第3章 线性代数

本章将介绍线性代数的基本内容。从矢量和矩阵的基本知识入手，将学习矩阵运算。接下来，学习矢量的线性组合，这是矩阵秩和矩阵独立等重要概念的基础。接着使用这些概念来对线性方程系统进行求解。然后，将学习线性代数在对线性变换过程进行建模时发挥的重要作用。本章还将学习特征值和特征向量，从而对其形成一个直观的理解。在计算机网络系统建模时经常会用到随机矩阵，因此，本章最后给出一个关于随机矩阵的描述。

3.1 矢量和矩阵

考虑一个两轮的实验，该实验用于为研究人员收集某条互联网链路的数据包。假设第一次记录到 312 个 TCP 数据包和 39 个 UDP 数据包，第二次记录到 432 个 TCP 数据包和 21 个 UDP 数据包。我们可以用有序二元组来表示实验的结果：[312， 39]和[432， 21]。这里，二元组中的元素所处的位置隐含地分别与“TCP 数据包数量”和“UDP 数据包数量”关联对应。我们把这种集合元素的有序表示方式称为**矢量**（vector）。

如果一个矢量有 n 个元素，我们称之为 n **维**。一个由 n 个实数值组成的 n 维矢量与 n 维度实数空间中的点存在一一对应的关系。回到刚才的例子中，矢量[432， 21]对应于二维实数空间中的一个坐标值为(432， 21)的点，该实数空间中 X 轴代表“TCP 数据包数量”，Y 轴代表“UDP 数据包数量”。如果我们要对该元素组添加另一个测量值，比如添加“ICMP 数据包数量”，我们可以用一个类似[432， 21， 12]的矢量来表示记录到的数据包数，这对应于一个三维实数空间中的一个点。

矢量可以用两种方式表示：即行矢量，如[312， 12， 88]，以及列矢量，如$\begin{bmatrix}312\\12\\88\end{bmatrix}$。我们将[0 0 0…0]定义为 n 维的**零矢量**，用 $\mathbf{0}$ 表示。在本书中，矢量用小写黑斜体字母表示，矢量元素用小写斜体字母表示。

回到我们的例子中，我们可以用矢量同时表示两次实验记录的数据包数量，例如：

$$\begin{bmatrix}312 & 39\\432 & 21\end{bmatrix}$$

我们将这种表达方式称为**矩阵**（matrix）。在本书中，矩阵用大写黑斜体字母表示，矩阵元素用小写斜体字母表示。

矢量的元素之间没有关联性。与矢量不同，一个矩阵的同一列的元素间往往是相互关联的。在我们的例子中，第一列的元素都表示 TCP 数据包数量。通常，一个 m 行 n 列的阵列称为一个 $m\times n$ 矩阵。通常，我们用符号 a_{ij} 来表示矩阵 $\boldsymbol{A}$ 中 i 行 j 列的元素。在前面的

例子中，a_{12}是 39，而 a_{21}是 432。

虽然矢量和矩阵的表示形式适用于任意的元素类型，如字符串等。在我们的讨论中，我们假定一个矢量或者矩阵的元素是数学**域**(field)的成员。出于完整性考虑，域的正式定义阐述如下；需注意，这个定义从本质上对我们直觉理解的实数特征进行了形式化：域 F 是一个有限或者无限**集合**，且集合中元素的加**运算**(用＋表示)和乘运算(用 $*$ 表示)满足以下六条公理。

(1) 加运算和乘运算的**闭合性**(Closure)：对于域 F 中的 a 和 b 来说，如果 $a+b=c$ 且 $a*b=d$，则 c 和 d 也包含在域 F 中。

(2) 加运算和乘运算的**交换性**(Commutativity)：对于域 F 中的 a 和 b 来说，$a+b=b+a$ 且 $a*b=b*a$。

(3) 加运算和乘运算的**结合性**(Associativity)：对于域 F 中的 a，b 和 c 来说，$(a+b)+c=a+(b+c)$。

(4) 集合中存在独特的加法恒等和乘法**恒等**(identity)元素：域 F 中存在独特的元素 0 和 1，对于 F 中所有的元素 a 有：$a+0=a$ 且 $a*1=a$。

(5) 存在加运算和乘运算的**逆运算**(inverse)：对于域 F 中的任一元素 a 来说，域 F 中总是同时存在一个元素 b，$a+b=0$。对于域 F 中的任一非 0 元素 a 来说，域 F 中总是同时存在一个元素 c，$a*c=1$。

(6) **乘法分配率**(distributivity)：对于所有的 a、b 和 c 来说，存在以下等式：$a*(b+c)=(a*b)+(a*c)$。

3.2 矢量和矩阵的代数运算

本节介绍一些基本的矢量运算和矩阵运算。

3.2.1 加法

同维度的两个**矢量之和**是一个新矢量，它的每个元素值是每个矢量对应元素的和。不同维度的矢量的和没有定义。

行数列数相同的两个**矩阵之和**是一个新矩阵，它的每个元素值是每个矩阵对应元素的和。行数或列数不同的矩阵的和没有定义。

由于矢量和矩阵的元素都属于域，且矢量和矩阵的加运算是逐个元素进行的，因此矢量和矩阵的加法运算继承了域的闭合性、交换性、结合性等特征，且同样存在加法的逆元以及加法的恒等元。

3.2.2 转置

行矢量 $\boldsymbol{x}$ 的转置矢量是一个列矢量，用 $\boldsymbol{x}^{\mathrm{T}}$ 表示。$\boldsymbol{x}^{\mathrm{T}}$ 的 j 行元素是 x 的 j 列元素。一个 $m\times n$ 的矩阵 $\boldsymbol{A}$ 的转置矩阵是一个 $n\times m$ 的矩阵，用 $\boldsymbol{A}^{\mathrm{T}}$ 表示。$\boldsymbol{A}^{\mathrm{T}}$ 的[j,i]元素就是 $\boldsymbol{A}$ 的[i,j]元素，即 a_{ij}。

例 3.1：转置

矢量[1，3，5，1]的转置是$\begin{bmatrix}1\\3\\5\\1\end{bmatrix}$。矩阵$\begin{bmatrix}23 & 2 & 9\\98 & 7 & 89\\34 & 9 & 1\end{bmatrix}$的转置是$\begin{bmatrix}23 & 98 & 34\\2 & 7 & 9\\9 & 89 & 1\end{bmatrix}$。

3.2.3　乘法

矢量 $\boldsymbol{x}$ 与一个实数，或者说一个**标量** s 的乘积，是一个由 $\boldsymbol{x}$ 的每个元素与该实数的乘积所组成的矢量。即：

$$s[x_1, x_2, \ldots, x_n] = [Sx_1, x_2, \ldots, sx_n] \tag{3.1}$$

类似地，一个矩阵 $\boldsymbol{A}$ 与一个实数（标量）s 的乘积，是一个由矩阵 $\boldsymbol{A}$ 的每个元素与该实数的乘积所组成的矩阵。即：

$$s\begin{bmatrix}a_{11} & \cdots & a_{1n}\\ \cdots & \cdots & \cdots\\ a_{m1} & \cdots & a_{mn}\end{bmatrix} = \begin{bmatrix}sa_{11} & \cdots & sa_{1n}\\ \cdots & \cdots & \cdots\\ sa_{m1} & \cdots & sa_{mn}\end{bmatrix} \tag{3.2}$$

两个矢量的乘积可以用**点**（dot）**乘积**或者**叉**（cross）**乘积**来定义。矢量 $\boldsymbol{x}$ 有元素 x_i，矢量 $\boldsymbol{y}$ 有元素 y_i，则矢量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的点乘积是一个标量 s，s 的值等于对应元素逐个相乘的和。即：

$$s = x,\quad y = \sum_{i=1}^{n} x_i y_i \tag{3.3}$$

如果两个矢量的维度不同，则无法定义它们的点乘积。两个矢量的叉乘积与计算机网络不相关，因此我们不在这里做深入的讨论。

与两个矢量的点乘积不同，两个矩阵的乘积不是一个标量，而仍是一个矩阵。该矩阵[i，j]元素值等于第一个矩阵的 i 行矢量与第二个矩阵的 j 列矢量的点乘积。即，如果 $\boldsymbol{C}=\boldsymbol{AB}$，则

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{3.4}$$

注意矩阵 $\boldsymbol{A}$ 的列数（即矩阵 $\boldsymbol{A}$ 每行矢量的维度）必须等于矩阵 $\boldsymbol{B}$ 的行数（即矩阵 $\boldsymbol{B}$ 每列矢量的维度）。因此，一个的 $m\times n$ 矩阵乘以一个 $n\times o$ 的矩阵，得到一个 $m\times o$ 的矩阵。因此，一个 n 维行矢量——即一个 $1\times n$ 矩阵——乘以一个 $n\times n$ 矩阵得到一个 n 维行矢量。

例 3.2：矩阵乘法

$\begin{bmatrix}23 & 2 & 9\\98 & 7 & 89\\34 & 9 & 1\end{bmatrix}$与$\begin{bmatrix}2 & 5 & -2\\4 & 9 & 8\\3 & 0 & 1\end{bmatrix}$的乘积是$\begin{bmatrix}81 & 133 & -21\\491 & 553 & -51\\107 & 251 & 5\end{bmatrix}$。例如，求 c_{11} 的方法是：

$$c_{11} = 23\times 2 + 2\times 4 + 9\times 3 = 46 + 8 + 27 = 81$$

矩阵乘法具有结合性，即$(\boldsymbol{AB})\boldsymbol{C}=\boldsymbol{A}(\boldsymbol{BC})$，但不具备交换性。即，通常来说：

$$\boldsymbol{AB} \neq \boldsymbol{BA} \tag{3.5}$$

以上结论可以通过对矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 的简单观察得出：尽管 $\boldsymbol{AB}$ 可能存在，如果矩阵 $\boldsymbol{B}$ 的列

数与 $\boldsymbol{A}$ 的行数不相等，就不能给出 $\boldsymbol{BA}$ 的定义。此外，如果 $\boldsymbol{AB}=0$，并不意味着矩阵 $\boldsymbol{A}$ 或 $\boldsymbol{B}$ 是空矩阵。这与标量不同。当 $ab=0$ 时，意味着 a 或者 b 其中之一必为 0。由此推论，如果 $\boldsymbol{AB}=\boldsymbol{AC}$，但 $\boldsymbol{B}$ 不一定与 $\boldsymbol{C}$ 相同。

3.2.4 方阵

一个具有相同行数和列数的矩阵称为**方阵**。一个只有在主对角线存在非零元素的方阵称为**对角矩阵**。

一个 $n\times n$ 对角方阵 $\boldsymbol{I}$ 的主对角线元素均为 1，其他元素都为 0 具备以下特性：$\boldsymbol{I}$ 与任意 $n\times n$ 的矩阵 $\boldsymbol{A}$ 相乘仍然得到 $\boldsymbol{A}$ 本身。即，$\boldsymbol{AI}=\boldsymbol{IA}=\boldsymbol{A}$。因此，$\boldsymbol{I}$ 也称为**单位**（identity）**矩阵**。

3.2.5 矩阵幂运算

定义一个方阵 $\boldsymbol{A}$，它与自己的乘积表示为 $\boldsymbol{A}^2$。如果 $\boldsymbol{A}$ 是 $n\times n$ 的，则 $\boldsymbol{A}^2$ 也是 $n\times n$ 的。以此类推，可以求得方阵 $\boldsymbol{A}$ 的更高阶幂，且也是 $n\times n$ 的。

例 3.3：矩阵幂运算

定义矩阵 $\boldsymbol{A}$ 为 $\begin{bmatrix}5&0&0\\0&7&0\\0&0&2\end{bmatrix}$，则 $\boldsymbol{A}^2=\begin{bmatrix}25&0&0\\0&49&0\\0&0&4\end{bmatrix}$，$\boldsymbol{A}^2=\begin{bmatrix}125&0&0\\0&343&0\\0&0&8\end{bmatrix}$。在这个例子中，很容易求出矩阵 $\boldsymbol{A}$ 的更高次幂（大家可以思考一下为什么?）。通常而言，如果是 $\boldsymbol{A}$ 一个对角矩阵，$\boldsymbol{A}$ 的 k 阶幂的 (i,i) 元素就是 a_{ii}^k。

3.2.6 矩阵指数

定义一个方阵 $\boldsymbol{A}$，可以求得它的更高次幂，且也是 $n\times n$ 的。在此基础上，矩阵指数 $e^{\boldsymbol{A}}$ 定义如下：

$$e^{\boldsymbol{A}}=1+\boldsymbol{A}+\frac{\boldsymbol{A}^2}{2!}+\frac{\boldsymbol{A}^3}{3!}+\cdots \tag{3.6}$$

3.3 线性组合、独立性、基和维度

本节介绍一些重要的基础概念，这些概念将在后续的章节中被使用到。

3.3.1 线性组合

定义一个包含 k 个实数变量的集合 $x_1,x_2,\ldots,x_k$，且假设已知一个 k 个实数权重集合 $w_1,w_2,\ldots,w_k$。我们定义 $s=w_1x_1+w_2x_2+\cdots+w_kx_k$。$s$ 变量是对变量 $\boldsymbol{x}$ 的加权线性"混合"。因此，我们将 s 定义为变量的**线性组合**。

我们可以将线性组合的概念推广到矢量中。定义每个 $\boldsymbol{x}_i$ 是一个矢量，因此，它们的线性组合 $\boldsymbol{s}$ 也是一个矢量。当然，每个矢量必须有相同数量的元素。请注意，$\boldsymbol{s}$ 的每个分量都

是一个由被组合矢量的相对应元素构成的线性组合。

例 3.4：标量的线性组合

定义标量为 2,4,1,5,权重为 0.1,0.4,0.25,0.25。求其线性组合 s。

解：线性组合 $s=0.1\times2+0.4\times4+0.25\times1+0.25\times5=0.2+1.6+0.25+1.25=3.3$。

例 3.5：矢量的线性组合

定义矢量为 $[2\quad4\quad1\quad5]$,$[3\quad5\quad1\quad2]$,$[5\quad6\quad2\quad1]$,$[9\quad0\quad1\quad3]$,权重为 0.1,0.4,0.25,0.25。求其线性组合 $\boldsymbol{s}$。

解：线性组合 $\boldsymbol{s}=0.1\times[2\quad4\quad1\quad5]+0.4\times[3\quad5\quad1\quad2]+0.25\times[5\quad6\quad2\quad1]+0.25\times[9\quad0\quad1\quad3]$。$\boldsymbol{s}$ 的第一个元素是 $0.1\times2+0.4\times3+0.25\times5+0.25\times9=0.2+1.2+1.25+2.25=4.49$。相应地,其他元素分别是 3.9,1.25 和 2.3。因此 $\boldsymbol{s}=[4.9,3.9,1.25,2.3]$。

3.3.2 线性无关

定义一个由 k 个矢量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_k$ 组成的集合,假设其中一个矢量——如 $\boldsymbol{x}_i$——是其他矢量的线性组合,则 $\boldsymbol{x}_i$ 的值由其他矢量值所决定：一旦设定其余矢量的值,则 $\boldsymbol{x}_i$ 的值不可能是任意值,而是由此被确定的某个值。这意味着在给矢量赋值的时候我们在某种程度上失去了一些自由度。

特别是,当我们适当地选择$(k-1)$个权重值,使得 $\boldsymbol{x}_i$ 是剩余$(k-1)$个矢量的线性组合时,我们可以将 $\boldsymbol{x}_i$ 写为：

$$\boldsymbol{x}_i=w_1\boldsymbol{x}_1+w_2\boldsymbol{x}_2+\cdots+w_{i-1}\boldsymbol{x}_{i-1}+w_{i+1}\boldsymbol{x}_{i+1}+\cdots+w_k\boldsymbol{x}_k \tag{3.7}$$

或者,可以做以下的改写：

$$w_1\boldsymbol{x}_1+w_2\boldsymbol{x}_2+\cdots+w_{i-1}\boldsymbol{x}_{i-1}-\boldsymbol{x}_i+w_{i+1}\boldsymbol{x}_{i+1}+\cdots+w_k\boldsymbol{x}_k=0 \tag{3.8}$$

由此推导出一个矢量组线性无关的定义：一个矢量组,当且仅当权重 $\boldsymbol{w}=0$ 时,才能满足式(3.8),则我们称这个矢量组是线性无关的。

注意如果一个矢量集合不是线性无关的,则其中的任何一个矢量都能够用其他矢量的组合来表达。

例 3.6：线性无关

有三个矢量：

$$\begin{aligned}x_1&=[3\quad0\quad2]\\x_2&=[-3\quad21\quad12]\\x_3&=[21\quad-21\quad0]\end{aligned}$$

由于 $6\boldsymbol{x}_1-\boldsymbol{x}_2-\boldsymbol{x}_3=0$,则它们不是线性无关的。由于 $\boldsymbol{x}_1$ 的第三个元素不能从 $\boldsymbol{x}_3$ 的第三个元素生成,因此 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_3$ 是线性无关的。

在 3.4.6 节中,我们会看到,如果一个矢量集合是线性无关的,则由这些矢量形成的矩阵是**非奇异**(nonsingular)**矩阵**,即它对应的**行列式**(determinant)具有非零值。

3.3.3 矢量空间、基和维度

给定一个由 k 个矢量组成的矢量组，假设我们能从其中找出 r 个线性无关的矢量组成的子集。即，剩下的矢量能够用这 r 个矢量的线性组合来表示。我们将这 r 个矢量称为该矢量组的**基**(basis)。它们组成了矢量组的基本核，从中我们可以导出其他的矢量。在此意思上，剩余的矢量可以被认为是冗余的。例如，在例 3.6 中，第一和第三个矢量构成一个基，第二个矢量能由基矢量来生成，即 $\boldsymbol{x}_2=6\boldsymbol{x}_1-\boldsymbol{x}_3$。注意，任意 r 个线性独立的矢量子集都能构成一个有效的基，因此基矢量集不是唯一的。

现在我们能对此观察结果做如下推广。假定我们有 r 个线性无关的矢量组成的矢量集，则由该集合中的矢量的线性组合所生成矢量集合是什么。显然，线性组合生成的矢量有无数个。我们将这样的一个无限集合称为从基矢量集生成的**矢量空间**(vector space)(注意，矢量空间是一个被精确定义的数学对象；这里所说的只是非正式的概念定义)。基矢量的数量——即基矢量集的基数(cardinality)——称为空间的**维度**(dimension)。

例 3.7：基和维度

有一个简单的方法可以保证一组矢量是相互线性无关的，即一个矢量中除一个元素外其余全都设为零元素。例如，矢量 $\boldsymbol{x}_1=[1\quad 0\quad 0]$，$\boldsymbol{x}_2=[0\quad 1\quad 0]$，$\boldsymbol{x}_3=[0\quad 0\quad 1]$确定是相互线性无关的。我们来计算有这个基矢量集生成的矢量空间。考虑任意矢量 $\boldsymbol{x}=[a\quad b\quad c]$，这个矢量可以用基矢量集的线性组合来表示，即 $\boldsymbol{x}=a\boldsymbol{x}_1+b\boldsymbol{x}_2+c\boldsymbol{x}_3$。因此，该基矢量集可以生成所有由三个实数元素组成的矢量，这就是它的矢量空间。

如果我们考虑用一个具有三个实数元素的矢量来对应三维空间中的一个点，该矢量的三个元素值就是对应点的笛卡儿坐标值，基矢量组能够生成三维空间所有的点。容易看到基矢量组对应于三条坐标轴。现在我们应该清楚为什么说基矢量组生成了一个空间，以及基矢量组的基数就是该空间的维度。

3.4 使用矩阵代数求解线性方程组

我们现在将注意力转移到一个重要运用上来，即使用矩阵代数来求解线性方程组。

3.4.1 表示

线性方程组可以很方便的用矩阵来表示。考虑以下线性方程式：

$$3x+2y+z=5$$
$$-8x+y+4z=-2$$
$$9x+0.5y+4z=0.9$$

我们可以用矩阵来表示以上线性方程组：

$$\begin{bmatrix} 3 & 2 & 1 & 5 \\ -8 & 1 & 4 & -2 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$$

实际上，一个数字在矩阵中所处的位置也暗示了它是方程式中某个变量的系数或者等式右边的一个值。这种矩阵的表示方式可以适用于任何线性等式组。如果矩阵最右侧的列是 0，我们称其为**齐次**(homogeneous)线性方程组。对应于线性等式的左侧子矩阵称为**系数矩阵**(coefficient matrix)。

3.4.2　初等行运算与高斯消元法

定义一组等式，我们可以通过简单的运算生成新的等式。例如，在等式左侧和右侧同乘一个标量从而生成一个新的等式。此外，我们将任意一对方程的等式左侧和右侧同加、减来生成新的等式。

在我们先前的例子中，前两个等式是 $3x+2y+z=5$ 和 $-8x+y+4z=-2$。我们能够将第一个等式乘以 3 得到新的等式 $9x+6y+3z=15$。也可以将等式 1 和 2 相加获得一个新的等式 $(3-8)x+(2+1)y+(1+4)z=(5-2)$，即 $-5x+3y+5z=3$。

我们也能将这些运算相互组合。例如，将等式 2 乘以 2，然后从等式 1 减去它，得到：

$$(3-16)x+(2-2)y+(1-8)z=(5-4)$$
$$19x-7z=9$$

在得到的等式中，y 变量被消去了(即等式中不再出现 y 变量)。类似地，我们可以将等式 3 乘以 4，然后将等式 1 减去它，从而获得另一个同样也消去了 y 变量的等式。我们现在得到两个含有两个变量的等式，从中可以简单的求出变量 x 和 z。将结果代入任何一个等式从而求出 y 的值。

从本质上来说，这种方法就是著名的**高斯消元法**(Gaussian elimination)。使用这个方法，我们选择任一变量，对方程组运用乘法和加法，从而将该变量从等式中消去。从而将一个具有 n 个变量以及 m 个等式的方程组转化有$(n-1)$个变量和$(m-1)$个等式的方程组。以此类推，最终[①]我们将获得只包含一个变量的等式，从而求得该变量的解。通过将求得变量值代入简化后的方程组，我们可以求得系统的解。

当我们使用矩阵来表示线性方程组时，对等式乘以标量和将两个等式的加法这些初等运算对应于两个矩阵**行运算**(row operations)。第一个行运算是将一行所有的元素乘以一个标量，第二个行运算是将两行元素逐个相加。很容易看出，这些操作与前面章节所述的相类似。高斯消元法使用基本的行运算对线性方程组的矩阵表示进行操作，从而使一行呈现以下形式：$[0\ 0\ \ldots\ 0\ 1\ 0\ \ldots\ 0\ a]$，从中我们可以得出对应变量的值。将该变量值代入其他等式，从而我们得到一个少了一个未知变量的方程组，通过递归，进一步求得所有变量的值。

例 3.8：高斯消元法

请使用行运算和高斯消元法来求解用矩阵表述的以下方程组

$$\begin{bmatrix} 3 & 2 & 1 & 5 \\ -8 & 1 & 4 & -2 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$$

解：将第 2 行减去第 3 行，得到：

① 假定方程组是自洽的(self-consistent)且至少有一组解。后面我们还将进一步介绍。

$$\begin{bmatrix} 3 & 2 & 1 & 5 \\ -17 & 0.5 & 0 & -2.9 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$$

第 3 行乘以 0.25 后，第 1 行减去它，得到：

$$\begin{bmatrix} 0.75 & 1.875 & 0 & 4.775 \\ -17 & 0.5 & 0 & -2.9 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$$

请注意，第 1 行和第 2 行代表一个二元的方程组。第 2 行乘以 1.875/0.5＝3.75，第 1 行减去它，得到：

$$\begin{bmatrix} 64.5 & 0 & 0 & 15.65 \\ -17 & 0.5 & 0 & -2.9 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$$

从而可以得出 x 的值为 15.65/66.525＝0.2426。将 x 的值代入第 2 行，我们得到 $-17\times0.2426+0.5y=-2.9$，从而得出 $y=2.4496$。将 x 和 y 的值代入第 3 行，我们得到 $9\times0.2426+0.5\times2.4496+4z=0.9$，从而得出 $z=0.6271$。代入检查，$3\times0.2426+2\times2.4496-0.6271=4.9975$，误差为 0.0005。

在实际操作中，选择最先被消去的变量非常重要。不明智的选择会使我们需要在计算中对矩阵元素保持很高的精度，这样开销很大。现有大量算法方面的工作用来谨慎选择要消去的变量，这被称之为**主元**（pivot）。标准的矩阵软件包，如 MATLAB 等可以实现这些算法。

3.4.3 秩

到此为止，我们都假设一个线性方程组有一组**一致**（consistent）的解。但实际情况中并非总是如此。如果一个线性方程组是**超定方程组**或者是**欠定方程组**，则它无解或者有无穷个数的解。如果在一个系统中，同一个变量被假定具有不一致的值，我们称之为超定系统。例如，一个简单的超定系统是等式 $x=1$ 和 $x=2$ 的集合。高斯消元法对这样系统无能为力。

如果一个方程组中的变量允许有多个值，该系统称为欠定系统。欠定系统的一个简单例子是由线性等式 $x+y=1$ 构建的系统，因为满足该等式的 x 和 y 值有无限个。对这样的方程组运用高斯消元法，得到的是一个变量组，这个变量组可表示为相互独立的变量的线性组合。对独立变量的每次赋值会得出该系统的一组一致解。

定义一个由 n 个变量组成的 m 个线性方程组，如果 $m<n$，则方程组是欠定的。如果 $m\geqslant n$，方程组有可能是欠定的，也有可能不是，这取决于该方程组中是否有等式是重复的。尤其是，当定义一个方程，它可以用一组其他方程的线性组合来表达，则我们称它与一组其他方程**线性相关**：对应于这个方程的矢量是由对应于其他方程的矢量组的线性组合。如果线性方程组中的一个方程与其他方程线性相关，则通过适当运用加法和乘法，我们可以将这个方程转变为 0＝0 的等式。这样的话，这个方程不再为我们提供任何附加的信息，也能够在不改变方程解的前提下，将它从方程式中删除。

如果一个系统由 m 个方程表示，其中有 k 个方程能用其他 $m-k$ 个方程的线性组合来表达，则我们实际只有 $m-k$ 个方程式用于求解。这个值称为系统的**秩**，用 r 表示。如果

$r<n$，方程组是欠定方程组，如果 $r=n$，方程组有唯一的解，如果 $r>n$，方程组是超定方程组因而无一致解。注意一个矩阵的秩等同于其相对应的行矢量组的基矢量集的基数。

例 3.9：秩

我们已经求出方程组 $\begin{bmatrix} 3 & 2 & 1 & 5 \\ -8 & 1 & 4 & -2 \\ 9 & 0.5 & 4 & 0.9 \end{bmatrix}$ 中的 x、y 和 z，有一个唯一一致的解。因此，它的秩是 3。

定义一个方程组 $\begin{bmatrix} 3 & 2 & 1 & 5 \\ -8 & 1 & 4 & -2 \\ 6 & 4 & 2 & 10 \end{bmatrix}$，可以看出第三行是第一行乘以 2 的结果。因此，由于第三行没能为方程组增加任何额外的信息，它能被移除。该方程组的秩是 2（它是欠定方程组），剩余方程组有无限个解。

现在，定义一个方程组 $\begin{bmatrix} 3 & 2 & 1 & 5 \\ -8 & 1 & 4 & -2 \\ 9 & 0.5 & 4 & 0.9 \\ 3 & 2 & 1 & 4 \end{bmatrix}$。我们已知前三行是线性无关的，秩为 3，第四行与第一行不一致，因此该方程组是超定方程组，无解。该方程组对应的系统秩为 4。

现在已知有多种确定系统方程组秩的技术。然而，这超出了本书讨论的范围。我们在此学习秩的目的是，帮助理解高斯消元法以及可在求得不一致解时报告系统是超定的，即其秩不小于 n；在方程组有无穷多组解时报告系统是欠定的，即方程组的秩小于 n。如果系统是欠定的，系统的秩等于对应方程组中不能约简为 $0=0$ 的方程个数。

3.4.4 行列式

我们现在将注意力转移到学习矩阵的行列式上来。即便是对数学充满热情的数学家也不得不承认学习行列式是一项非常枯燥的课题。并且，一个矩阵的行列式其本身没有过多的实际操作价值。尽管我们可以用行列式来简洁的表示一个线性方程组，但使用行列式来求方程组的解并不可行。掌握行列式是深入学习矩阵特征值的前提，这是我们为什么要坚持学习掌握行列式的真正原因。

一个 2 行 2 列的矩阵 $\boldsymbol{A}$ 的行列式是一个标量，我们用 $D=\det\boldsymbol{A}$ 来表示：

$$D=\det\boldsymbol{A}=\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}=a_{11}a_{22}-a_{12}a_{21} \tag{3.9}$$

请注意，这里我们使用垂直线代替方括号来表示矩阵的行列式，而非矩阵本身。一个 2 行 2 列的矩阵的行列式称为 2 阶行列式。

我们需要引入**子阵**的概念来帮助描述更大型方阵的行列式，子阵是相对于元素 a_{jk} 而言的。将矩阵 $\boldsymbol{A}$ 中 a_{jk} 所在的 j 行和 k 列元素删除，剩余矩阵称为 a_{jk} 的对应子阵，用 $S_{jk}(\boldsymbol{A})$ 表示。该子阵的行列式，即 $\det S_{jk}(\boldsymbol{A})=|S_{jk}(\boldsymbol{A})|$，它是一个标量，称为元素 a_{jk} 的**余子式**，用 M_{jk} 表示。

请注意一个矩阵的子阵比对应的矩阵少一行一列。一个 n 行 n 列的矩阵的行列式是

n 阶的，因此它的每个余子式是 $n-1$ 阶的。

现在我们定义另一个辅助术语，元素 a_{jk} 的**代数余子式**，用 C_{jk} 表示。C_{jk} 的定义如下：

$$C_{jk} = (-1)^{j+k} M_{jk} \tag{3.10}$$

现在我们可以给出矩阵的行列式的定义了：

$$D = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{k=1}^{n} a_{ki} C_{ki} \tag{3.11}$$

这里 i 是任意行或者列。可以证明，无论我们选择哪一行或者哪一列做展开式，矩阵的行列式 D 都不会有变化。并且，从 3.7 节可知，矩阵转置后，它的行列式仍然不会改变，即 $|\boldsymbol{A}| = |\boldsymbol{A}^T|$。

例 3.10：行列式

请计算矩阵 $\begin{bmatrix} 2 & 5 & -2 \\ 4 & 9 & 8 \\ 3 & 0 & 1 \end{bmatrix}$ 的行列式。

解：由于元素 $a_{32}=0$，我们可以忽略掉中间的一个代数余子式，因此我们通过对矩阵的第三行进行展开来计算其行列式更为方便。该矩阵的行列式是：

$$\begin{aligned} a_{31}C_{31} + a_{33}C_{33} &= 3(-1)^{3+1}M_{31} + 1(-1)^{3+3}M_{33} = 3\begin{vmatrix} 5 & -2 \\ 9 & 8 \end{vmatrix} + 1\begin{vmatrix} 2 & 5 \\ 4 & 9 \end{vmatrix} \\ &= 3(40-(-18)) + 1(18-20) = 174 + (-2) = 172 \end{aligned}$$

我们通过扩展第二列来计算行列式，以此作为验证：

$$\begin{aligned} a_{12}C_{12} + a_{22}C_{22} &= 5(-1)^{1+2}M_{12} + 1(-1)^{2+2}M_{22} = -5\begin{vmatrix} 4 & 8 \\ 3 & 1 \end{vmatrix} + 9\begin{vmatrix} 2 & -2 \\ 3 & 1 \end{vmatrix} \\ &= -5(4-24) + 9(2-(-6)) = 100 + 72 = 172 \end{aligned}$$

这里给出关于行列式的一些有用的特性：

- 可以通过展开一个矩阵的任意一行或者任意一列来计算它的行列式。因此，如果一个矩阵有一行或一列为 0，它的行列式也为 0。
- 一个矩阵的某一行或者某一列中每个元素都乘以一个常量 c，它的行列式等于原矩阵的行列式乘以相同的因子。
- 将矩阵 $\boldsymbol{A}$ 的任意两行或者两列相互交换从而得到矩阵 $\boldsymbol{B}$，则 $|\boldsymbol{B}| = -|\boldsymbol{A}|$。由于 0 是唯一一个正负相等的数字，因此当一个矩阵有相同的两行或者两列时，它的行列式一定为 0。
- 当且仅当一个 n 行 n 列的方阵有一个非零的行列式时，它的阶是 n。
- 当且仅当一个方阵有一个非零的行列式时，它有一个逆矩阵（即该方阵是非奇异的）。

3.4.5 克莱姆定理

通过计算一个矩阵的行列式，我们可以轻松地求解一个方程组（至少在理论上是这样的）。在实践中，通过计算行列式来求解方程组要比通过高斯消元法来求解消耗更多资源，因此，大多时候我们使用**克莱姆定理**来深入了解方程求解过程的本质。

克莱姆定理表明：对于一个包含 n 个变量的，n 个线性方程式组成的方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 来说，如果系数矩阵 $\boldsymbol{A}$ 有一个非零的系数行列式 $D=\det\boldsymbol{A}$，则该方程组恰有唯一的解

$$x_i = D_i / D$$

这里 D_i 是将 $\boldsymbol{b}$ 代入矩阵 $\boldsymbol{A}$ 的第 i 列所获得的新矩阵的行列式。因此，如果我们已知相应的行列式，可以利用以上公式直接计算出 x_i 的值。克莱姆定理也称为**克莱姆法则**。

当 $\boldsymbol{b}=0$ 时，我们称以上方程组是齐次方程组。在这种情况下，每个 D_i 都是 0，因此每个 x_i 也是 0。如果系数矩阵 $\boldsymbol{A}$ 的行列式 D 为 0，且这个方程组是齐次的，则根据克莱姆法则，每个变量 x_i 将被赋予 0/0 的无穷值。然而在这种情况下，方程组实际上存在非零的解。这一重要的事实是我们计算矩阵特征值的出发点。

3.4.6　逆矩阵

一个方阵 $\boldsymbol{A}$ 的逆矩阵表示为 $\boldsymbol{A}^{-1}$，它满足以下条件：$\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$。

例 3.11：逆矩阵

请验证一个矩阵的逆矩阵是唯一的。

解：如果矩阵 $\boldsymbol{A}$ 有一个逆矩阵 $\boldsymbol{B}$，同时还有一个逆矩阵 $\boldsymbol{C}$，则 $\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{AC}=\boldsymbol{CA}=\boldsymbol{1}$。因此 $\boldsymbol{B}=\boldsymbol{BI}=\boldsymbol{B}(\boldsymbol{AC})=(\boldsymbol{BA})\boldsymbol{C}=\boldsymbol{IC}=\boldsymbol{C}$。

并非所有的方阵都是可逆的：一个没有逆矩阵的矩阵称为**奇异矩阵**(singular matrix)。所有的奇异矩阵的行列式都为 0。如果一个矩阵不是奇异矩阵，则它的逆矩阵为：

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}[C_{jk}]^{\mathrm{T}} = \frac{\boldsymbol{1}}{|\boldsymbol{A}|}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \tag{3.12}$$

这里，C_{jk} 是元素 a_{jk} 的代数余子式。注意，相对于 $\boldsymbol{A}$ 来说，矩阵代数余子式 C_{jk} 是转置的。作为特例，一个 2 行 2 列的矩阵 $\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$，其逆矩阵是

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{3.13}$$

例 3.12：逆矩阵

请计算矩阵$\begin{bmatrix} 2 & 3 \\ 6 & 2 \end{bmatrix}$的逆矩阵。

解：该矩阵的行列式是 $2\times2-6\times3=-14$。我们可以使用式(3.13)，计算得出它的逆矩阵是$\frac{1}{-14}\begin{bmatrix} 2 & -3 \\ -6 & 2 \end{bmatrix}$。

3.5　线性变换、特征值和特征向量

本节讨论线性变换这个重要问题，以及如何使用特征值和特征向量来进行线性变换。

3.5.1　矩阵的线性变换

回忆一个 $n\times n$ 矩阵与一个 n 维列矢量(即一个 $n\times1$ 矩阵)的乘积是另一个 n 维列矢

量。因此我们可以将此矩阵看作是输入矢量到输出矢量的**变换**(transforming)。

注意,输出列矢量的第 k 个元素是将输入矢量的所有元素与矩阵的第 k 行元素进行加权求和得到。这个过程正好是对输入矢量元素的一个线性组合。表示这样一组线性组合(相应的权值)的方阵称为是对输入矢量的一个**线性变换**(linear transformation)。

例 3.13:使用矩阵进行线性变换

定义一个矩阵$\begin{bmatrix}2 & 3\\6 & 2\end{bmatrix}$,一个输入矢量$\begin{bmatrix}a\\b\end{bmatrix}$。矢量和矩阵的乘积是输出矢量$\begin{bmatrix}2 & 3\\6 & 2\end{bmatrix}\times\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}2a+3b\\6a+2b\end{bmatrix}$。输出矢量的第一个元素是输入元素 a 和 b 分别与权值 2 和 3(即矩阵的第一行元素)相乘的和,输出矢量的第二个元素是是输入元素 a 和 b 分别与权值 6 和 2(即矩阵的第二行元素)相乘的和。

根据矩阵乘法的定义,我们可以用矩阵乘积来表示两次线性变换的组合。特别是假设矩阵 $\boldsymbol{A}$ 将一个列矢量 x 变换成另一个列矢量 $\boldsymbol{x}'$,接着矩阵 $\boldsymbol{B}$ 将 $\boldsymbol{x}'$ 变换成另一个矢量 $\boldsymbol{x}''$。则 $\boldsymbol{x}''=\boldsymbol{Bx}=\boldsymbol{B}(\boldsymbol{Ax})=(\boldsymbol{BA})\boldsymbol{x}=\boldsymbol{Cx}$,这里 $\boldsymbol{C}=\boldsymbol{BA}$ 就是两次变换矩阵的组合。也就是说,我们可以用两个变转矩阵的乘积来表示两次变换的组合。

例 3.14:线性变换的组合

我们已知矩阵$\begin{bmatrix}2 & 3\\6 & 2\end{bmatrix}$将一个输入矢量$\begin{bmatrix}a\\b\end{bmatrix}$变换成输出矢量$\begin{bmatrix}2a+3b\\5a+2b\end{bmatrix}$。假设我们运用矩阵$\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$对输出矢量再次进行线性变换,得到的结果是$\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\times\begin{bmatrix}2a+3b\\5a+2b\end{bmatrix}=\begin{bmatrix}4a+6b\\12a+4b\end{bmatrix}$。同时我们计算两个变换矩阵的乘积为$\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\times\begin{bmatrix}2 & 3\\6 & 2\end{bmatrix}=\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}$,然后使用这个乘积矩阵对输入矢量进行线性变换,得到$\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}\times\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}4a+6b\\12a+4b\end{bmatrix}$,与前面的运算结果一致。

3.5.2 矩阵的特征值

定义一个矩阵$\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$和一个输入矢量$\begin{bmatrix}a\\b\end{bmatrix}$,输入矢量和矩阵相乘得到输出矢量$\begin{bmatrix}2a\\2b\end{bmatrix}=2\times\begin{bmatrix}a\\b\end{bmatrix}$。因此,这个矩阵对输入矢量做了个翻倍的变换:将该矩阵作用于任何矢量的结果,等同于将矢量乘以标量 2。

当一个矩阵与一个特殊矢量的乘积与一个标量与该矢量的乘积相同时,我们将这个标量称为为矩阵的**特征值**(eigenvalue),相应的矢量称为**特征向量**(eigenvector)。更加精确地,我们定义方阵 $\boldsymbol{A}$ 的特征值为标量 λ,此时对于某个非零的列矢量 $\boldsymbol{x}$ 存在以下等式

$$\boldsymbol{Ax}=\lambda\boldsymbol{x} \tag{3.14}$$

特征值的大小标志着矩阵运算对一个特征向量的缩放程度:特征值越大,矩阵运算对

特征向量的缩放影响越大。

例 3.15：特征值和特征向量

请计算矩阵$\begin{bmatrix} 4 & 6 \\ 12 & 4 \end{bmatrix}$的特征值以及相对应的特征向量。

解：定义特征向量是$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$，则$\begin{bmatrix} 4 & 6 \\ 12 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\lambda\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$。将等式左侧展开，我们得到由两个方程式组成的一个方程组：

$$\begin{cases} 4x_1 + 6x_2 = \lambda x_2 \\ 12x_1 + 4x_2 = \lambda x_2 \end{cases} \tag{3.15}$$

移项，得到

$$(4-\lambda)x_1 + 6x_2 = 0$$
$$12x_1 + (4-\lambda)x_2 = 0$$

这是一个有两个方程式组成的齐次方程组。从 3.4.5 节可知，当且仅当系数矩阵的行列式为 0 时，该方程组有一个非零的解。即

$$\begin{vmatrix} (4-\lambda) & 6 \\ 12 & (4-\lambda) \end{vmatrix} = 0$$

将其展开得到二次方程式：

$$(4-\lambda)^2 - 72 = 0$$
$$\lambda = 4 \pm \sqrt{72} = 4 \pm 6\sqrt{2}$$

根据给出的特征值，我们来计算相应的特征向量，过程如下。首先将特征值$\lambda=4\pm6\sqrt{2}$带代入式(3.15)，得到

$$(4-4-6\sqrt{2})x_1 + 6x_2 = 0$$
$$12x_1 + (4-4-6\sqrt{2})x_2 = 0$$
$$-6\sqrt{2}x_1 + 6x_2 = 0$$
$$12x_1 - 6\sqrt{2}x_2 = 0$$

两式约简后都可得到

$$\sqrt{2}x_1 - x_2 = 0$$

我们有两个变量但仅有一个包含这两个变量的等式，它对应于无限多个特征向量的解，用一个自由变量来表示可得$\begin{bmatrix} x_1 \\ \sqrt{2}x_1 \end{bmatrix}$。例如，我们定义 $x_1=1$，一个可能的特征向量是$\begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$。特征向量的集合也可用 $a\begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$来表示，其中 a 是任意标量。

作为验证，注意 $Ax = \begin{bmatrix} 4 & 6 \\ 12 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} 4+6\sqrt{2} \\ 12+4\sqrt{2} \end{bmatrix}$，而 $\lambda x = (4+6\sqrt{7}) \times \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} 4+6\sqrt{2} \\ 12+4\sqrt{2} \end{bmatrix}$。

我们从几何的角度对此作如下的解释。假定我们将笛卡儿坐标系中的一个点(x,y)用矢量$\begin{bmatrix}x\\y\end{bmatrix}$表示。参数为$x_1$的特征向量$\begin{bmatrix}x_1\\\sqrt{2}x_1\end{bmatrix}$代表的是直线$y=\sqrt{2}x$上所有点的集合。由于矩阵对特征向量的影响等同于一个标量对特征向量的影响,因此,该直线上所有点与矩阵进行线性变换后,得到的点仍在同一条直线上。并且,矩阵对特征向量的缩放程度就是特征值$4+\sqrt{72}$。也就是说,在这条直线上距离原点为一个单位距离的点(即一个单位特征向量)经矩阵线性变换后,变成了一个仍在该直线上且距离原点为$4+\sqrt{72}$的点。这个结果如图3.1所示。

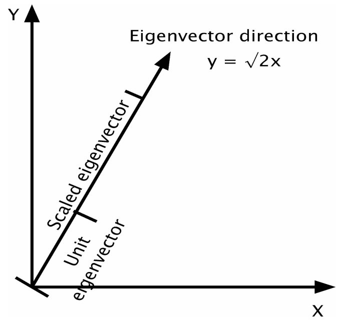

图3.1 矩阵对特征向量的影响就是沿着原方向对其进行缩放

如果要得到其他特征向量,我们只要将特征值$\lambda=4-\sqrt{72}$代入式(3.15)中,

$$(4-4-\sqrt{72})x_1+6x_2=0$$
$$12x_1+(4-4-\sqrt{72})x_2=0$$
$$\sqrt{72}x_1+6x_2=0$$
$$12x_1+\sqrt{72}x_2=0$$

从而得到参数化的特征向量解$\begin{bmatrix}x_1\\-\sqrt{2}x_1\end{bmatrix}$,它也可以用$a\begin{bmatrix}1\\-\sqrt{2}\end{bmatrix}$来表示,其中$a$是任意的一个标量。

3.5.3 计算矩阵的特征值

现在我们对以上例子中计算特征值和特征向量的方法进行总结。定义一个矩阵$\boldsymbol{A}$,它的特征向量$\boldsymbol{x}$,对应的特征值是λ。根据定义我们知道

$$\boldsymbol{Ax}=\lambda\boldsymbol{x}$$

将以上等式改写成

$$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{x}=0 \tag{3.16}$$

这是一个齐次方程式,从3.4.5节可知,当且仅当系数矩阵的行列式为0时,方程式有非平凡的解:

$$|\boldsymbol{A}-\lambda\boldsymbol{I}|=0 \tag{3.17}$$

该行列式称为矩阵的**特征行列式**(characteristic determinant),式(3.17)称为**特征方程式**(characteristic equation)。通常来说,展开行列式可以到一个关于λ的n阶多项式,这就是矩阵的**特征多项式**(characteristic polynomial)。如我们所见,矩阵的特征值就是特征多项式的根。这个重要结论允许我们计算任何矩阵的特征值。同时注意,如果矩阵被转置,矩阵的特征行列式的值不会改变。因此,一个矩阵与其转置矩阵的特征值是一致的。

一般来说,n阶多项式的根值可能是实数或者复数。而且,从代数学的基础理论可知,

此多项式至少有一个、最多有 n 个相异的根。因此，n 阶方阵具有一个 1 到 n 个不同的特征值，如果矩阵是实的，则其中的某些根可能是复数并构成复共轭对。此外，有些特征值是重复的。每个特征值对应于一个特征向量族，这些特征向量至少有一个自由变量参数。

我们将一个矩阵的特征值集合称为它的**谱**(spectrum)。值最大的特征值称为**主特征值**(principal eigenvalue)，或者是矩阵的**谱半径**(spectral radius)。每个特征值对应一个特征向量族。可以证明，对应不同特征值的特征向量之间通常是相互独立的。所有特征向量的线性组合所构成的矢量集合称为矩阵的**特征空间**(eigenspace)。

例 3.16：复数特征值

我们来计算矩阵 $\mathbf{A}=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$的特征向量。我们通过将特征行列式置为 0 来求取：

$$\begin{vmatrix}-\lambda & 1\\ -1 & -\lambda\end{vmatrix}=0$$

从而得出特征方程式

$$\lambda^2+1=0$$

得到 $\lambda=\pm i$。这就是矩阵的谱，且它的谱半径是 1。

令

$$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=i\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

$$x_2=ix_1$$

$$-x_1=ix_2$$

这相当于设方程组的秩为 1；即第二个方程式是第一个方程式的变形。此时，我们可求得 $\lambda=i$ 对应的特征向量。

一个满足条件的特征向量是$\begin{bmatrix}1\\ i\end{bmatrix}$。将其代入等式可得到验证，$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}1\\ i\end{bmatrix}=i\begin{bmatrix}1\\ i\end{bmatrix}$。类似地，可以计算出，特征向量 $\lambda=-i$ 对应的特征向量是$\begin{bmatrix}x_1\\ -ix_1\end{bmatrix}$。

由于两个特征向量族都是复数，因此线性变换矩阵绝对不会让实值向量的方向保持不变。事实上，如我们所期望的那样，本例中矩阵对应于一个 90°的旋转。但我们未曾预料的是，该旋转矩阵并未改变复值向量的方向。这个结果没有明显直观的解释。

例 3.17：一个对角矩阵的特征值

定义一个对角矩阵 $\mathbf{A}=\begin{bmatrix}a & 0 & 0\\ 0 & b & 0\\ 0 & 0 & c\end{bmatrix}$，求其特征值。

矩阵 $\mathbf{A}$ 对应的特征行列式是

$$\begin{bmatrix}a-\lambda & 0 & 0\\ 0 & b-\lambda & 0\\ 0 & 0 & c-\lambda\end{bmatrix}=0$$

从而得出特征方程式

$$(a-\lambda)(b-\lambda)(c-\lambda)=0$$

这表明该矩阵的特征值就等于对角元素 a,b,c。这个结论可推而广之：对于任何对角矩阵而言，它的特征值就是它的矩阵对角元素。

3.5.4 特征值的重要性

当我们对一个输入矢量进行重复多次线性变换运算时，矩阵特征值就将显示其重要作用了。假设我们用一个矢量表示一个系统的状态，其中矢量的每个元素分别对应该系统的某些方面，例如缓冲区占用程度(我们将在 8.2.2 节进行更加详细讨论)。进一步假定我们可以将时间上的一步状态变换等价地表示对输入矢量应用状态变换操作(和等效的变换矩阵)。因而，我们能够通过对初始矢量重复进行矩阵变换运算，以得到系统的稳定状态，或者是最终状态。这说明了变换矩阵的特征值能被用来表示系统的稳态特征。

为证明这一点，首先考虑用矩阵 $\boldsymbol{A}$ 对特征值为 λ 的特征向量 $\boldsymbol{x}$ 重复进行变换的情况。按照定义，对于矢量 $\boldsymbol{x}$ 的单次矩阵变换是 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$。重复 n 次进行矩阵变换可以转换成计算 $\lambda^n\boldsymbol{x}$，这要比直接进行 n 次矩阵变换计算简单得多!

现在，考虑一个初始状态矢量 $\boldsymbol{v}$，$\boldsymbol{v}$ 可以表示为矩阵 $\boldsymbol{A}$ 的两个特征向量 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的线性组合，即：

$$\boldsymbol{v}=c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2$$

假定特征向量对应的特征值分别是 λ_1 和 λ_2，则 $\boldsymbol{A}^n\boldsymbol{v}$ 能够用以下公式表示：

$$\boldsymbol{A}^n\boldsymbol{v}=\boldsymbol{A}^n(c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2)=c_1\boldsymbol{A}^nx_1+c_2\boldsymbol{A}^nx_2=c_1\lambda_1^n\boldsymbol{x}_1+c_2\lambda_2^n\boldsymbol{x}_2$$

我们看到，原来是很复杂的在矢量 $\boldsymbol{v}$ 上重复进行矩阵 $\boldsymbol{A}$ 变换的计算，被替换成计算一个标量的 n 次幂，这样要相对容易很多。

这个直观的结果很容易进一步推广。如果我们可用矩阵的特征向量的线性组合来表示初始矢量，能够看到重复进行矩阵运算变得轻而易举。下面的例子说明了这种情况。

例 3.18：使用特征向量计算 $\boldsymbol{A}^nX$

从前面的例子我们可知矩阵 $\boldsymbol{A}=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$有两个特征值 i 和 $-i$，其对应的单位特征向量是$\begin{bmatrix}1\\ i\end{bmatrix}$以及$\begin{bmatrix}1\\ -i\end{bmatrix}$。定义一个状态矢量$\begin{bmatrix}10\\ 0\end{bmatrix}$，它可以特征向量表示为 $5\times\begin{bmatrix}1\\ i\end{bmatrix}+5\times\begin{bmatrix}1\\ -i\end{bmatrix}$。因此，$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}^{100}\begin{bmatrix}10\\ 0\end{bmatrix}$，这个直接计算很复杂的运算，可转换为计算 $5\times i^{100}\times\begin{bmatrix}1\\ i\end{bmatrix}+5\times i^{100}\times\begin{bmatrix}1\\ -i\end{bmatrix}=5\times\begin{bmatrix}1\\ i\end{bmatrix}+5\times\begin{bmatrix}1\\ -i\end{bmatrix}=\begin{bmatrix}10\\ 0\end{bmatrix}$。从几何学的角度来看则很容易理解，对矢量运用一次矩阵 $\boldsymbol{A}$ 运算相当于将它旋转 $90°$，应用任意 4 的倍数次矩阵运算，可保持原始矢量不发生变化。

当原始矢量可表示为特征向量的线性组合时，运用以上方法计算重复多次线性变换的结果是很有用的。在这种情形下，以下事实将有其用途：对应于一组不同特征值的特征向量构成一个**线性独立集**(linearly independent set)。我们已经看到，任意 n 个线性独立的复

数矢量形成复数超空间 C^n 的一个基。因此，如果一个 n 阶方阵有 n 个不同的特征值，我们可以用该矩阵的特征向量的线性组合来表示任何初始矢量。这是为什么知道矩阵特征值非常有用的理由之一。

如果初始状态矢量不能用特征向量的线性组合来表示，那将如何？也就是说，如果初始状态矢量不在矩阵的特征空间中，那该如何？在这种情况下，一种更加复杂的计算，仍涉及到特征值，可让我们求得 $\boldsymbol{A}^n\boldsymbol{x}$。

3.5.5　主特征值的作用

考虑一个矩阵 $\boldsymbol{A}$，它有 m 个特征值 λ_i，$i=1,2,\dots,m$，分别对应单位特征向量 $\boldsymbol{x}_i$。假设我们选择该矩阵特征空间中的一个矢量 $\boldsymbol{v}$，从而该矢量可以表示为所有特征向量的线性组合：

$$\boldsymbol{v}=\sum_{i=1}^{m}c_i\boldsymbol{x}_i$$

然后，我们对该矢量进行 n 次矩阵变换运算，得到以下矢量

$$\sum_{i=1}^{m}c_i\lambda_i^n\boldsymbol{x}_i$$

当 $n\to\infty$，矢量和被取值最大的特征值所控制，该特征值称为**主特征值**(principle，或 dominant eigenvalue)(如果有多个特征值幅度相同，则它们都被认为是主特征值)。在计算极限值时，我们可以忽略所有其他的特征值。

例 3.19：主特征值

定义一个矩阵 $\boldsymbol{A}=\begin{bmatrix}4 & 6\\ 12 & 4\end{bmatrix}$，从例 3.15 已知该矩阵有两个特征值，$\lambda=4\pm\sqrt{72}$，大约是 12.48 和 −4.48。矩阵的单位特征向量是 $\begin{bmatrix}\frac{1}{\sqrt{2}}\end{bmatrix}$ 和 $\begin{bmatrix}1\\ \frac{-1}{\sqrt{2}}\end{bmatrix}$。定义初始矢量是 $\begin{bmatrix}0\\ 3\sqrt{2}\end{bmatrix}$，我们可以表示为 $2\times\begin{bmatrix}\frac{1}{\sqrt{2}}\end{bmatrix}-2\times\begin{bmatrix}1\\ \frac{-1}{\sqrt{2}}\end{bmatrix}$。则 $\boldsymbol{A}^{10}\begin{bmatrix}0\\ 3\sqrt{2}\end{bmatrix}=2\times 12.48^{10}\times\begin{bmatrix}\frac{1}{\sqrt{2}}\end{bmatrix}-2\times(-4.48)^{10}\begin{bmatrix}1\\ \frac{-1}{\sqrt{2}}\end{bmatrix}$。估算值为 $\begin{bmatrix}1.83\times 10^{11}\\ 2.59\times 10^{11}\end{bmatrix}$。如果我们忽略第二项(即，忽略特征值 −4.48 的贡献)，结果值仅在小数点后第三位有变化。很明显，主特征值发挥主要作用。

3.5.6　寻找特征值和特征向量

在实践中经常需要我们计算矩阵的特征值和特征向量。下面列举的一些事实，可以帮助我们识别矩阵特征值的本质特性。

- 如果一个矩阵是对角方阵，则它的特征值就是它的对角元素。
- 如果一个矩阵是对称方阵(即 $\boldsymbol{A}^T=\boldsymbol{A}$)，则它的特征值是实数。
- **盖氏“圆盘”定理**(Gerschgorin’s “circle” theorem)表明，一个复数矩阵的所有特征

值都在复平面的“圆盘”集合上，这个圆盘以对角线元素为中心，半径等于矩阵非对角元素幅度之和。直观来说，如果非对角元素都“不太大”，矩阵的特征值就是它的对角元素。

例 3.20：寻找特征值

矩阵 $\boldsymbol{A}=\begin{bmatrix}9.1 & 0.8 & 0.3\\ 0.8 & 5.8 & 0.2\\ 0.3 & 0.2 & 6.5\end{bmatrix}$是对称矩阵，因此它的特征值是实数。它有三个盖氏圆盘：(1)中心 $9.1+0i$，半径 1.1，(2)中心 $5.8+0i$，半径 1.0，(3)中心 $6.5+0i$，半径 0.5。由于特征值是实数，它们必然位于[8, 10.2]，[4.8, 6.8]，[6, 7]三个区间中间的一个。由于第二个区间与第三个区间有部分重叠，因此我们知道特征值位于[4.8, 7]或者[8, 10.2]区间内。

使用**幂(power)方法**，可以近似计算一个矩阵[①]的主特征值。在此方法中，我们从任意一个初始矢量 $\boldsymbol{x}_0$ 开始，对其重复进行矩阵 $\boldsymbol{A}$ 的运算。每一步我们计算**瑞利比**(Rayleigh ratio)$\dfrac{\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{x}_k}=\dfrac{\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{x}_{k+1}}{\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{x}_k}$，它将收敛于矩阵的主特征值。这个直观的想法是对任意矢量应用矩阵 $\boldsymbol{A}$ 进行变换在多维度对该矢量进行缩放，但其中主特征值决定了缩放的效果。重复进行矩阵 $\boldsymbol{A}$ 运算放大了主特征值的贡献度，从而使其显示出来。

例 3.21：使用幂方法计算主特征值

请使用幂方法计算矩阵$\begin{bmatrix}4 & 6\\ 12 & 4\end{bmatrix}$的主特征值。

解：假定我们从任意初始矢量 $\boldsymbol{x}_0=\begin{bmatrix}1\\ 1\end{bmatrix}$开始，对其进行一次矩阵运算，我们得到 $\boldsymbol{x}_1=\begin{bmatrix}10\\ 16\end{bmatrix}$。瑞利比值为$\left([1\quad 1]\times\begin{bmatrix}10\\ 16\end{bmatrix}\right)\Big/\left([1\quad 1]\times\begin{bmatrix}1\\ 1\end{bmatrix}\right)=26/2=13$。重复进行以上运算，我们得到 $\boldsymbol{x}_2=\begin{bmatrix}138\\ 184\end{bmatrix}$，对应的瑞利比值为 $4304/356=12.08$。再一次迭代，我们得到 $\boldsymbol{x}_3=\begin{bmatrix}1648\\ 2368\end{bmatrix}$，瑞利比值为 $659\ 840/52\ 352=12.60$。回忆一下例 3.19 中计算出的主特征值是 12.48。我们仅用了三次迭代计算就得到了一个误差在 1%以内的近似值。

可以证明，以上计算方法的收敛速度取决于主特征值和第二大特征值之间的差距大小。两者之间差距越大，收敛得越快。直观来说，如果第二大特征值与主特征的取值相接近，其缩放效果不会轻易消失，则需要进行多次迭代计算。

幂方法也被用于计算主特征向量。为求出主特征向量，在每次迭代计算后，对矢量 $\boldsymbol{x}_i$ 进行缩放：即将矢量的最大的元素设置成 1，其余每个元素除以此最大的元素，如下例所示。

① 通常，但并非必然是主特征值。事实证明，对于大多数实际情况，我们将得到主特征值。

例 3.22：使用幂法计算主特征向量

请使用幂法计算矩阵$\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}$的主特征向量。

解：从例 3.21 可知，$x_1=\begin{bmatrix}10\\16\end{bmatrix}$。我们将矢量所有元素除以 16 得到新矢量$\begin{bmatrix}0.625\\1\end{bmatrix}$。将其作为 x_1，迭代计算得到 $x_2=\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}\times\begin{bmatrix}0.625\\1\end{bmatrix}=\begin{bmatrix}8.5\\11.5\end{bmatrix}$。再将 x_2 进行缩放得到$\begin{bmatrix}0.696\\1\end{bmatrix}$。回忆以前例子可知该矩阵特征向量是$\begin{bmatrix}1\\\sqrt{2}\end{bmatrix}$，我们可以将它缩放成$\begin{bmatrix}0.707\\1\end{bmatrix}$。同样的，经过三次迭代，我们得到的最终结果误差在 1.5%以内。

3.5.7　相似性与对角化

如果两个矩阵有相同的特征值集合，我们称这两个矩阵**相似**。给定一个矩阵 A，在某些情况下如果能够求出 A 的相似的对角矩阵 D 是很有用处的（在例 3.23 中我们展示了一个有用的例子）。

对于一个 n 阶矩阵来说，它是对角矩阵的充分但非必要条件是该矩阵有 n 个完全不同的特征值；为了验证该条件不是必要条件，考虑一个对角矩阵，其对角元素都是重复的。在这种情况下，定义矩阵 X，它的列都是矩阵 A 的特征向量。此时，通过展开以下的项，很容易看出矩阵

$$D = X^{-1}AX \tag{3.18}$$

是一个对角矩阵，其对角元素是矩阵 A 的特征值。

我们很容易求出一个对角矩阵的 m 阶幂。根据式(3.18)，请注意

$$D^2 = (X^{-1}AX)(X^{-1}AX) = X^{-1}A^2X$$

简单推导可以得到：

$$D^m = X^{-1}A^mX$$

因此有

$$A^m = XD^mX^{-1}$$

由于 D 是对角矩阵，因此等式右边比较容易计算。因此我们能够很容易计算出 A^m。

例 3.23：矩阵对角化

定义一个矩阵 $A=\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}$，从例 3.15 可知，它有两个特征值，$\lambda=4\pm6\sqrt{2}$，对应的特征向量是$\begin{bmatrix}1\\\sqrt{2}\end{bmatrix}$和$\begin{bmatrix}1\\-\sqrt{2}\end{bmatrix}$。从而我们可以得到矩阵 $X=\begin{bmatrix}1 & 1\\\sqrt{2} & -\sqrt{2}\end{bmatrix}$。从式(3.13)可知，$X^{-1}=\dfrac{1}{(-\sqrt{2}-\sqrt{2})}\begin{bmatrix}-\sqrt{2} & -1\\-\sqrt{2} & 1\end{bmatrix}$。因此，我们将矩阵 A 作如下对角化

$$\frac{1}{-2\sqrt{2}}\begin{bmatrix}-\sqrt{2} & -1\\-\sqrt{2} & 1\end{bmatrix}\begin{bmatrix}4 & 6\\12 & 4\end{bmatrix}\begin{bmatrix}1 & 1\\\sqrt{2} & -\sqrt{2}\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}\sqrt{2} & 1\\\sqrt{2} & -1\end{bmatrix}\begin{bmatrix}2 & 3\\6 & 2\end{bmatrix}\begin{bmatrix}1 & 1\\\sqrt{2} & -\sqrt{2}\end{bmatrix}$$

$$=\begin{bmatrix}4+6\sqrt{2} & 0\\ 0 & 4-6\sqrt{2}\end{bmatrix}$$

请注意矩阵 $\boldsymbol{A}$ 的对角元素正是它的特征值。

根据以上定义，我们来计算 $\boldsymbol{A}^5$，过程如下：

$$\boldsymbol{XD}^m\boldsymbol{X}^{-1}=\begin{bmatrix}1 & 1\\ \sqrt{2} & -\sqrt{2}\end{bmatrix}\begin{bmatrix}4+6\sqrt{2} & 0\\ 0 & 4-6\sqrt{2}\end{bmatrix}^5\frac{1}{-2\sqrt{2}}\begin{bmatrix}-\sqrt{2} & -1\\ -\sqrt{2} & 1\end{bmatrix}$$

根据对角矩阵的特性，矩阵的幂运算退化为一个标量的指数运算。经过简化并在计算中保留十位数字的运算精度，我们得到计算结果为

$$\begin{bmatrix}150\,783.99 & 107\,903.99\\ 215\,807.99 & 150\,783.99\end{bmatrix}$$

该值在以下真实值的舍入误差内。

$$\begin{bmatrix}150\,784 & 107\,904\\ 215\,808 & 150\,784\end{bmatrix}$$

如果矩阵有重复的特征值，则该矩阵不能被对角化。相反，我们能做的最佳处理是将其变换为**约当标准型**（Jordan canonical form），即与对角矩阵类似，对角元素都是矩阵的特征值。然而，某些紧邻且在主对角线之上的元素可能是 1。在高等线性代数课本中我们可以找到关于约当标准型运算的更多细节内容。

3.6 随机矩阵

随机矩阵，或者**马尔可夫矩阵**是一类特殊的矩阵。**右随机矩阵**是一个方阵，其元素都是非负实数，且每行元素的和为 1。**左随机矩阵**是一个方阵，其元素都是非负实数，且每列元素的和为 1。当我们提到随机矩阵时，除非有其他特别说明，通常指右随机矩阵。

在计算机网络环境下随机矩阵非常有用，因为随机矩阵 $\boldsymbol{A}$ 的每行对应于一个用来表示网络协议或者路由器的缓冲区的有限状态机，或者是一个马尔可夫链的状态。每个元素 a_{ij} 可以看作从状态 i 转变到状态 j 的概率。每行元素和为 1 的约束条件表明事实：状态转移的结果或者是保持不变，或者是转变为其他状态。我们在学习马尔可夫链、随机过程和排队论时随机矩阵会经常出现。

例 3.24：随机矩阵

矩阵 $\boldsymbol{A}=\begin{bmatrix}0.25 & 0.5 & 0.25\\ 0.1 & 0.9 & 0\\ 0 & 0 & 1.0\end{bmatrix}$，由于它是一个方阵，所有元素均为非负实数，且每行元素和为 1，因此该矩阵是一个随机矩阵。我们将第一行的数值解释为，如系统处于状态 1，经过一次转换，仍然在状态 1 的概率是 0.25，变为状态 2 的概率是 0.5，变为状态 3 的概率是 0.25。注意一旦系统进入状态 3，则它再也不会转换到其他状态，因此我们将状态 3 称为一个**吸收态**。