

第一章 线性变换的表示与分解

在上一章中，我们研究了线性映射的基础理论。本章中，我们进一步，寻求对线性映射的表示和分解。

我们已经看到：选定一个映射和两个基，就可以写出矩阵。但是，其实这些矩阵中有些是容易使用的，而有些则是难以处理的。如果有一个基下映射的矩阵我们觉得是良好的，我们常常会称这类矩阵为标准型

本章中我们不研究所有线性映射，而是聚焦一种特殊的线性映射——线性变换。我们会看到，如果域是代数闭的，在特定条件下我们可以将其写成对角矩阵；而更一般地，它一定可以写成 Jordan 块组成的分块对角矩阵。

本章的第一节我们讲述对角标准型的基础工作——特征值、特征向量；第二节讲解对角标准型的定义及判定是否可对角化的方式；

第三至第七节，我们逐渐搭建出 Jordan 标准型的理论。

1.1 线性变换的特征值和特征向量

在本节和下一节中，我们将解决对角标准型的问题。

对角标准型的想法是：尽管并非所有线性映射呈现出 $A(\alpha) = \lambda\alpha$ 的形式，但是很多线性映射都呈现为将基向量仅仅进行伸缩的形式。这样的线性映射会在指定基上呈现为对角矩阵，其对角线元素我们一般称为特征值。

我们首先构建特征值和其基向量——特征向量的理论。

1.1.1 特征值和特征向量的定义

定义 1.1.1: 线性变换的特征值、特征向量、特征子空间

设 V 是一个 F 上的线性空间, $A \in \text{hom}(V, V)$

那么如果 $\exists \lambda \in F, \alpha \in V, \alpha \neq \mathbf{0}$ 使得

$$A(\alpha) = \lambda\alpha \quad (1.1)$$

成立

那么我们称 λ 是 A 的一个特征值, α 是 A 的隶属于 λ 的一个特征向量

同时我们定义:

$$V_\lambda := \{\alpha | A(\alpha) = \lambda\alpha\} \quad (1.2)$$

称为 A 的属于特征值 λ 的特征子空间

特征值代表着一个线性映射对一个向量的伸缩程度, 这里我们要求至少能找到一个向量不为零是因为: 如果一个特征值只对于零向量成立, 那么它过于平凡, 而且会扰乱后续我们一些关于特征值数量的命题

类似地, 我们自然可以定义矩阵的特征值

定义 1.1.2: 矩阵的特征值、特征向量

设 $A \in M_n(F)$

那么如果 $\exists \lambda \in F, \alpha \in F^n$, 使得

$$A\alpha = \lambda\alpha \quad (1.3)$$

成立。

那么我们称 λ 是 A 的一个特征值, α 是 A 的隶属于 λ 的一个特征向量

1.1.2 特征值和特征向量的性质

我们探讨一些相关的基本性质

1. 我们首先验证, 特征子空间的确是子空间

命题 1.1.1. $\forall A \in \text{hom}(V, V)$, V_λ 是 V 的一个线性子空间

证明: 取 $\forall \alpha, \beta \in V_\lambda, k \in F$

$$A(\alpha + \beta) = \lambda\alpha + \lambda\beta = \lambda(\alpha + \beta) \Rightarrow \alpha + \beta \in V_\lambda$$

$$A(k\alpha) = k\lambda\alpha = \lambda(k\alpha) \Rightarrow k\alpha \in V_\lambda$$

于是命题得证 □

请注意: 特征子空间并非全体特征向量的集合, 而是全体特征向量和零向量的集合

2. 从特征值的定义可以看出, 特征值就像是把线性变换的一部分转换为纯量乘法。我们猜想: 每一个特征值体现了映射的不同的“方向”, 不同特征子空间的线性无关向量组的并也应该线性无关

以下命题证实了这个猜想

命题 1.1.2. 设 $A \in \text{hom}(V, V)$, λ_1, λ_2 是 A 的两个特征值, $\lambda_1 \neq \lambda_2$

如果 $\alpha_1, \dots, \alpha_m \in V_{\lambda_1}$ 线性无关, $\beta_1, \dots, \beta_n \in V_{\lambda_2}$ 线性无关

那么 $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n$ 也线性无关

证明: 取线性组合并设其为零:

$$k_1\alpha_1 + \dots + k_m\alpha_m + l_1\beta_1 + \dots + l_n\beta_n = \mathbf{0}$$

$$\text{将 } A \text{ 在其上进行变换得: } k_1A(\alpha_1) + \dots + k_mA(\alpha_m) + l_1A(\beta_1) + \dots + l_nA(\beta_n) = \mathbf{0}$$

$$\Rightarrow k_1\lambda_1\alpha_1 + \dots + k_m\lambda_1\alpha_m + l_1\lambda_2\beta_1 + \dots + l_n\lambda_2\beta_n = \mathbf{0}$$

但是, 如果我们把最初的线性组合乘以 λ_1 , 得:

$$k_1\lambda_1\alpha_1 + \dots + k_m\lambda_1\alpha_m + l_1\lambda_1\beta_1 + \dots + l_n\lambda_1\beta_n = \mathbf{0}$$

将两式相减, 得:

$$l_1(\lambda_2 - \lambda_1)\beta_1 + \dots + l_n(\lambda_2 - \lambda_1)\beta_n = \mathbf{0}$$

$$\Rightarrow \forall i, l_i(\lambda_2 - \lambda_1) = 0 \Rightarrow l_i = 0$$

$$\Rightarrow k_1\alpha_1 + \dots + k_m\alpha_m = \mathbf{0} \Rightarrow \forall i, k_i = 0$$

于是命题得证 □

一个显然的推论是此结论的 n 个子空间的版本:

推论 1.1.1

$A \in \text{hom}(V, V)$, $\lambda_1, \dots, \lambda_n$ 是 A 的 n 个互不相同的特征值

如果 $\forall i, \alpha_{ir_1}, \dots, \alpha_{ir_m} \in V_{\lambda_i}$ 线性无关

那么向量组 $\{\alpha_{jr_k}\}$ 也线性无关

证明: 由数学归纳法易证 □

3. 容易注意到, 矩阵的特征值和线性映射的特征值其实是一样的, 正如下面的命题:

命题 1.1.3. 设 V 是一个 F 上的线性空间, $\dim V = n < \aleph_0$, $\{\alpha_1, \dots, \alpha_n\}$ 是 V 的一个基

$A \in \text{hom}(V, V)$ 在 $\{\alpha_1, \dots, \alpha_n\}$ 下的矩阵是 A

那么, λ 是 A 的一个特征值, α 是 A 的一个隶属于 λ 的一个特征向量 \Leftrightarrow

λ 是 A 的一个特征值, 并且 α 在 $\{\alpha_1, \dots, \alpha_n\}$ 下的坐标 \mathbf{x} 是 A 的一个隶属于 λ 的特征向量

证明: $A(\alpha) = \lambda\alpha$

$\Leftrightarrow A\mathbf{x} = \lambda\mathbf{x}$ (参见命题 2.3.23)

于是命题得证。 □

1.1.3 特征矩阵与特征多项式

前面我们研究了特征值的性质, 接下来我们想知道: 是否可以直接去寻找计算特征值的直接方法? 事实上, 是可以的, 我们指出: 特征值是特征多项式的一个根, 而特征向量是对应映射的核的一个元素

先给出定义:

定义 1.1.3: 线性变换的特征多项式

设 $A \in \text{hom}(V, V)$, 我们称

$$\chi_A(\lambda) = \det(\lambda I - A)$$

为 A 的特征多项式

类似地, 我们可以定义矩阵的特征矩阵和特征多项式:

定义 1.1.4: 矩阵的特征矩阵和特征多项式

设 $A \in M_n(F)$, 我们称 $\lambda I - A$ 是 A 的特征矩阵

并称 $\chi_A(\lambda) = \det(\lambda I - A)$ 是 A 的特征多项式

下面的性质指出了我们想要的结果:

1.

命题 1.1.4. 设 V 是一个 F 上的线性空间, $\dim V = n < \aleph_0$, $\{\alpha_1, \dots, \alpha_n\}$

如果 $A \in \text{hom}(V, V)$ 在 $\{\alpha_1, \dots, \alpha_n\}$ 下的矩阵是 A

那么我们断言: $\chi_A(\lambda) = \chi_A(\lambda)$

证明: 这是显然的, 因为矩阵的行列式的定义就是其矩阵的行列式 □

2.

命题 1.1.5. $\forall A \in M_n(F), \chi_A \in F[\lambda]$

证明: 设 $A = (a_{ij}) \in M_n(F)$

$$\text{那么, } \lambda I - A = \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix}$$

由行列式的置换展开可知, $\chi_A(\lambda) \in F[\lambda]$ □

推论 1.1.2

$\forall A \in \text{hom}(V, V), \dim V = n < \aleph_0, \chi_A(\lambda) \in F[\lambda]$

证明: 这是显然的, 因为我们只需要任取一个基 $\{\alpha_1, \dots, \alpha_n\}$, 并利用 A 在此基下的矩阵 A

利用前面线性映射与其矩阵的特征多项式相同的命题, 即可得证。 □

定理 1.1.3: 线性变换的特征值即是特征多项式在域内的根

设 V 是一个 F 上的 n 维线性空间, $A \in \text{hom}(V, V)$

那么: λ 是 A 的一个特征值, $\alpha \in V_\lambda \Leftrightarrow \chi_A(\lambda) = 0, \alpha \in \ker(\lambda I - A)$

3.

证明: λ 是 A 的一个特征值, $\alpha \in V_\lambda$

$$\Leftrightarrow A(\alpha) = \lambda\alpha, V_\lambda \neq \{0\}$$

$$\Leftrightarrow (\lambda I - A)(\alpha) = 0, \ker(\lambda I - A) \neq \{0\}$$

$$\Leftrightarrow \text{rank}(\lambda I - A) < n, \alpha \in \ker(\lambda I - A)$$

$$\Leftrightarrow \chi_A(\lambda) = 0, \alpha \in \ker(\lambda I - A)$$
 □

其自然推论是其矩阵版本:

推论 1.1.4

设 $A \in M_n(F)$

那么: λ 是 A 的一个特征值, $\mathbf{x} \in F^n$ 是 A 从属于 λ 的特征向量 $\Leftrightarrow \chi_A(\lambda) = 0, (\lambda I - A)\mathbf{x} = \mathbf{0}, \mathbf{x} \neq \mathbf{0}$

这个推论也指出了一下结果:

推论 1.1.5

设 $A, B \in M_n(F), A \sim B$

那么 A 和 B 具有相同的特征值, 并且 $\chi_A(\lambda) = \chi_B(\lambda)$

我们之后不再完全讨论矩阵版本的特征值相关定理, 因为上面的结果已经说明了: 相似矩阵, 以及相互对应的矩阵和映射在特征值理论中并无区别

4.

命题 1.1.6. 设 V 是一个 F 上的 n 维线性空间, $A \in \text{hom}(V, V)$

那么:

$$\chi_A(\lambda) = \lambda^n - \text{tr}(A)\lambda^{n-1} + \cdots + (-1)^{n-k} \left(\sum_{j_1 < \cdots < j_{n-k}} A_{j_1, \cdots, j_{n-k}}^{j_1, \cdots, j_{n-k}} \right) \lambda^k + \cdots + (-1)^n \det(A) \quad (1.4)$$

证明: 事实上, λ^k 项即是下面的行列式求和: $\sum_{j'_1 < \cdots < j'_k} \begin{vmatrix} -a_{11} & \cdots & 0 & \cdots & 0 & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ -1 & \cdots & \lambda & \cdots & 0 & \cdots & 1 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ -1 & \cdots & 0 & \cdots & \lambda & \cdots & 1 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & 0 & \cdots & 0 & \cdots & -a_{nn} \end{vmatrix}$ 其

中行列式的 j'_1, \cdots, j'_k 列为仅在第 j_i 个元素为 λ , 其他位置都是 0 的列。

这是因为, 如果想要产生 λ^k , 那么求和中的每一项必须选取对角线上的 k 个元素; 同时, 如果在进一步展开中选取了 $-a_{ii}$ 而不是 λ , 那么会导致次数降低

注意到:

$$\begin{vmatrix}
 -a_{11} & \cdots & 0 & \cdots & 0 & \cdots & -a_{1n} \\
 \vdots & \ddots & \vdots & & \vdots & & \vdots \\
 -1 & \cdots & \lambda & \cdots & 0 & \cdots & 1 \\
 \vdots & & \vdots & \ddots & \vdots & & \vdots \\
 -1 & \cdots & 0 & \cdots & \lambda & \cdots & 1 \\
 \vdots & & \vdots & & \vdots & \ddots & \vdots \\
 -a_{n1} & \cdots & 0 & \cdots & 0 & \cdots & -a_{nn}
 \end{vmatrix}$$

$$= (-1)^{(j'_1 + \cdots + j'_k) + (j'_1 + \cdots + j'_k)} \det(\text{diag}\{\lambda, \cdots, \lambda\}) (-1)^{n-k} A_{j_1, \cdots, j_{n-k}}^{j'_1, \cdots, j'_{n-k}}$$

这个结果只需要对 j'_1, \cdots, j'_k 列展开即可得出, 其中 $(-1)^{n-k}$ 是将 $-a_{ij}$ 前面的负号提出去得到的, 而 j_1, \cdots, j_{n-k} 是与 j'_1, \cdots, j'_k 互补的有限序列

于是命题得证。 \square

其推论是以下结论:

推论 1.1.6

设 F 是一个代数闭域, V 是一个 F 上的 n 维线性空间, $A \in \text{hom}(V, V)$

设 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征值 ($\chi_A(\lambda)$ 中的重根按重数计算)

那么有:

$$\prod_{i=1}^n \lambda_i = \det(A) \quad (1.5)$$

$$\sum_{i=1}^n \lambda_i = \text{tr}(A) \quad (1.6)$$

证明: 对 $\chi_A(\lambda)$ 作唯一分解:

$$\begin{aligned}
 \chi_A(\lambda) &= \prod_{i=1}^n (\lambda - \lambda_i) \\
 &= \lambda^n - (\lambda_1 + \cdots + \lambda_n) \lambda^{n-1} + \cdots + (-1)^n \prod_{i=1}^n \lambda_i
 \end{aligned}$$

结合前面的命题既得。 \square

请注意: 这个命题必须要求 F 代数闭, 否则可能 $\chi_A(\lambda)$ 的根并非全体特征值, 而导致论证失效。

1.2 线性变换的对角标准型

本节中我们构建对角标准型的理论。对角标准型的理论是完全自然的，我们无需更多讨论，只需给出定义并直接推导充要条件即可。

1.2.1 对角标准型的定义

定义 1.2.1: 线性映射的对角标准型

设 V 是一个 F 上的一个线性空间， $\dim V = n < \aleph_0$

如果存在一个基 $\{\alpha_1, \dots, \alpha_n\}$ ，使 A 在其上的矩阵 \mathbf{A} 是一个对角矩阵

那么我们称 \mathbf{A} 为 A 的对角标准型，此时称 A 可对角化。

类似地，我们可以定义矩阵的对角标准型

定义 1.2.2: 矩阵的对角标准型

设 $\mathbf{A} \in M_n(F)$ ，如果存在 $\mathbf{P} \in GL_n(F)$ 和对角矩阵 $\mathbf{D} \in M_n(F)$ ，使得

$$\mathbf{A} = \mathbf{P}^{-1} \mathbf{D} \mathbf{P}$$

那么我们称 \mathbf{D} 为 \mathbf{A} 的对角标准型，此时称 \mathbf{A} 可对角化。

容易看出，线性映射的对角标准型与矩阵的对角标准型是等价的。特别地，如果认为 \mathbf{A} 是 $A \in \text{hom}(F^n, F^n)$ 在 $\{\mathbf{e}_i\}$ 下的矩阵，那么容易发现：其实 \mathbf{P}^{-1} 的列向量就是 A 的特征向量。

1.2.2 线性变换可对角化的条件

观察对角矩阵的结构，既得以下显然的条件：

定理 1.2.1: 线性映射可对角化的条件 (1)

设 V 是一个 F 上的一个线性空间， $\dim V = n < \aleph_0$

那么： $A \in \text{hom}(V, V)$ 可对角化 \Leftrightarrow 存在一个由 A 的特征向量组成的 V 的基

证明： $A \in \text{hom}(V, V)$ 可对角化

\Leftrightarrow 存在一个基 $\{\alpha_1, \dots, \alpha_n\}$ ，使得 A 在其上的矩阵为 $\text{diag}\{\lambda_1, \dots, \lambda_n\}$

$\Leftrightarrow A(\alpha_1, \dots, \alpha_n) = \text{diag}\{\lambda_1, \dots, \lambda_n\}(\alpha_1, \dots, \alpha_n)$

$$\Leftrightarrow A(\alpha_1, \dots, \alpha_n) = (\lambda_1 \alpha_1, \dots, \lambda_n \alpha_n)$$

$$\Leftrightarrow \lambda_i \text{ 是 } A \text{ 的特征值, } \alpha_i \in V_{\lambda_i}$$

$$\Leftrightarrow \text{存在一个由 } A \text{ 的特征向量组成的 } V \text{ 的基}$$

□

推论 1.2.2: 线性映射可对角化的条件 (2)

设 V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

那么: $A \in \text{hom}(V, V)$ 可对角化 \Leftrightarrow 存在 n 个线性无关的特征向量

证明: 这是显然的, 因为 V 中的 n 个线性无关的向量一定组成一个基

□

定理 1.2.3: 线性映射可对角化的条件 (3)

设 V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

那么: $A \in \text{hom}(V, V)$ 可对角化 $\Leftrightarrow A$ 的全体不同的特征值 $\lambda_1, \dots, \lambda_s$, 满足 $V_{\lambda_1}, \dots, V_{\lambda_s}$ 的基的并集是 V 的基

证明: 必要性由之前的命题是显然的。

接下来证明充分性。

如果 A 可对角化, 那么, 它一定能选取一个由特征向量组成的基 $\{\alpha_{11}, \dots, \alpha_{1r_1}, \dots, \alpha_{s1}, \dots, \alpha_{sr_s}\}$,

其中 $\alpha_{ij} \in V_{\lambda_i}, r_1 + \dots + r_s = n$

我们只需证明, 其实 $\{\alpha_{i1}, \dots, \alpha_{ir_i}\}$ 就是 V_{λ_i} 的一个基

首先, 它一定线性无关, 因为它是 V 的一个基的一个子集。不妨假设它不是 V_{λ_i} 的基, 那么, 一定 $\exists \beta \in V_{\lambda_i}$, 使得它不能被 $\{\alpha_{i1}, \dots, \alpha_{ir_i}\}$ 线性表出

那么, 一定有 $\{\alpha_{i1}, \dots, \alpha_{ir_i}, \beta\}$ 线性无关。因为如果它线性相关, 那么 $\{\alpha_{i1}, \dots, \alpha_{ir_i}\}$ 一定是它的一个极大线性无关组, 这使得 β 可以被 $\{\alpha_{i1}, \dots, \alpha_{ir_i}\}$ 线性表出, 矛盾。

但是, 我们之前已经指出, 若干个特征子空间的线性无关向量的并集必定也是线性无关的, 这说明 $\alpha_{11}, \dots, \alpha_{1r_1}, \dots, \alpha_{s1}, \dots, \alpha_{sr_s}, \beta$ 线性无关。但是 $\dim V = n$, 其中的 $n+1$ 个向量不可能线性无关, 矛盾。

于是命题得证

□

其推论是接下来的两个条件:

推论 1.2.4: 线性映射可对角化的条件 (4)

设 V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

那么: $A \in \text{hom}(V, V)$ 可对角化 \Leftrightarrow 如果 $\lambda_1, \dots, \lambda_s$ 是 A 的全体不同的特征值, 那么 $V_{\lambda_1} \oplus \dots \oplus V_{\lambda_s} = V$

证明: 这是显然的, 因为前面的命题指出了 V 的一个基正是 V_{λ_i} 的基组成的, 而这正是直和的充要条件。□

推论 1.2.5: 线性映射可对角化的条件 (5)

V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

那么: $A \in \text{hom}(V, V)$ 可对角化 \Leftrightarrow 如果 $\lambda_1, \dots, \lambda_n$ 是 A 的全体不同的特征值, 那么 $\dim V_{\lambda_1} + \dots + \dim V_{\lambda_s} = n$

证明: 充分性由上面的命题是显然的。

必要性只需注意到前面命题的结果, 即存在一个 V 的基由全体特征子空间的基组成即可。□

至此, 我们基本上得到了全部结果, 最后我们将看到一个最终结果, 它也是最常用的充要条件

其实, 我们已经注意到: 如果特征子空间的维数之和不足, 那么它就不可对角化。我们的一个天然的猜想是: 特征子空间的维数是不是受限于特征值的重数? 如果是的话, 其实只需要填满这个重数就可以对角化。

现在给出一个引理

引理 1.2.6: 几何重数小于代数重数

设 V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

设 $\lambda_i \in F$ 是 $A \in \text{hom}(V, V)$ 的一个特征值, 并且其特征多项式有分解:

$$\chi_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$$

那么我们断言: $\dim V_{\lambda_i} \leq r_i$

其中 $\dim V_{\lambda_i}$ 也常常称为 λ_i 的几何重数, r_i 称为 λ_i 的代数重数

证明: 不妨设 $\dim V_{\lambda_i} = d$

取 V_{λ_i} 的一个基 $\{\alpha_1, \dots, \alpha_d\}$, 并补全为 V 的一个基 $\{\alpha_1, \dots, \alpha_d, \dots, \alpha_n\}$

考虑 A 在此基上的矩阵 A , 它一定具备以下形式:

$$A = \begin{pmatrix} \lambda_i I_d & B \\ O & C \end{pmatrix}$$

因为 A 在 $\{\alpha_1, \dots, \alpha_d\}$ 上的作用即是按照相同的特征值 λ_i 进行纯量乘法

$$\text{那么, 特征矩阵 } \lambda I - A = \begin{pmatrix} (\lambda - \lambda_i) I_d & -B \\ O & \lambda I_{n-d} - C \end{pmatrix}$$

于是有: $\chi_A(\lambda) = \det((\lambda - \lambda_i) I_d) \det(I_{n-d} - C)$

$$= (\lambda - \lambda_i)^d \det(I_d) \det(I_{n-d} - C)$$

因此一定有 $d \leq r_i$, 命题得证 □

那么有以下命题:

定理 1.2.7: 线性映射可对角化的条件 (6)

设 F 是一个代数闭域, V 是一个 F 上的一个线性空间, $\dim V = n < \aleph_0$

那么 $A \in \text{hom}(V, V)$ 可对角化 \Leftrightarrow 如果 $\chi_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$, 那么 $\dim V_{\lambda_i} = r_i$

证明: 必要性是显然的, 因为这正是前面的命题给出的结果

充分性只需注意到, 如果可对角化必定有 $\dim V_{\lambda_1} + \dots + \dim V_{\lambda_s}$, 但是引理已经指出 $\dim V_{\lambda_i} \leq r_i$, 又因为 $r_1 + \dots + r_s = n$, 那么一定有 $\dim V_{\lambda_i} = r_i$

于是命题得证。 □

对角标准型的性质就是对角矩阵的性质, 它们都很显然故无需阐述。可以看出来, 对角标准型的条件是极其严苛的——这也是它优良性质的缘由。

1.3 线性变换的根子空间、不变子空间

我们转向不可对角化的线性映射的表示。

我们观察可对角化的条件 (4), 它的另一种表述是:

$$V = \ker(\lambda_1 I - A) \oplus \dots \oplus \ker(\lambda_s I - A)$$

但是, 我们知道, 直和并不总是成立。我们的一个想法是: 如果我们调整 $\lambda_i I - A$ 的幂, 是否能让它构成直和呢? 如果能构成直和, 这个幂有办法快速求出来吗?

一个相当胡扯的想法是: 特征多项式的分解 $\chi_A(\lambda) = (\lambda - \lambda_1)^{r_1} \dots (\lambda - \lambda_s)^{r_s}$ 似乎很符合我们的结构

我们从特征多项式在映射上的作用，及其分解入手。

1.3.1 线性变换的多项式

此处我们希望讨论的，其实是一个具体的映射的多项式，也就是以下定义：

定义 1.3.1: 线性映射的行列式

设 F 是一个域， $A \in \text{hom}(V, V)$

我们定义：

$$F[A] := \left\{ \sum_{k=0}^n a_k A^k \mid a_k \in F, n \in \mathbb{N} \right\} \quad (1.7)$$

称为 A 的多项式环，其元素称为 A 的多项式

一个显然的事实是： $F[A]$ 是 $\text{hom}(V, V)$ 的子环

接下来探讨线性变换的多项式的性质。

1. Hamilton-Caylay 定理

定理 1.3.1: Hamilton-Caylay 定理

设 V 是一个 F 上的 n 维线性空间， $A \in \text{hom}(V, V)$

那么有：

$$\chi_A(A) = 0 \quad (1.8)$$

证明： 设 M 是一个 $F[A]$ 上的一个模，其元素和 V 一致，保持 V 中的加法，而其纯量乘法则为：

$$f(A) \cdot \alpha := f(A)(\alpha), f(A) \in F[A], \alpha \in V$$

那么容易注意到，此时 $A \in \text{hom}(M, M)$ ，即成为 M 的一个自同态

接下来我们尝试给出 $\det(\lambda I - A)$ 在自同态下的一个表示。此处，我们尝试借助矩阵来完成这一步

由于 M 其实和 V 有相同的元素，并且 F 本就是 $F[A]$ 的子环，因此只需取 V 的一个基 $\{e_1, \dots, e_n\}$ ，它既是 M 的生成元

那么，一定能写出 A 的在这一组生成元下的矩阵 $\mathbf{A} \in M_n(F[A])$

那么， $\det(\lambda I - A) = \det(\lambda \mathbf{I} - \mathbf{A})$ （请注意，此处的 \mathbf{I}, \mathbf{A} 是环上的矩阵，而非域上的矩阵）

此时只需考察 $\chi_A(A) = \det(AI - A)$ (由于此时矩阵已经是建立在环 $F[A]$ 上的了, A 成为了纯量, 可以直接代入)

记 $P = AI - A, \text{adj}(P) = (\frac{1}{|P|} P_{N-\{j\}}^{\mathbb{N}-\{i\}})$, 并注意:

$$\chi_A(A)(e_1, \dots, e_n) = (e_1, \dots, e_n) \det(P) = (e_1, \dots, e_n) P \text{adj}(P)$$

$$(A(e_1, \dots, e_n) - A(e_1, \dots, e_n)) \text{adj}(P) = 0$$

于是命题得证。 \square

这个定理看起来非常令人惊讶: 特征多项式竟然可以直接把映射本身化为零。

这个定理的一个常见的伪证是: 直接把 A 代入, 得到 $\det(AI - A) = \det(0) = 0$ 。但是, 其实 λ 是一个纯量, 并不能将映射代入。

我们此处的证明正是基于这种想法进行了一个修改: 将 F 视为环 $F[A]$ 的子环, 从而实现了正确的代入。

2.

命题 1.3.1. 设 $A \in \text{hom}(V, V), f(x), f_1(x), f_2(x) \in F[x], f(x) = f_1(x)f_2(x), (f_1(x), f_2(x)) = 1$

那么有: $\ker f(A) = \ker f_1(A) \oplus \ker f_2(A)$

证明: 我们首先证明, $\ker f_1(A)$ 和 $\ker f_2(A)$ 的确是 $\ker f(A)$ 的子空间

$$\forall \alpha \in \ker f_1(A)$$

$$\Rightarrow f_1(A)(\alpha) = 0 \Rightarrow f(A)(\alpha) = (f_2(A)f_1(A))(\alpha) = 0 \Rightarrow \alpha \in \ker f(A) \Rightarrow \ker f_1(A) \subseteq \ker f(A)$$

同理可证 $\ker f_2(A) \subseteq \ker f(A)$

于是我们证明了 $\ker f_1(A) + \ker f_2(A) \subseteq \ker f(A)$ 。现在我们证明 $\ker f(A) \subseteq \ker f_1(A) + \ker f_2(A)$

$$\forall \alpha \in \ker f(A)$$

由于 $(f_1(x), f_2(x)) = 1$, 一定存在 $u(x), v(x) \in F[x], u(x)f_1(x) + v(x)f_2(x) = 1$

于是有: $\alpha = I(\alpha) = (u(A)f_1(A) + v(A)f_2(A))(\alpha) = (u(A)f_1(A))(\alpha) + (v(A)f_2(A))(\alpha)$

记 $\alpha_1 = (u(A)f_1(A))(\alpha), \alpha_2 = (v(A)f_2(A))(\alpha)$

注意到: $f_2(A)(\alpha_1) = (f_2(A)u(A)f_1(A))(\alpha) = (u(A)f(A))(\alpha) = u(A)(0) = 0 \Rightarrow \alpha_1 \in \ker f_2(A)$

同理可证 $\alpha_2 \in \ker f_1(A)$ 。于是我们证明了 $\ker f(A) = \ker f_1(A) + \ker f_2(A)$

我们最后证明它们的确构成直和。

$$\forall \beta \in \ker f_1(A) \cap \ker f_2(A)$$

$$\text{注意到: } \beta = I(\beta) = (u(A)f_1(A) + v(A)f_2(A))(\beta) = (u(A)f_1(A))(\beta) + (v(A)f_2(A))(\beta) = 0$$

于是命题得证。 \square

这个命题的直接推论是：

推论 1.3.2: –

设 $A \in \text{hom}(V, V)$, $f(x), f_1(x), \dots, f_s(x) \in F[x]$, $f(x) = \prod_{i=1}^s f_i(x)$,
 $\forall i, j, (f_i(x), f_j(x)) = 1$
 那么有: $\ker f(A) = \bigoplus_{i=1}^s \ker f_i(A)$

至此，其实我们已经可以看出来一些方向了。我们已经证明了多项式如果分解为多个互素多项式，那么他们的核也可以构成直和。我们又注意到，任意映射在特征多项式下为零。

那么，其实特征多项式的唯一分解一定可以直和为零映射的核——即线性空间本身，这正是我们所期望的结果

1.3.2 线性变换的根子空间

我们将这种理想的子空间为根子空间，因为它的常数是从特征多项式求出来的

定义 1.3.2: 根子空间

设 F 是一个代数闭域， V 是 F 上的一个线性空间， $\dim V = n < \aleph_0$

$$\chi_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$$

那么我们称 $\ker(A - \lambda_i I)^{r_i}$ 为 A 的一个根子空间（又称广义特征子空间）

我们现在正式叙述我们观察到的结果

定理 1.3.3: 子空间分解定理

设 F 是一个代数闭域, V 是 F 上的一个线性空间, $\dim V = n < \aleph_0$

$$\chi_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$$

那么有:

$$V = \bigoplus_{i=1}^s \ker(A - \lambda_i I)^{r_i} \quad (1.9)$$

至此, 我们证实了, 代数闭域上的任意有限维线性映射, 都可以形成一个根子空间分解。

接下来我们转向矩阵表示。在之前的对角标准型中, 从特征值的定义可以看出来, 每一个特征子空间表示为一个对角线上的元素。那么, 现在我们希望: A 在根子空间上的限制, 都可以写成一个矩阵; 特别地, 如果每一个限制所属的子空间不同, 此时我们便有可能写成分块对角矩阵

1.3.3 线性变换的不变子空间

一个首要的问题是: $A|_{\ker(A - \lambda_i I)^{r_i}}$ 到底能不能写成一个方阵的形式。如果它可以, 那么, 我们将不仅可以把映射写成分块对角矩阵, 而且每一个块的具体形式可以借助线性变换来求得。

总之, 我们希望验证: 是否任意代数闭域上的线性变换在根子空间上的限制是一个线性变换, 而且这些线性变换在指定基下的矩阵, 恰好可以分块地组成线性变换的矩阵

我们首先给出不变子空间的定义, 它其实就是使得变换的限制依旧为变换的子空间

定义 1.3.3: 不变子空间

设 V 是 F 上的一个线性空间, $A \in \text{hom}(V, V)$

如果 V 的子空间 $W \subseteq V$ 满足:

$$\forall \alpha \in W, A(\alpha) \in W$$

那么我们称 W 是 A 的一个不变子空间

显然, V 和 $\{0\}$ 都是 A 的一个不变子空间, 它们一般被称为是平凡的不变子空间

接下来我们给出一些性质

1.

命题 1.3.2. 设 V 是 F 上的一个线性空间, $A \in \text{hom}(V, V)$, λ 是 A 的一个特征值

那么 $\ker A, \operatorname{Im} A, V_\lambda$ 是 A 的不变子空间

证明: $\forall \alpha \in \ker A, A(\alpha) = \mathbf{0} \in \ker A$

$\forall \alpha \in \operatorname{Im} A, A(\alpha) \subseteq \operatorname{Im} A$

$\forall \alpha \in V_\lambda, A(\alpha) = \lambda\alpha \in V_\lambda$

□

2.

命题 1.3.3. 设 V 是 F 上的一个线性空间, $A, B \in \operatorname{hom}(V, V), AB = BA$, λ 是 A 的一个特征值

那么 $\ker A, \operatorname{Im} A, V_\lambda$ 是 B 的不变子空间

证明: $\forall \alpha \in \ker A, A(B(\alpha)) = B(A(\alpha)) = B(\mathbf{0}) = \mathbf{0} \Rightarrow B(\alpha) \in \ker A$

$\forall A(\alpha) \in \operatorname{Im} A, B(A(\alpha)) = A(B(\alpha)) \in \operatorname{Im} A$

$\forall \alpha \in V_\lambda, A(B(\alpha)) = B(A(\alpha)) = B(\lambda\alpha) = \lambda B(\alpha) \Rightarrow B(\alpha) \in V_\lambda$

于是命题得证。

□

3.

命题 1.3.4. 一个线性变换的两个不变子空间的和与交, 也是不变子空间

证明: 设 $A \in \operatorname{hom}(V, V)$, W, U 是 A 的不变子空间

注意到: $\forall \alpha + \beta \in W + U, \alpha \in W, \beta \in U, A(\alpha + \beta) = A(\alpha) + A(\beta) \in W + U$

$\forall \gamma \in W \cap U, A(\gamma) \in W, A(\gamma) \in U \Rightarrow A(\gamma) \in W \cap U$

于是命题得证

□

4. 最后一条性质, 是我们所希望的结果: 即如果把映射限制在一系列构成直和的不变子空间, 那么矩阵也可以自然构成分块

命题 1.3.5. 设 V 是 F 上的一个线性空间, $\dim V = n < \aleph_0, A \in \operatorname{hom}(V, V)$, 那么:

$V = W_1 \oplus \cdots \oplus W_s$, 其中 W_i 是 A 的不变子空间, 且 $A|_{W_i}$ 在基 $B_i = \{\alpha_{i1}, \cdots, \alpha_{ir_i}\}$ 下的矩阵为 A_i

$\Leftrightarrow A$ 在 $\{\alpha_{11}, \cdots, \alpha_{1r_1}, \cdots, \alpha_{s1}, \cdots, \alpha_{sr_s}\}$ 下的矩阵为

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_s \end{pmatrix}$$

证明: 先证充分性。

首先, 由于 W_1, \cdots, W_s 构成直和, 因此 $B_1 \cup \cdots \cup B_s$ 的确是 V 的一个基

因为 $A|_{W_i}$ 在 B_i 下的矩阵为 \mathbf{A}_i , 因此有: $A|_{W_i}(\alpha_{ij}) = \sum_{k=1}^{r_i} \mathbf{A}_i(k; j) \alpha_{ik}$

考察 A 在同样的向量上的作用:

$$A(\alpha_{ij}) = A|_{W_i}(\alpha_{ij}) = \sum_{k=1}^{r_i} \mathbf{A}_i(k; j) \alpha_{ik}$$

由基的线性无关性可知, 以上是 $A(\alpha_{ij})$ 的唯一表示。于是, A 在基 $B_1 \cup \dots \cup B_s$ 下的矩阵 \mathbf{A} 分块后的第 i 列, 当且仅当在 i 行为 \mathbf{A}_i , 其他均为 \mathbf{O}

$$\text{于是一定有 } \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_s \end{pmatrix}, \text{ 充分性得证。}$$

接下来证明必要性。

取 $W_i = \text{span}(\alpha_{i1}, \dots, \alpha_{ir_i})$

由直和的充要条件可知, 的确有 $V = W_1 \oplus \dots \oplus W_s$

且由分块矩阵的结构可见:

$$(A|_{W_i})(\alpha_{i1}, \dots, \alpha_{ir_i}) = (\alpha_{i1}, \dots, \alpha_{ir_i})\mathbf{A}$$

于是命题得证。 □

5. 我们本节提出的最后一个命题, 是指出根子空间也是不变子空间, 自然可以应用上面关于分块矩阵的命题

命题 1.3.6. 设 V 是 F 上的一个线性空间, $\dim V = n < \aleph_0$,

如果 $A \in \text{hom}(V, V)$ 的特征多项式在 F 内有以下分解:

$$\chi_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$$

那么 A 的根子空间 $\ker(A - \lambda_i I)^{r_i}$ 是 A 的不变子空间

证明: 这其实是显然的, 我们只需要注意到: $A(A - \lambda_i I) = A^2 - \lambda_i A = (A - \lambda_i I)A$

即 A 和 $A - \lambda_i I$ 是可交换的

那么, 按照数学归纳法, 容易证明 $A(A - \lambda_i I)^{r_i} = (A - \lambda_i I)^{r_i} A$

于是有:

$$\forall \alpha \in \ker(A - \lambda_i I)^{r_i}, (A - \lambda_i I)^{r_i}(\alpha) = \mathbf{0}$$

$$\Rightarrow A(A - \lambda_i I)^{r_i}(\alpha) = \mathbf{0} \Rightarrow (A - \lambda_i I)^{r_i}(A(\alpha)) = \mathbf{0}$$

$$\Rightarrow A(\alpha) \in \ker(A - \lambda_i I)^{r_i}$$

于是命题得证。 □

1.4 线性变换的最小多项式

前面我们指出了：只要特征多项式可分解为一次因式的乘积，那么线性空间就是根子空间的直和；进一步，我们只需求出线性变换在根子空间上的限制的矩阵，就可以对角地写出线性变换的矩阵

本节中我们依旧选取多项式作为一个出发点，试图找到这些分块的具体形式。

1.4.1 最小多项式的定义

定义 1.4.1: 最小多项式

设 V 是 F 上的一个线性空间， $A \in \text{hom}(V, V)$

如果 $f(x) \in F[x], f \neq 0$ 满足：

1. $f(A) = 0$
2. $\forall g \in F[x] \wedge g(A) = 0, \deg f \leq \deg g$
3. f 是首一多项式

那么我们称 f 是 A 的最小多项式。

1.4.2 最小多项式的性质

1.5 幂零变换的 Jordan 标准型

1.6 线性变换的 Jordan 标准型