
线性算符的行列式、迹和特征值，都是关于一个线性算符的标量，而且这些标量的取值不依赖基的选择。

线性算符的行列式

定义 0.1 ($n \times n$ 矩阵的行列式). 设 A 是数域 \mathbb{F} 上的 $n \times n$ 矩阵 (记为 $A \in \mathbb{F}^{n \times n}$), $a_i = (A_{i1}, \dots, A_{in})$, $i = 1, \dots, n$ 是 A 的第 i 行的有序数组, 如果函数 $D: \mathbb{F}^{n \times n} \rightarrow \mathbb{F}$ 满足:

1. D 是关于 (a_1, \dots, a_n) 的 n 重线性函数, 即

$$D(a_1, \dots, \lambda a_i + a'_i, \dots, a_n) = \lambda D(a_1, \dots, a_i, \dots, a_n) + D(a_1, \dots, a'_i, \dots, a_n), \forall \lambda \in \mathbb{F}$$

2. 当 A 的任意两行相等则 $D(A) = 0$
3. 设 A' 是 A 的任意两行调换后的矩阵, 则 $D(A') = -D(A)$
4. $D(I) = 1$, 其中 I 是 $n \times n$ 单位矩阵

则称 D 是矩阵 A 的行列式 (*determinant*), 记为 $\det A$ 。

上述定义中的条件 1 是本身是 n 重线性函数 (n -linear function) 的定义, 条件 1 和条件 2 一起则是交替 (*alternating*) n 重线性函数的定义。条件 3 是可由条件 1 和 2 独立证明得到的 (此略^{[1]§5.2}), 加到了定义中只是为了便于理解。我们可以说, 行列式就是一个关于 $n \times n$ 矩阵的交替 n 重线性标量值函数且附加满足条件 4。

上面的定义隐含默认了任意一个 $n \times n$ 矩阵的行列式是唯一存在的要求 (因为“函数”的定义本身要求这一条), 但这未被证明。本讲义暂不列出详细证明的证明过程, 只简述证明的思路。证明存在性往往相当于找到这一存在。回顾本科的线性代数课本, 我们发现那里的定义方式^{[2]§1.3} 定义 3.1 就帮我们找出了这一存在:

$$\det A = \sum_{\sigma} (\operatorname{sgn} \sigma) A(1, \sigma_1) \cdots A(n, \sigma_n)$$

其中 $A(i, j) \equiv A_{ij}$; $\sigma_i, i = 1, \dots, n$ 表示 n 阶排列 σ 的一种; $\operatorname{sgn} \sigma$ 是根据 σ 的奇偶性取值 1 或 -1^* 。作为存在性的证明, 提出这一公式, 然后验证其满足行列式定义中的条件, 就完成了。而唯一性的证明则需要如下引理: 任意 $\mathbb{F}^{n \times n}$ 上的交替 n 重线性函数 D 均满足 $D(A) = \det A D(I), \forall A \in \mathbb{F}^{n \times n}$, 其中 $\det A$ 在此只需存在即可。该引理的证明也可以从上述的行列式公式直接得出, 使得行列式定义中的条件 4 实际成为行列式唯一性的必要条件。

以下定理列出了 $n \times n$ 矩阵的行列式的一些性质, 在本讲义中也不作证明而直接承认其成立, 因为它们都已经大学一年级的线性代数课程中介绍过了。

*关于 n 阶排列的知识可参见^{[2]§1.2}。

定理 0.1.

- $\det(AB) = \det A \det B, \forall A, B \in \mathbb{F}^{n \times n}$
- $\det(A^T) = \det A, \forall A \in \mathbb{F}^{n \times n}$
- A 可逆 $\Leftrightarrow \det(A) \neq 0$
- $\det(A^{-1}) = (\det A)^{-1}, \forall A \in \mathbb{F}^{n \times n}$
- 如果 B 是 A 的第 i 行加上第 j 行的倍数得到的矩阵, 则 $\det B = \det A$
- 克拉默法则 (Cramer's rule)^{[2]§1.5}

定义在数域 \mathbb{F} 上的有限维向量空间 \mathcal{V} 上的线性算符 $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 在不同有序基下的坐标矩阵之间有变换公式。设 B, B' 是 \mathcal{V} 的两组有序基, S 是从 B 到 B' 的过渡矩阵, 则由定理??有 $(\mathbf{T}) = S(\mathbf{T})'S^{-1}$, 其中 $(\mathbf{T}), (\mathbf{T})'$ 分别是 \mathbf{T} 在 B, B' 下的坐标矩阵。由行列式的性质有 $\det(\mathbf{T}) = \det[S(\mathbf{T})'S^{-1}] = \det(\mathbf{T})'$, 因此一个线性算符在任意基下的坐标矩阵的行列式都相等。由此我们可以直接定义“线性算符的行列式”, 记为 $\det \mathbf{T}$, 为其在任一基下的坐标矩阵的行列式。正式定义如下。

定义 0.2 (线性算符的行列式). 设 \mathbf{T} 是 n 维向量空间 \mathcal{V} 上的线性算符, B 是 \mathcal{V} 的一组有序基, 则 \mathbf{T} 的行列式 $\det \mathbf{T} \equiv \det(\mathbf{T})$, 其中 (\mathbf{T}) 是 \mathbf{T} 在 B 下的坐标矩阵。

线性算符的迹

从线性算符的行列式的定义过程, 我们发现, 与本科线性代数课上直接定义成一个运算公式不同, 我们总是先定义一组代数规则, 再去证明这样的代数规则唯一对应一个具体的运算公式。线性算符的迹也可按类似的方式重新定义。不过, 既然已经通过行列式的例子来了解这种思想, 此处关于迹的定义仍采用简化的方式。值得注意的是, 迹是定义在内积空间上的线性算符上的。

定义 0.3 (线性算符的迹). 设 \mathbf{A} 是数域 \mathbb{F} 上的内积空间 \mathcal{V} 上的线性算符, 则 \mathbf{A} 的迹 (trace) 为 $\text{tr} \mathbf{A} \equiv \sum_k (\mathbf{A} \hat{\mathbf{e}}_k | \hat{\mathbf{e}}_k)$, 其中 $\{\hat{\mathbf{e}}_k\}$ 是 \mathcal{V} 的任意一组规范正交基。

易验, 上述定义的迹的值不依赖基的选择而改变, 从而任一线性算符唯一对应一个迹。我们还能进一步获得如下性质。

定理 0.2. 设 \mathbf{A}, \mathbf{B} 是数域 \mathbb{F} 上的内积空间 \mathcal{V} 上的任意两个线性算符, 则

1. $\text{tr}(\alpha \mathbf{A} + \mathbf{B}) = \alpha \text{tr} \mathbf{A} + \text{tr} \mathbf{B}$
2. $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$

我们在此还可通过迹来定义 $\mathcal{L}(\mathcal{V})$ 空间上的一种内积。

定义 0.4 (线性算符的标准内积). 设 \mathbf{A}, \mathbf{B} 是数域 \mathbb{F} 上的内积空间 \mathcal{V} 上的任意两个线性算符, 令 $(\mathbf{A}|\mathbf{B}) \equiv \text{tr}(\mathbf{A}^*\mathbf{B})$, 可验证, 该定义满足内积规定, 称为线性算符的标准内积 (*standard inner product*), 经常记作 $\mathbf{A} : \mathbf{B}$ 。

读者可尝试自行推导两个线性算符的标准内积在给定基下的坐标运算公式。

线性算符的特征值

我们已经知道, 向量空间上的线性算符在选定某组基下, 它的性质就全由其坐标矩阵决定。线性算符的特征值分析, 其实是对一个线性算符的“内部结构”像对钟表那般进行解剖, 搞清楚它的运作方式。

尽管线性算符的特征值分析能给出很丰富的结论, 但目前我们也许可以先从这样一种动机去引入线性算符的特征值定义。比如我们关心一个线性算符是否可逆, 从线性变换的维数定理的角度, 需要知道线性算符的秩, 即其在任一选定基下的坐标矩阵的秩。从定理0.1的角度, 需要知道线性算符的行列式, 即其在任一选定基下的坐标矩阵的行列式。如果一个数域 \mathbb{F} 上的 n 维向量空间 \mathcal{V} 上的线性算符 \mathbf{T} 在基 $\{\mathbf{a}_i\}$ 下的坐标矩阵是对角矩阵, 即

$$(\mathbf{T}) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \lambda_i \in \mathbb{F}, i = 1, \dots, n$$

就称线性算符 \mathbf{T} 是可对角化的 (*diagonalizable*)。此时有

$$\mathbf{T}\mathbf{a}_i = \lambda_i \mathbf{a}_i, \quad i = 1, \dots, n$$

\mathbf{T} 的值域就是由 $\lambda_i \neq 0$ 对应的 \mathbf{a}_i 线性生成的, $\lambda_i \neq 0$ 的个数就是 $\text{rank}\mathbf{T}$, 要想 \mathbf{T} 可逆, 必须使得 $\text{rank}\mathbf{T} = n$ (满秩), 或由 $\det\mathbf{T} = \prod_{i=1}^n \lambda_i \neq 0$ 要求 (\mathbf{T}) 的对角元素全不为零。总之矩阵 \mathbf{T} 的很多性质就能确定了。

为了这一目标, 我们先讨论一般的形如 $\mathbf{T}\mathbf{a} = \lambda\mathbf{a}$ 的情况。

定义 0.5 (特征值、特征向量、特征空间). 设 \mathcal{V} 是数域 \mathbb{F} 上的有限维向量空间, $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 是一个线性算符。若存在 $\mathbf{a} \neq \mathbf{0}, \mathbf{a} \in \mathcal{V}, \lambda \in \mathbb{F}$ 满足 $\mathbf{T}\mathbf{a} = \lambda\mathbf{a}$, 则称 λ 是 \mathbf{T} 的一个特征值 (*characteristic value*), \mathbf{a} 是 \mathbf{T} 对应于特征值 λ 的一个特征向量 (*characteristic vector*)。 \mathbf{T} 关于同一特征值 λ 的所有不同的特征向量的集合称为 \mathbf{T} 的特征空间 (*characteristic space*)。

由特征值和特征向量的定义可知, 复数域上的 n 维向量空间上的线性算符可以有 n 个特征值。对应于同一个特征值, 可能有多个特征向量。能够证明, 关于同一特征值的特征空间是向量空间的一个子空间, 作为定理如下。

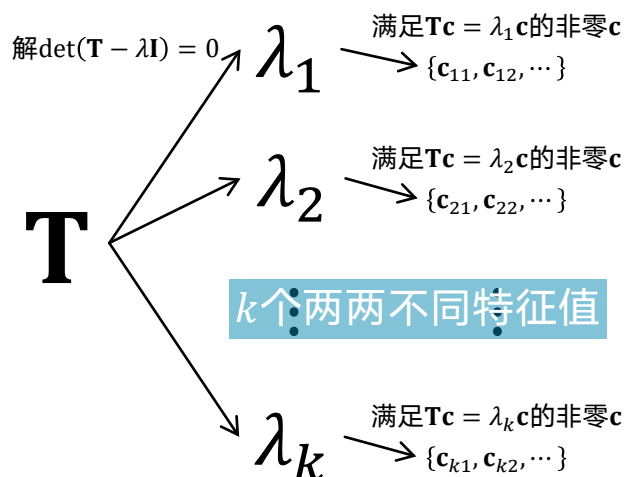


图 0.1: 复数域上的 n 维向量空间上的线性算符 \mathbf{T} 有 n 个特征值, 其中有 $k \leq n$ 个两两不同特征值。对应每个不同特征值, 都可能有一个或以上的特征向量。对应于每个不同特征值的所有特征向量的集合称为关于该特征值的特征空间。特征空间具有向量运算的封闭性 (需证明), 故它们都是所在向量空间的子空间。

定理 0.3. 设 \mathcal{V} 是数域 \mathbb{F} 上的有限维向量空间, $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 是一个线性算符, 则 \mathbf{T} 的特征空间是 \mathcal{V} 的子空间。

证明. 设 $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ 是关于 \mathbf{T} 的特征值 λ 的其中两个特征向量, 则对任意 $\alpha \in \mathbb{F}$ 有 $\mathbf{T}(\alpha\mathbf{a} + \mathbf{b}) = \alpha\mathbf{T}\mathbf{a} + \mathbf{T}\mathbf{b} = \alpha\lambda\mathbf{a} + \lambda\mathbf{b} = \lambda(\alpha\mathbf{a} + \mathbf{b})$ 。 \square

定义 0.5 只告诉我们, 如果有满足定义的东西, 则把它称为什么; 它没有告诉我们这样的东西存不存在, 如何去求它们。为使特征值的概念更实用, 我们进一步考虑一个重要的线性算符表达式。设 \mathcal{V} 是数域 \mathbb{F} 上的有限维向量空间, $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 是一个线性算符, λ 是 \mathbf{T} 的一个特征值, \mathbf{a} 是 \mathbf{T} 关于 λ 的任一特征向量, 则总有

$$(\mathbf{T} - \lambda \mathbf{I}) \mathbf{a} = \mathbf{T}\mathbf{a} - \lambda \mathbf{I}\mathbf{a} = \mathbf{0}$$

即 \mathbf{T} 关于 λ 的特征空间就是线性算符 $\mathbf{T} - \lambda \mathbf{I}$ 的零空间。根据线性变换的维数定理, \mathbf{T} 的特征空间的维数就是 $\mathbf{T} - \lambda \mathbf{I}$ 的零化度, 进而有如下定理。

定理 0.4. 设 \mathcal{V} 是数域 \mathbb{F} 上的有限维向量空间, $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 是一个线性算符, 则以下命题相互等价:

1. λ 是 \mathbf{T} 的特征值
2. $\mathbf{T} - \lambda \mathbf{I}$ 是奇异的
3. $\det(\mathbf{T} - \lambda \mathbf{I}) = 0$

证明. 我们在本科的线性代数课中已经了解, 矩阵行列式为零, 就是该矩阵相对就的一个线性方程组无非全零解, 亦即该矩阵是不可逆的。这些结论的详细证明就是对 $2 \Leftrightarrow 3$ 的证明。此略。 $1 \Leftrightarrow 3$ 可利用行列式的性质简单证得, 此略。 \square

这一定理的第 3 个命题实际提供了一个求算给定线性算符的特征值的方法。因为

$$\det(\mathbf{T} - \lambda \mathbf{I})$$

必是一个 $\dim \mathcal{V}$ 阶首一多项式^{*}。因此可将其定义为 \mathbf{T} 的特征多项式 (*characteristic polynomial*), 从而 \mathbf{T} 的任一特征值均为其特征多项式的一个根。在复数域上, 由代数基本定理, n 阶首一多项式有且必有 n 个复根 (可能有重根)[†]。因此我们可以说复数域上任一 n 维线性空间上的线性算符必有 n 个特征值; 解其特征多项式就可以得到它们全部。

本节开头提出的, 希望线性变换在某组基下的坐标矩阵是对角矩阵, 这相当于要求线性变换的特征向量直接就是一组基。我们可以从这一要求出发, 找出与其等价的一系列命题, 如下定理所示。

定理 0.5. 设 \mathcal{V} 是数域 \mathbb{F} 上的 n 维向量空间, $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 是一个线性算符, $\{\theta_1, \dots, \theta_k\}$ 是 \mathbf{T} 的两两不同的特征值, \mathcal{W}_i 是线性算符 $\mathbf{T} - \theta_i \mathbf{I}$ 的零空间, 则以下命题相互等价:

1. \mathbf{T} 是可对角化的;
2. \mathbf{T} 的特征多项式是

$$(\theta - \theta_1)^{d_1} \cdots (\theta - \theta_k)^{d_k}$$

且

$$\dim \mathcal{W}_i = d_i, \quad i = 1, \dots, k$$

3. $\dim \mathcal{V} = \sum_{i=1}^k \dim \mathcal{W}_i$

这一定理第 1、2 条之间的等价性可见 [2]§5.2 定理 2.1, 第 2、3 条之间的等价性要参考其他线性代数教材 [1]§6.2。

若考察第 1、2 条之间等价的证明过程, 我们将看到, 复数域上的向量空间上的任一线性算符必有 n 个特征值。但只有当这一线性算符可对角化时, 才能从这些特征值对应的特征空间中找到 n 个线性无关的向量。如果线性算符 \mathbf{T} 可对角化, 它的对角坐标矩阵的对角元素就是 \mathbf{T} 的 n 个特征值。如果 \mathbf{T} 不可对角化, 那么 \mathbf{T} 仍可以有 n 个特征值, 但就无法找到任何

^{*}这一命题的结论需要依靠多项式代数的知识来证明。后文还要用到代数基本定理, 这个定理也是通过多项式代数证明的重要结论, 此略 [1]。

[†]这样的数域又叫代数闭域。例如实数域就不是代数闭域。所以线性算符所有关于特征值的性质, 在实数域上都需要小心是否会丧失。

一组基，使 \mathbf{T} 在其下的坐标矩阵是一个对角矩阵，更别说想要是以 \mathbf{T} 的特征值作为对角元素的对角矩阵了。以下是一个例子：一个可逆矩阵，在关于其某重根数大于 1 的特征值的特征空间中，找不到足够的个数（等于重根数）的线性无关向量组来作为这个特征空间的一组基，从而违反定理 0.5 而不可对角化。

例 0.1. 考虑矩阵

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

首先看到， A 是可逆的，因为它的行列式

$$\begin{aligned} \det A &= 2 \times (2 \times 2 - 1 \times 0) - 1 \times (0 \times 2 - 1 \times 0) + 0 \times (0 \times 1 - 2 \times 0) \\ &= 2 \times 4 \\ &= 8 \neq 0 \end{aligned}$$

我们再考察 A 的特征值，考虑其特征多项式

$$p(\lambda) = \det(A - \lambda I) = \det \begin{pmatrix} 2 - \lambda & 1 & 0 \\ 0 & 2 - \lambda & 1 \\ 0 & 0 & 2 - \lambda \end{pmatrix} = (2 - \lambda)^3$$

故 A 只有一个不同特征值 $\lambda = 2$ ，重根数是 3。关于 $\lambda = 2$ 的特征向量 $v = (v_1, v_2, v_3)^T$ 应该满足 $v \neq (0, 0, 0)^T$ 且

$$\begin{aligned} Av &= 2v \\ \Leftrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ \Leftrightarrow v_1 \neq 0, \quad v_2 = v_3 = 0 \end{aligned}$$

可见，关于 $\lambda = 2$ 的特征空间维数是 1。而 $\lambda = 2$ 是 3×3 矩阵 A 的唯一一个不同特征值了，在这个特征值的特征空间中都找不齐 3 个线性无关的向量。定理 0.5 中的第 3 条不满足，故 A 是不可对角化的。

对于不可对角化的线性算符，我们仍可通过 Jordan 标准形 (Jordan canonical form) 来简化讨论。Jordan 标准形的知识需要阅读稍深的线性代数教材*。这方面知识本讲义不作介绍了。

*例如：李尚志 (2006), 线性代数 (数学专业用), 高等教育出版社

若考察第 2、3 条之间等价的证明，我们将看到，在复数域上，无论一个线性算符 \mathbf{T} 是否可对角化，它的 n 个特征值中都有可能重复的（即 \mathbf{T} 的特征多项式有重根）。设 $\theta_1, \dots, \theta_k$ 为 \mathbf{T} 的两两不同特征值，第 j 个特征值重复 d_j 次。如果 \mathbf{T} 可对角化，则 \mathbf{T} 的对角化坐标矩阵可以写成

$$(\mathbf{T}) = \begin{pmatrix} \theta_1 I_{d_1} & & \\ & \ddots & \\ & & \theta_k I_{d_k} \end{pmatrix}$$

其中 I_j 表示 $d_j \times d_j$ 单位矩阵。证明过程明确了这一结论：每个两两不同特征值的重复次数 d_j 就是它对应的特征空间的维数，进而有 $d_1 + \dots + d_k = n$ 。此外，对应于不同特征值的特征向量线性无关。也就是说从若干个两两不同的特征值所对应的特征空间中分别各取一个向量所组成的向量组必为一个线性无关向量组^{[2]§5.1 性质 1.4}，但是反过来我们不能说任意两个线性无关的特征向量必对应于两个不同特征值。

以下定理^{[2]§5.1 性质 1.1、1.2}给出了特征多项式最重要的三项的系数：

定理 0.6. 设 $\{\lambda_i\}$ 是线性算符 $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ 的 n 个特征值，则 $\text{tr} \mathbf{T} = \sum_{i=1}^n \lambda_i$ ， $\det \mathbf{T} = \prod_{i=1}^n \lambda_i$ ， $\det(\mathbf{T} - \lambda \mathbf{I}) = \lambda^n - \text{tr} \mathbf{T} \lambda^{n-1} + \dots + (-1)^n \det \mathbf{T}$

证明. 本科线性代数课本^[2]第五章性质 1.2 已证明了本定理，但是比较简短。这里不写完整的证明过程，只作一些补充性说明。

矩阵的迹和矩阵的行列式，都有独立的定义。它们之所以能与矩阵的特征值建立联系，是需要说明的。矩阵的迹定义为矩阵对角元素之和。由矩阵的性质可知，矩阵的迹在相似变换下不变。具体地，任何矩阵都能在有限次初等行变换后化成上三角矩阵。如果矩阵可对角化，则可化为对角矩阵。无论是上三角矩阵还是对角矩阵，对角元素都是原矩阵的特征值。因此矩阵的迹就是矩阵的特征值之和。

至于矩阵的行列式为何是特征值之积，也是通过特征值在由原矩阵变换得到的上三角或对角矩阵的对角元素中出现来证明的。但是，这一证明必须知道，所谓的“有限次初等行变换”得到上三角矩阵，更正式的说法是通过舒尔分解（Schur decomposition）。给定某矩阵 A ，总存在一个幺正矩阵 Q ，使得 $Q^T A Q$ 是一个上三角矩阵。这个幺正矩阵，是通过针对 A 的特征值逐个选定一组规范正交基来变换坐标得到的。这个过程中，矩阵的行列式是不变的，因此矩阵的行列式就是矩阵的特征值之积。具体可以参考舒尔分解的其他资料。

以上两段想要告诉我们，矩阵的迹和行列式与特征值的关系已经建立了。

现在我们考虑矩阵的特征多项式。在本科线性代数课本^[2]第五章性质 1.2 的证明中已经解释过，特征式项总是首一多项式。根据韦达公式（Viète formulas）??，首一多项式的系数与它的根之间有确定的关系。这一关系就是定理0.6的内容。□

特别地，当 $n = 3$ 时， $\det(\mathbf{T} - \lambda \mathbf{I}) = -\lambda^3 + \text{tr} \mathbf{T} \lambda^2 - \frac{1}{2}(\text{tr}^2 \mathbf{T} - \text{tr} \mathbf{T}^2) \lambda + \det \mathbf{T}$ 。我们令 $I_{\mathbf{T}} = \text{tr} \mathbf{T}$, $II_{\mathbf{T}} = \frac{1}{2}(\text{tr}^2 \mathbf{T} - \text{tr} \mathbf{T}^2)$, $III_{\mathbf{T}} = \det \mathbf{T}$ ，称为 \mathbf{T} 的第一、第二和第三主不变量 (*principal invariants*)。易见这些主不变量都是标量，且它的值不依赖基的选择，与 \mathbf{T} 唯一对应。我们还常用 $J_1 = I_{\mathbf{T}}$, $J_2 = I_{\mathbf{T}}^2 - 2II_{\mathbf{T}}$, $J_3 = I_{\mathbf{T}}^3 - 3I_{\mathbf{T}}II_{\mathbf{T}} + 3III_{\mathbf{T}}$ ，中文也常译为主不变量 (*main invariants*)。

定理0.6和0.1告诉我们验证线性算符是否可逆的一个方法：可逆线性算符所有特征值不为零。