个问题：设 M 是 $\mathrm{Mat}_n(\mathbb{R})$ 的一个子空间，如果 M 中的矩阵都是可逆矩阵，问 M 的维数最大是多少？注意这里的基域是 $\mathbb{R}$，因为在复数域上这个答案只能是 1。

这个问题的答案是猜不到的，叫做 Radon-Hurwitz 数。设 $\dim M = n$，把 n 写成 $n = 2^a \cdot b$ 的形式，这里 b 是奇数。再设 $a = 4c + d$ $(0 \le d \le 3)$，则 Radon-Hurwitz 数定义为

$$\rho(n) = 2^d + 8c - 1.$$

这就是 M 维数的最大值。

这个问题的解决要用到微分拓扑和 Clifford 代数的表示，这属于研究生阶段的内容，所以文档中就不再介绍了。在它的背后，是一副波澜壮阔的数学画卷。

6

覆盖定理

定理 6.1. 数域 F 上的有限维向量空间 V 不能被它的有限个真子空间覆盖，即不存在真子空间 $V_1, V_2, \ldots, V_m$ 使得 $V \subseteq V_1 \cup \cdots \cup V_m$。

证明. 设 $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ 是 V 的一组基，考察

$$\alpha_i = \varepsilon_1 + i\varepsilon_2 + \cdots + i^{n-1}\varepsilon_n,$$

由 Vandermonde 行列式的知识知道无穷向量序列 $\{\alpha_n\}_{n=1}^{\infty}$ 中任何 n 个都构成 V 的一组基。而 $V_1, \ldots, V_k$ 只能包含其中有限多个向量，从而有无穷多个 α_i 不能被覆盖。 $\square$

这个定理用到的地方很少，不过每次应用都可以带来惊喜。以下是两个例子：

例 6.2 (2009南开大学). 设 F 是一个数域，T 是 $\mathrm{Mat}_n(F)$ 上的线性变换满足对任何 $A, B \in \mathrm{Mat}_n(F)$，$T(AB) = T(A)T(B)$ 和 $T(AB) = T(B)T(A)$ 至少有一个成立。求证：要么对所有的 $A, B \in \mathrm{Mat}_n(F)$ 都有 $T(AB) = T(A)T(B)$ 成立，要么对所有的 $A, B \in \mathrm{Mat}_n(F)$ 都有 $T(AB) = T(B)T(A)$ 成立。

证明. 对于一个固定的矩阵 A，如果对任何矩阵 B 都有 $T(AB) = T(A)T(B)$ 成立，就称 A 是正手型的。相应地，如果对任何矩阵 B 都有 $T(AB) = T(B)T(A)$ 成立，就称 A 是反手型的。证明分两步，首先证明 $\mathrm{Mat}_n(F)$ 中的方阵必然是正手型或者反手型之一，其次证明正手型和反手型不能同时出现，就完成了证明。

首先对任一矩阵 A，设 U_A 是所有满足 $T(AB) = T(A)T(B)$ 的方阵 B 构成的集合，W_A 是所有满足 $T(AC) = T(C)T(A)$ 的方阵 C 构成的集合，不难验证 U_A, W_A 都是 $\mathrm{Mat}_n(F)$ 的子空间。而且根据已知，$\mathrm{Mat}_n(F) = U_A \cup W_A$，从而 U_A, W_A 不可能都是 $\mathrm{Mat}_n(F)$ 的真子空间，所以必有 $U_A = \mathrm{Mat}_n(F)$ 或者 $W_A = \mathrm{Mat}_n(F)$ 之一成立。从而 A 不是正手型就是反手型。

其次设 U_1 是全体正手型的 A 构成的集合，U_2 是全体反手型的 A 构成的集合，显然 U_1, U_2 也都是子空间。注意由刚刚证得的结论，$\mathrm{Mat}_n(F) = U_1 \cup U_2$，从而 U_1, U_2 不能都是真子空间，所以要么 $U_1 = \mathrm{Mat}_n(F)$，要么 $U_2 = \mathrm{Mat}_n(F)$。如果前者成立（都是正手型的），那么对任何 A, B 都有

$$T(AB) = T(A)T(B).$$

同理如果后者成立（都是反手型的），则

$$T(AB) = T(B)T(A).$$

这样就得到了证明。 □

定理 6.3. 设 A 是 n 维向量空间 V 上的线性变换。对 $v \in V$，记 $m_v(x)$ 为使得 $m_v(A)v = 0$ 成立的次数最低的首一多项式，这个多项式也叫做 v 的极小多项式。求证：存在 $v \in V$ 使得 v 的极小多项式 $m_v(x)$ 恰好等于 A 的极小多项式 $m(x)$。

证明. 显然对任何 $v \in V$，$m_v(x)$ 都整除 A 的极小多项式 $m(x)$。但是 $m(x)$ 只有有限多个因子，所以当 v 跑遍 V 时，$m_v(x)$ 只有有限多个不同的可能 $m_1(x), \ldots, m_k(x)$。考虑

$$V_i = \{\, v \in V \,\big|\, m_i(A)v = 0 \,\},$$

那么 V 等于 $V_1, V_2, \ldots, V_k$ 的并，由于 $V_1, V_2, \ldots, V_k$ 不能都是真子空间，所以必有某个 V_i 使得 $V = V_i$，从而 $m(x) = m_i(x)$，得证。 □

定理 6.4. A 的特征多项式 $f(x)$ 与极小多项式 $m(x)$ 相等的充要条件是存在 $v \in V$ 使得 $v, Av, \ldots, A^{n-1}v$ 构成 V 的一组基。

证明. 必要性：由已知，$m(x)$ 的次数是 n。根据定理 6.3，有一个向量 v 使得 $m_v(x) = m(x)$，从而 $v, Av, \ldots, A^{n-1}v$ 必然线性无关。否则的话设

$$a_0 v + a_1 Av + \cdots + a_{n-1}A^{n-1}v = 0,$$

这就说明 $m_v(x)$ 的次数小于等于 $n-1$，矛盾！
充分性：显然 $m_v(x)$ 是 n 次的，而 $m_v(x)|m(x)|f(x)$，所以 $m_v(x) = m(x) = f(x)$。 □

以下是一些记号上的约定。

设 A 是向量空间 V 上的线性变换。如果存在向量 $v \in V$ 使得 $v, Av, \ldots, A^{n-1}v$ 构成 V 的一组基，我们就称 V 在 A 的作用下是一个循环空间，v 是一个生成元，并用记号 $V = \{v\}$ 来表示整个空间 V 可以被 v 这一个向量生成。特别注意：我们证明了在循环空间 $V = \{v\}$ 中有 $m_v(x) = m(x) = f(x)$ 成立。这个时候 V 中的任何向量 u 都可以表示为 $u = g(A)v$ 的形式，这里 $g(x)$ 是一个多项式（我们干脆就简称为 u 可以表示为 v 的多项式）。这种表示方式不是唯一的，但是如果

$$u = g_1(A)v = g_2(A)v,$$

则 $[g_1(A) - g_2(A)]v = 0$，从而 $g_1 - g_2$ 可以被 A 的极小多项式整除。（验证一下！）

我们已经证明了 V 是 A 作用下的循环空间等价于 A 的极小多项式等于其特征多项式。我们下一章就会证明这个事情还等价于任何与 A 交换的矩阵都是 A 的多项式。

7

有理标准形和交换的矩阵

7.1 Frobenius 有理标准形定理

定理 7.1. 设 V 是域 F 上一个 n 维向量空间，A 是 V 上的线性变换，则存在一组向量 $v_1, \ldots, v_r$ 使得下面两个条件同时成立：

- $V = \{v_1\} \oplus \{v_2\} \oplus \cdots \oplus \{v_r\}$。
- 设 A 在子空间 $\{v_i\}$ 上的特征多项式为 $p_i(x)$，则 $p_r|\cdots|p_2|p_1$。

这组 v_i 一般不是唯一的，但是这 r 个多项式 $p_1(x), \ldots, p_r(x)$ 是被 A 唯一决定的，它们叫做 A 的不变因子。两个线性变换 A 和 B 是相似的，当且仅当它们有相同的不变因子。

思考题 7.2. 在定理 7.1 的结论下，证明 A 的特征多项式 $f(x) = p_1p_2\cdots p_r$，A 的极小多项式 $m(x) = p_1(x)$。

Frobenius 有理标准形定理是线性代数最重要的定理之一，也是教材中最难学的部分。教材上 $\lambda-$ 矩阵的途径是最不直观的，但却是唯一能够在证明分解的存在性和不变因子唯一性的同时又给出计算不变因子具体方法的，是必须掌握的。所以如果你阅读这一部分的内容有困难的话，我也没有什么好的办法，多读多揣摩，时间长了就会好些。如果你有机会学习抽象代数学中“主理想整环上的有限生产模”的知识，就会真正理解不变因子的概念。

不过呢，如果只想证有理标准形定理的存在性，而不考虑不变因子的计算的话，还是有其它的办法的。

Frobenius 标准形定理的证明

这个证明使用的是空间分解和提升的思想。这个思想可以表示为一个三段论：先找合适的子空间降阶，然后转移到商空间使用归纳假设，最后拉回到全空间得出结论。我们来看看它是怎么做的。

对空间的维数 n 归纳。根据定理 6.3，存在一个向量 $v_1 \in V$ 使得 v_1 的极小多项式 $p_1(x)$ 等于 A 的极小多项式 $m(x)$。这个 v_1 我们已经找好了，接下来是要找满足要求的 $v_2, \ldots, v_r$。

考虑商空间 $V/\{v_1\}$，这个空间维数严格小于 V，因此根据归纳假设，存在向量 $\overline{v_2}, \ldots, \overline{v_r} \in V/\{v_1\}$ 满足

1. $V/\{v_1\} = \{\overline{v_2}\} \oplus \cdots \oplus \{\overline{v_r}\}$。

2. 设 $\overline{v_2}, \ldots, \overline{v_r}$（在诱导变换 $\overline{A}$ 下）的极小多项式为 $p_2, \ldots, p_r$，则 $p_r|\cdots|p_2$。

$p_r|\cdots|p_2|p_1$ 是显然的（你自己想一下），看起来我们只要给每个 $\overline{v_i}\,(i \geq 2)$ 都指定一个代表元 v_i，然后证明 $V = \{v_1\} \oplus \{v_2\} \oplus \cdots \oplus \{v_r\}$ 就好了。但是 $\overline{v_i}\,(i \geq 2)$ 的代表元都是无穷多个，选取哪一个才能让直和这个要求满足呢？

引理 7.3. 对每个 $i \geq 2$，存在 $\overline{v_i}$ 在 V 中的代表元 v_i 使得 v_i 在 V 中的极小多项式等于 $p_i(x)$。

让我们先承认这个引理的正确性，把证明继续下去：

下面来证明

$$V = \{v_1\} \oplus \{v_2\} \oplus \cdots \oplus \{v_r\}.$$

首先证明 $V = \{v_1\} + \cdots + \{v_r\}$。由于

$$V/\{v_1\} = \{\overline{v_2}\} \oplus \cdots \oplus \{\overline{v_r}\},$$

以及

$$\{\overline{v_i}\} = \overline{\{v_i\}} = \{\, x_i + \{v_1\} \,|\, x_i \in \{v_i\} \,\},$$

所以对任何 $v \in V$，

$$v + \{v_1\} = c_2(x_2 + \{v_1\}) + \cdots + c_r(x_r + \{v_r\}).\ \ x_i \in \{v_i\}.$$

因此 $v - c_2x_2 - \cdots - c_rx_r \in \{v_1\}$，这就证明了 $V = \{v_1\} + \cdots + \{v_r\}$。

接下来还要证明是直和。设 $u_i \in \{v_i\}$ 满足

$$a_1u_1 + a_2u_2 + \cdots + a_ru_r = 0.$$

我们之前说过，循环空间 $\{v_i\}$ 中的向量 u_i 可以表示为 v_i 的多项式：$u_i = f_i(A)v_i$。转移到商空间 $V/\{v_1\}$ 中去：

$$a_2f_2(\overline{A})\overline{v_2} + \cdots + a_rf_r(\overline{A})\overline{v_r} = \overline{0}.$$

然而由归纳假设，$\{\overline{v_2}\} \oplus \cdots \oplus \{\overline{v_r}\}$ 是直和，所以必须 $a_if_i(\overline{A})\overline{v_i} = \overline{0}\,(i \geq 2)$。因此 $a_if_i(x)$ 可以被 $\overline{v_i}$ 的极小多项式 $p_i(x)$ 整除。然而 v_i 的极小多项式也是 p_i，从而 $a_iu_i = a_if_i(A)v_i = 0$，从而每一个和项都是零向量，这就证明了直和。

因此所有的焦点集中到证明引理 7.3 上来：

引理 7.3 的证明：$\overline{v_i}$ 的任何两个代表元的差都是 $\{v_1\}$ 中的一个向量，而 $\{v_1\}$ 中的向量都可以写成 v_1 的多项式。因此我们就先随便给 $\overline{v_i}$ 指定一个代表元 v_i，然后找合适的多项式 $g_i(x)$ 调整它，使得新的代表元 $v_i + g_i(A)v_1$ 满足我们的要求。

我们以 $i = 2$ 为例：$p_2(\overline{A})(\overline{v_2}) = \overline{0}$ 说明 $p_2(A)v_2 \in \{v_1\}$。不妨设

$$p_2(A)v_2 = r_2(A)v_1,\ \ r_2(x) \in F[x].$$

由于 $p_2(x)|p_1(x)$，设 $p_1 = p_2m_2$，上式两边同时用 $m_2(A)$ 作用得到

$$0 = r_2(A)m_2(A)v_1.$$

因此 $p_1|r_2m_2$，即 $p_2|r_2$。再令 $r_2(x) = p_2(x)l_2(x)$，又有

$$p_2(A)(v_2 - l_2(A)v_1) = 0.$$

我们说 $v_2 - l_2(A)v_1$ 就是我们想要的代表元：它被 $p_2(A)$ 零化，因此设它的极小多项式为 $h(x)$，则 $h(x)|p_2(x)$。另一方面

$$h(A)(v_2 - l_2(A)v_1) = 0.$$

转移到商空间 $V/\{v_1\}$ 中：

$$h(\overline{A})\overline{v_2} = \overline{0}.$$

因此 $p_2|h_2$，于是 $h_2(x) = p_2(x)$。

至此我们就完成了 Frobenius 标准形定理的证明。

7.2 何时与 A 交换的矩阵总是 A 的多项式？

何时与矩阵 A 交换的矩阵总能表示为 A 的多项式？下面将从变换和矩阵两个角度入手，对这个问题进行一下深入的探讨，力求把想法写清楚。

变换的角度

如果任何与 A 交换的变换 B 总能表示为 A 的多项式，那么 A 应该有怎样的性质？我们来观察 V 在 A 的作用下的循环子空间分解

$$V = \{v_1\} \oplus \{v_2\} \oplus \cdots \oplus \{v_r\}.$$

这里的 $v_1, \ldots, v_r$ 满足的条件与定理 7.1 中一致。

设 B 是任一与 A 交换的变换。由于 B 可以表示为 A 的多项式，所以 B 必然也以 $\{v_1\}, \ldots, \{v_r\}$ 为其不变子空间。直观上说，就是 B 不能把子空间 $\{v_i\}$ 中的向量映到别的 $\{v_j\}$ 中去。

然而我们可以证明这样一个结论：

引理 7.4. 假设 $r > 1$，则存在与 A 交换的变换 B 使得 $Bv_1 = v_2$。

假设这个引理是正确的，那么 B 肯定不是 A 的多项式，所以必须 $r = 1$，即 A 是由一个向量生成的循环空间。

证明. 我们已经知道 $\{v_1\}$ 中的元素都形如 $f(A)v_1$，这里 $f(x)$ 是一个多项式。既然已经规定了 B 在生成元 v_1 处的值，因此我们只要“顺水推舟”地定义 B 就好了：对于 $u = f(A)v_1 \in \{v_1\}$，令

$$Bu = f(A)v_2 \in \{v_2\}.$$

而把 B 在其它的 $\{v_i\}$ $(i \neq 1)$ 上的作用都定义为零。这就是一个线性变换，把 v_1 映为 v_2，而且与 A 交换。

这样定义 B 有个问题，就是 $u = f(A)v_1$ 这个表示中 $f(x)$ 不是唯一的，我们还需要验证 Bu 的定义不依赖于 $f(x)$ 的选取。设 $u = f_1(A)v_1 = f_2(A)v_1$，我们要说明 $f_1(A)v_2 = f_2(A)v_2$。由 $f_1(A)v_1 = f_2(A)v_1$ 可得 $p_1(x)|(f_1 - f_2)$，当然就有 $p_2(x)|(f_1 - f_2)$，从而 $f_1(A)v_2 = f_2(A)v_2$。因此我们 B 的定义是合理的。 □

现在我们已经找到了 A 应该满足的必要条件：V 有一个关于 A 的循环向量 v。那么这个条件是不是充分的呢？答案是肯定的：

定理 7.5. 若 V 有一个关于 A 的循环向量 v，则任何与 A 交换的变换都可以写成 A 的多项式。

证明. 我们断言循环空间上一个与 A 交换的变换 B 由它在生成元 v 处的值 Bv 完全决定：由于存在多项式 $f(x)$ 使得 $Bv = f(A)v$，不难验证这时 $B = f(A)$ 成立。 □

从而我们已经从变换的角度完全解决了这一问题。

矩阵的角度

在定理 4.4 中令 $A = B$，那么 $i(A, A)$ 就是与 A 交换的矩阵构成的空间的维数。这个维数总是大于等于 n 的：

$$i(A, A) = \sum_{i=1}^{p}\sum_{j=1}^{p} \deg\Big(f_i(x), f_j(x)\Big) \geq \sum_{i=1}^{p} \deg\Big(f_i(x), f_i(x)\Big) = \sum_{i=1}^{p} \deg f_i(x) = n.$$

A 的多项式也构成一个线性空间，这个空间的维数总是小于等于 n 的，因为其中的元素总是 $I, A, \ldots, A^{n-1}$ 的线性组合。这两个空间的维数相等当且仅当上面式子中的等号成立，这时 A 仅有一个不变因子，也就是 A 的特征多项式等于其极小多项式，这就从矩阵的角度证明了定理 7.5。

例 7.6. 如果矩阵 A 有 n 个互不相同的特征值那么与 A 交换的矩阵一定可以表示为 A 的多项式.

证明. 设 $Av_i = \lambda_i v_i$，那么 $v_1 + \cdots + v_n$ 就是一个循环向量。 □

7.3 双中心化子定理

定理 7.7 (双中心化子). 设 A 是任一给定的方阵，如果方阵 C 和每一个与 A 可交换的方阵都可交换，则 C 可以表示为 A 的多项式。

证明. 首先把 A, C 都看成某个向量空间上的线性变换，如果结论成立的话，显然下面的条件是必须满足的：

引理 7.8. A 的不变子空间直和分解必定也是 C 的不变子空间直和分解。

引理的证明：设 V 在 A 的作用下分解为不变子空间的直和 $V = V_1 \oplus \cdots \oplus V_r$，$P$ 是从 V 到 V_i 上的投影，则有 $AP = PA$，那么根据已知有 $CP = PC$，所以 V_i 也是 C 的不变子空间，引理得证。

我们还是来考虑 V 在 A 的作用下的循环子空间分解

$$V = \{v_1\} \oplus \{v_2\} \oplus \cdots \oplus \{v_r\}.$$

这里的 $v_1, \ldots, v_r$ 满足的条件仍然与定理 7.1 中一致。由于 C 在每个 $\{v_i\}$ 上的限制与 A 交换，所以立刻想到用上定理 7.5 的结论：C 在子空间 $\{v_i\}$ 上可以表示为 A 的多项式 $C = g_i(A)$。问题是这些 $g_i(x)$ 相等吗？如果是同一个多项式那问题就解决了，但是这个不那么显然，所以得分析的再细一点。我们回忆前面证过的结论（参考引理 7.4）

引理 7.9. 对任何 $i > 1$，存在与 A 交换的变换 B 使得 $Bv_1 = v_i$。（道理是一样一样的）

现在是它派上用场的时候了。我们有

$$v_1 \xrightarrow{B} v_i \xrightarrow{C} g_i(A)v_i,$$

$$v_1 \xrightarrow{C} g_1(A)v_1 \xrightarrow{B} g_1(A)v_i.$$

由于 $CB = BC$ 所以两条路径的结果应该是一样的，也就是 $g_i(A)v_i = g_1(A)v_i$。再强调一遍：循环空间上一个与 A 交换的变换由它在生成元处的值完全决定，所以 $g_i(A)$ 和 $g_1(A)$ 在 $\{v_i\}$ 上相同，所以我们可以统一用 $g_1(x)$ 来代替所有的 $g_i(x)$，即 $C = g_1(A)$。 □

这个定理的意义就是在环 $\mathrm{Mat}_n(F)$ 中，对矩阵 A 生成的子环 R 求两次中心化子以后又得到了 R，所以叫做双中心化子定理。双中心化子性质是单代数的核心性质，在结合代数理论中会经常见到它。

8

解题的艺术

这一章选择了一些解法比较巧妙但又不是纯难度的一些题目。如果你是第一次见到它们，还是建议先自己思考一会。毕竟这样的题目比较少，浪费一道少一道。

例 8.1. 伴随矩阵 A^* 总可以表示为 A 的多项式.

证明. 首先 $\mathrm{r}(A) < n-1$ 和 $\mathrm{r}(A) = n$ 的情形是很容易的，$\mathrm{r}(A) = n-1$ 的情形才是关键。下面就考虑这种情形。

显然只要对 A 是 Jordan 标准形的情况加以证明即可。把 A 按特征值分块为

$$A = \begin{pmatrix} B & 0 \\ 0 & J \end{pmatrix},$$

这里 B 是 m 阶可逆矩阵，J 是一个特征值为 0 的 l 阶 Jordan 块，$m + l = n$。注意由于有 $\mathrm{r}(A) = n-1$ 的前提，所以 A 只能有一个特征值为 0 的 Jordan 块。

现在来看 A^*。由于 $AA^* = A^*A = 0$，所以 A^* 也被同时分块为

$$A^* = \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix}.$$

（因为 B 可逆，所以左上角也应该是 0）只需要算 C，也就是 J 的伴随矩阵乘以 $|B|$。而 J 只有左下角的 0 的代数余子式是 $(-1)^{l+1}|B|$，其它的都是 0，所以

$$C = |B|J^* = (-1)^{l+1}|B| \begin{pmatrix} 0 & 0 & \cdots & 1 \\ & 0 & \ddots & \vdots \\ & & \ddots & 0 \\ & & & 0 \end{pmatrix}_{l\times l} = (-1)^{l+1}|B|J^{l-1}.$$

好了，现在的任务是找一个多项式 $f(x)$ 使得

$$f(B) = 0,\quad f(J) = (-1)^{l+1}|B|J^{l-1}.$$

设 B 的特征多项式是 $g(x)$，那么

$$f(x) = (-1)^{n+1}x^{l-1}g(x)$$

就合乎要求（注意 $g(x)$ 的常数项是 $(-1)^m|B|$）。

这里的 $f(x)$ 一眼看不大出来，我是这么找的：设

$$f(x) = a_0 + a_1 x + \cdots + a_{l-1}x^{l-1} + a_l x^l + a_{l+1}x^{l+1} + \cdots,$$

并代入 $f(J) = (-1)^m |B| J^{l-1}$ 就会发现 $a_0 = a_1 = \cdots = a_{l-2} = 0$，$a_{l-1} = (-1)^{l+1}|B|$，所以

$$f(x) = (-1)^{m+l+1} x^{l-1} \Big((-1)^m |B| + a_l x + a_{l+1} x^2 + \cdots \Big).$$

这样就凑出来了。 □

例 8.2. 设 A, B 都是 n 阶矩阵，$C = AB - BA$。如果 $AC = CA$，求证 C 幂零。

这个问题的传统解法是归纳证明 $\operatorname{tr}(C^k) = 0$。不过让我们考虑另外一种解法：考虑 C 的 Jordan 标准形，并且把相同特征值的 Jordan 块合并在一起。即

$$C = \begin{pmatrix} C_1 & & & \\ & C_2 & & \\ & & \ddots & \\ & & & C_r \end{pmatrix}.$$

这样不同的 C_i 没有共同的特征值。这个时候 A 也被分块为

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{pmatrix}.$$

如果这个时候再把 B 划分为 r^2 个子矩阵 $B = (B_{ij})$ 的话，有意思的事情发生了：

$$C_i = A_i B_{ii} - B_{ii} A_i.$$

这说明 C_i 的迹是 0。但是 C_i 是上三角矩阵而且对角线上是它仅有的特征值 λ_i，所以 $\lambda_i = 0$，从而 C 所有特征值都是 0，从而幂零。

例 8.3. 已知方阵 A, B 的特征值都是正数且满足 $A^2 = B^2$，求证 $A = B$。

这个问题作为练习出现过，不了解第四章的结论的话真是很难想出来。

证明. 由已知可得 $A(A-B) = (A-B)(-B)$，所以若 $A - B \neq 0$ 则 A, B 必须有共同的特征值，矛盾！ □

下面这个例子的解法来自 Halmos 的书。

例 8.4. A, B 是数域 F 上的 n 阶方阵，若 $I - BA$ 可逆，则 $I - AB$ 也可逆，并求其逆。

首先用纯形式的推导找出这个逆来：把矩阵看成数，作幂级数展开：

$$\begin{aligned}(I-AB)^{-1} &= I+AB+(AB)^2+\cdots\\ &= I+A(I+BA+(BA)^2+\cdots)B\\ &= I+A(I-BA)^{-1}B.\end{aligned}$$

所以 $(I-AB)^{-1}=I+A(I-BA)^{-1}B$，这就从形式上找出来了结果，剩下的只是验证而已。

这个做法确实很巧妙，而且不用记忆答案，用到的时候立刻就能推出来。

例 8.5. 求证在 n 维欧式空间中两两夹钝角的单位向量的个数的最大值是 $n+1$。

证明. 分两步，第一步证明两两夹钝角的单位向量的个数不超过 $n+1$，第二步证明确实可以达到 $n+1$。两步都要用到归纳法。

首先证明在 n 维欧式空间中不能有超过 $n+1$ 个的两两夹钝角的单位向量。$n=1$ 显然。设小于 n 时结论成立。看 n 的情形：若这时有 $n+2$ 个两两夹钝角的单位向量 $\alpha_1,\ldots,\alpha_{n+2}$。考虑 $\beta_i=\alpha_i-(\alpha_i,\alpha_{n+2})\alpha_{n+2}\ (i=1,2,\ldots,n+1)$，则 $\beta_1,\ldots,\beta_{n+1}$ 都落在 α_{n+2} 的正交补空间中（这个空间维数是 $n-1$），且

$$(\beta_i,\beta_j)=(\beta_i,\alpha_j)=(\alpha_i,\alpha_j)-(\alpha_i,\alpha_{n+2})(\alpha_j,\alpha_{n+2})<0.$$

这与归纳假设矛盾。

其次证明确实有 $n+1$ 个两两夹钝角的单位向量。仍然对维数归纳：$n=1$ 时 α 和 $-\alpha$ 是两个夹钝角的向量。设小于 n 时结论成立，对 n 的情形：先任取一个非零单位向量 β，在 β 的正交补空间中取 n 个互相正交的单位向量 $\alpha_1,\ldots,\alpha_n$。我们证明存在一个负数 λ 使得

$$\alpha_1-\lambda\beta,\quad \ldots\alpha_n-\lambda\beta,\quad \beta$$

是 $n+1$ 个两两夹钝角的向量。这只要令 λ 满足

$$\lambda^2<\min\{-(\alpha_i,\alpha_j),\ 1\le i<j\le n\}$$

即可。□

例 8.6. A 是实方阵且 $AA'=A^2$，求证 A 是对称矩阵。

证明. 迹函数

$$(A,B)\to \mathrm{tr}(AB')=\sum_{i=1}^n\sum_{j=1}^n a_{ij}b_{ij}$$

其实就是 $\mathrm{Mat}_n(\mathbb{R})$ 作为 n^2 维欧式空间上的标准内积。特别地，

$$\mathrm{tr}(AA')=\sum_{i=1}^n\sum_{j=1}^n a_{ij}^2.$$

因此要证明 $A-A'$ 是零矩阵，只需验证 $\mathrm{tr}[(A-A')(A-A')']=0$ 即可。□

例 8.7. 设 A 是一个实对称矩阵，正负惯性指数分别为 p,q，T 是一个实矩阵。问矩阵 $B=T'AT$ 的正负惯性指数与 A 相比是大还是小？

这个题目我不写过程了，大家可以参考北大代数几何教研室编的《高等代数》第三版 234 页的补充题 3，这两个题目是一回事。答案是：当用一个不可逆的矩阵 T 给一个对称矩阵 A 做“合同变换”的话，正负惯性指数都会减小（不一定是严格减小）。

最后这个题目来自 Horn 的书：

例 8.8. 设 V 是一个向量空间，是否存在 V 上的线性变换 A,B 使得 $AB-BA=I$？

答案不难，V 是有限维的话就没有，两边求迹就得出矛盾。V 是无限维的话可能有，比如在 $C^1(\mathbb{R})$ 上定义

$$Af=f'(x)，Bf=xf(x)，$$

不难验证就有 $AB-BA=I$ 成立。

H.Wielandt 给出了如下证明：

如果有满足条件的 A,B 存在，不难用归纳法证明

$$A^{k+1}B-BA^{k+1}=(k+1)A^k.$$

设 $\|\cdot\|$ 是 $\mathrm{Mat}_n(F)$ 上任一范数，那么

$$(k+1)\|A^k\|\leq 2\|A^{k+1}\|\cdot\|B\|\leq 2\|A^k\|\cdot\|A\|\cdot\|B\|.$$

这样必有某个 k 使得 $\|A^k\|=0$。但是上面第一个不等号告诉我们 A^{k-1} 的范数被 A^k 的范数所控制，所以 $A^k=0\Rightarrow A^{k-1}=0\Rightarrow\cdots\Rightarrow A=0$，这就导出了矛盾。

这个证明的好处就在于它可以应用到 Banach 空间上去：在量子力学中有如下方程

$$XP-PX=\frac{ih}{2\pi}I.$$

X 是位置算子，P 是动量算子，h 是普朗克常数。Wielandt 的证明告诉我们这样的算子一定不是有界算子，这解决了一个物理学中长久以来悬而未决的问题。