矩阵分析 笔记

任云玮

目录

1 Eigenvalues, Eigenvectors, and Similarity

1.0 Introduction

1.1 The eigenvalue-eigenvector equation

p46 注意对于矩阵 A 和常数 k，有

$$(A+kI)x=\lambda' x \quad \Rightarrow \quad Ax=(\lambda'-k)x.$$

因此在计算特征值的时候可以先给矩阵加上 I 的某个倍数来得到一个特征值更方便计算的矩阵.

p46. Theorem 1.1.6 这一定理的第一部分可以用来证明 $\sigma(p(A))$ 非空并给出其中的某些值；而第二部分则限定了 $\sigma(p(A))$ 中的值的可选范围. 如果我们在已知 $p(A)$ 的特征值情况下讨论 A，则这两部分的效果是反过来的.

p47. The proof of Theorem 1.1.9 我们的目的是证明任意 $A\in M_n$ 有至少一个特征值 λ，且它所对应的特征向量可以被表示为 $g(A)y$ 的形式，其中 $g(t)\in\mathbb{P}_{n-1}$.

按照定理 1.1.6，我们可以找一个多项式 p，找 $p(A)$ 的特征值. 考虑之前的观察以及 p 有常数项这件事，我们只需要说明 $0\in\sigma(p(A))$)，即证 $p(A)$ 是奇异的. 对于一个 $p(A)$ 形式的矩阵，十分自然地引入了一组向量 $(A^n y,\dots,y)$. 对于任意的 $y\neq 0$，我们可以取恰当的 p 使得这组向量是线性相关，从而 $p(A)$ 是奇异的. 将上述讨论反过来叙述即证明了命题的第一部分，其中关于 p 需选取度数最小的那一个.

为得到所需要的 g，我们只需要考虑把 $p(A)y=0$ 进行因式分解，得

$$(A-\lambda I)(g(A)y)=0.$$

由于 p 是满足线性相关的条件的多项式中度数最小的，所以 $g(A)y$ 非奇异，从而它是一个特征向量.

1.2 The characteristic polynomial and algebraic multiplicity

p54. Notes on Definition 1.2.14 $S_k(\lambda_1,\dots,\lambda_n)$ 可以理解为从 $\lambda_1,\dots,\lambda_n$ 这 n 个数中取出 k 个，将它们相乘，然后把所有取法的结果相加.

p55. TODO

1.3 Similarity

p59. Exercise 1. 由于相似变换不改变矩阵的特征值，所以有 $S_k(S^{-1}AS)=S_k(A)$，因此成立 $E_k(S^{-1}AS)=S_k(S^{-1}AS)=S_k(A)=E_k(A)$.

p59. Theorem 1.3.7 这一定理表明，如果我们有 k 个线形无关的特征向量，那么我们就可以把一个矩阵的前 k 列相似地化成为一个对角线是对应特征值的对角阵；反之亦成立，而这就保证了只要一个矩阵可对角化，我们就可以通过恰当地摆放这些特征向量来得到 S，是的 $S^{-1}AS$ 的对角线恰为 $\lambda_1, \ldots, \lambda_n$ 的任意一个排列，其中不同的 λ_i 可能相同。

p60. Exercise 3. $\operatorname{rank}(A-\lambda I) > n-m$ 意味着 $\dim(\operatorname{nullspace}(A-\lambda I)) < m$，即和 λ 相关联的线形无关的特征向量不足 m 个. 所以 A 的线性无关特征向量总数不足 n 个.

p60. Exercise 4. 根据秩-零化度定理，$\operatorname{rank}(A-\lambda I) \le n-k$. 设 λ 的代数重数为 m，则根据 Theorem 1.2.18，有 $\operatorname{rank}(A-\lambda I) \ge n-m$. 所以有

$$n-m \le n-k \quad \Rightarrow \quad m \ge k.$$

p60. Lemma 1.3.8 不同特征值对应的特征向量组成的向量组线性无关。

p60. The proof of Lemma 1.3.8 设 (λ_i, u_i) 为对应的特征值-特征向量对，令 $x_1u_1 + \cdots + x_ku_k = 0$. 将 A 作用于上式两端 $0, 1, \ldots, k-1$ 次，得矩阵

$$\begin{bmatrix} x_1u_1 & \cdots x_ku_k \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \cdots & \lambda_k^{k-1} \end{bmatrix} = 0.$$

由于 λ_i 各不相同，所以其中第二个矩阵为 Vandermonde 矩阵，是非奇异的，所以在两边右乘它的逆后我们得到 $[x_1u_1, \cdots, x_ku_k] = 0$. 由于特征向量不是零向量，所以有 $x_i = 0$. 从而 $\{u_i\}$ 线性无关。

p62. Theorem 1.3.12 之所以可以让 A 有如此形式是有 Theorem 1.3.7 保证的。

p62. Definitions 1.3.16 称 $W \subset \mathbb{C}^n$ 为 F-invariant, 若将 $\mathcal{F} \subset M_n$ 左乘作用于 W，其值域被包含于 W，即在 $\mathcal{F}$ 中元素的作用下封闭.

疑问：A 和 B 可交换是否意味着它们对应的线性变换可交换?

p63. Exercise 1. Suppose the one-dimensional subspace $W \subset \mathbb{C}^n$ is A-invariant, then for any $x \in W$,

$$Ax \in W \quad \Rightarrow \quad Ax = kx, \quad \text{For some } k \in \mathbb{C}.$$

Hence, x is an eigenvector of A.

p63. Observation 1.3.18 有 k 维不变子空间 W 等价于可以相似地将化为 $\left[\begin{smallmatrix} B & C \\ 0 & D \end{smallmatrix}\right]$，其中 $B \in M_k$. 由于相似矩阵有相同的特征多项式，所以它同时表明 $p_A(t) = p_B(t)p_D(t)$.

同时在该不变子空间中可以找到一个 A 的特征向量. [1] 关于这一结论，首先我们考虑这样一件事情，我们是否可以找到这样一个不变子空间 W，使得 $A|_W : W \to W$ 是一个满射（从而是一个双射）。答案是可行的，首先考虑将 A 作用于 W，其值域 $\tilde{W}$ 是一个 W 的子空间，如果 $\dim \tilde{W} = \dim W$，则有 $\tilde{W} = W$. 若 $\dim \tilde{W} < \dim W$，则我们注意到 $\tilde{W}$ 仍是一个 A 的不变子空间，所以我们可以再做一遍同样的操作。由于 $\dim W$ 是一个有限的非负整数，所以迟早会有一个满足要求的 $\tilde{W}$. 接下来我们来考虑这一 $\tilde{W}$ 的维度的问题，考虑在 W 非零的情况下，它是否是一定是非零的。不幸的是这是不能保证的，显然 $W = \ker A$ 是一个不变子空间，且它可以不为零，但是按照这种方式找到的 $\tilde{W}$ 是 $\{0\}$.

但是我们仍然可以从这一结论出发得出结论。考虑非零不变子空间 W 和对应的 $\tilde{W}$。如果 $\tilde{W} = \{0\}$，则我们知道 A 至少把一个 W 中的向量 x 映到了 0 上，从而 x 是 A 的与 A 相关联的特征向量。而若 $\tilde{W} \neq \{0\}$，我们不妨设 $A = \tilde{W}$。接下来我们考虑 S_1 和 $B = [\beta_1, \ldots, \beta_k]$ 的含义。首先 S_1 是一个从 $\mathbb{C}^k$ 到 $W \subset \mathbb{C}^n$ 的线形双射。而 $AS_1 = S_1B$ 意味着 $As_i = S_1\beta_i$，即意味着对于任意 $x = \sum x_i s_i \in W$，有

$$Ax = \sum_{i=1}^{k} x_i(As_i) = S_1 \left(\sum_{i=1}^{k} x_i \beta_i \right) = S_1 B[x_1, \ldots, x_k]^T.$$

这说明我们可以把 A 作用于 $x \in W$ 的效果拆分成一次 B 作用于对应坐标组成的向量上的结果再加上一次用 S_1 把坐标向量转成对应的 W 中的向量。我们知道 B 一定有一个特征值 λ 以及一个相关的特征向量 $[x_1^*, \ldots, x_k^*]$，代入上式即得到

$$Ax^* = S_1 B[x_1^*, \ldots, x_k^*]^T = \lambda S_1 [x_1^*, \ldots, x_k^*]^T = \lambda x.$$

从而我们的到了一个特征向量。

这一讨论也说明了对于一个不变子空间，我们可以把 A 对于其中向量的作用效果用一个 M_k 中的矩阵来描述。

p63. The proof of Observation 1.3.18 首先假设 k 维子空间 $W \subset \mathbb{C}^n$ 是 A-invariant，我们所要做的即找非奇异的 S，使得 $S^{-1}AS$ 为左上角块属于 M_k. 设 $s_1, \ldots, s_k$ 是 W 的一组基，$S_1 = [s_1, \ldots, s_k]$ 由于 W A-invariant，所以对于任意 W 中的向量 x，Ax 是 $s_1, \ldots, s_k$ 的线形组合，分别取 $x = s_k$，我们有

$$As_i = \sum_{j=1}^{k} b_{ij} s_j = S_1 \beta_i, \quad 1 \leq i \leq k.$$

[1] 我意识到为了得出这一结论，下面的大部分讨论实际上是不需要，真正重要的只有之后第二节的后一半。

设 $B = [\beta_1, \ldots, \beta_k]$，则有 $AS_1 = S_1B$. 令 $S = [S_1, S_2]$，注意到 S_2 实际上并不是十分重要，所以我们只需要确保 S 非奇异即可，我们可以从 $s_1, \ldots, s_k$ 开始扩充出一组 $\mathbb{C}^n$ 的基，并用这组基的后半部分部分组成 S_2. 我们有

$$S^{-1}AS = [S^{-1}S_1BS^{-1}AS_2].$$

其中 $S^{-1}S_1 = [S^{-1}s_1, \ldots, S^{-1}s_k] = [e_1, \ldots, e_k]$.

反之，如果我们已知相似于 (1.3.17) 的形式，设变换矩阵为 $S = [S_1, S_2]$，可以证明 S_1 的列向量张成的空间为 A-invariant.

p63. Lemma 1.3.19 两两可交换的矩阵共有至少一个特征向量。

p64. The proof of Lemma 1.3.19 首先考虑交换族有什么性质。注意到有 $A(Bx) = B(Ax)$，若 x 是 A 的与 λ 相关联的特征向量，则有 $B(Ax) = \lambda(Bx)$，从而 Bx 也是 A 的一个特征向量。即对于任意一个 $A \in \mathcal{F}$ 和它的任意特征值 λ，设 $W_{A,\lambda}$ 是其对应的特征向量全体，则 $W_{A,\lambda}$ 是一个 $\mathcal{F}$-不变子空间。

同时我们分析 $W_{A,\lambda}$ 的性质。取定 $x_0 \in W_{A,\lambda}$，则对于 TODO

p64. Exercise 1. TODO: Why commuting implies $\dim W = 1$.

Theorem 1.3.21 通常我们会把某一个可对角化的矩阵 A 相似地化成 $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ 的形式，而此时的问题在于用这一个 S，是否能够将其他的 B 化成对角阵。注意命题的第一部分仅保证了使同时对角化的阵的存在性，没有说明是否某个给定的矩阵可以做到这件事情，而命题的第二部分保证了这个 S 可以对角化其他所有阵。

首先根据 Theorem 1.3.7 我们知道这个这个 S 由 A 的 n 个线性无关的特征向量组成，而由于对任意 $B \in \mathcal{B}$，$AB = BA$，所以 B 是一个和 A 共形的对角分块矩阵，我们只需要证明每一个矩阵块都是对角阵即可。而考虑到在 B 本身不是对角阵的情况，每个矩阵块的大小都要比 $n \times n$ 要小，所以自然的可以想到用归纳法来完成整个证明。

2 Cheat Sheet

1 定理 (rank-nullity) Let $A \in M_{m,n}(\mathbb{F})$ be given. $\operatorname{rank} A + \dim(\operatorname{nullspace} A) = n$.

2 引理 (full-rank factorization) Suppose $A \in M_{m,n}(\mathbb{F})$, then $\operatorname{rank} A = k$ iff $A = XY^T$ for some $X \in M_{m,k}(\mathbb{F})$ and $Y \in M_{n,k}(\mathbb{F})$ that each have independent columns.

3 引理 (rank-one perturbation) $\det(A + xy^*) = \det A + y^*(\operatorname{adj} A)x$.

4 引理 (特征多项式系数) 设 $p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$，有

$$a_k = \frac{1}{k!} p_A^{(k)}(0) = (-1)^{n-k} E_{n-k}(A).$$

5 引理

$$\begin{bmatrix} I_m & X \\ 0 & I_n \end{bmatrix}^{-1} = \begin{bmatrix} I_m & -X \\ 0 & I_n \end{bmatrix}.$$