线性代数讲义

《线性代数》这门课程是理工科方向学生的必修课程。本课程拟采用的教材是 Howard Anton 和 Chris Rorres 所著的 Elementary Linear Algebra, 这本书内容丰富，例证详实，而且介绍了线性代数在工程领域方面的诸多应用，学生认真阅读此书是会受益匪浅的。如果非要有什么吹毛求疵的地方，那可能是此书较厚，部分学生可能存有畏惧心理，而且在短短一学期的学习过程中可能不太容易抓住学习的重点。所以本讲义旨在帮助同学们去阅读 Howard Anton 和 Chris Rorres 的这本著作，在编写过程中完全遵循Elementary Linear Algebra 这本教材，没有任何创新的地方，当然不免有画蛇添足之嫌，但如果能在学生学习线性代数的过程中有一丝帮助，那也是值得的。另外，本讲义采用中文撰写，目的是给刚进大学的学生一个过段的阶段，不过如果考虑到同学们以后的学业以及个人发展，还是强烈建议学生认真阅读英文教材。

讲义使用指南：本讲义完全按照 Elementary Linear Algebra 的章节编排，为方便学生学习使用，定理的编号和教材中的编号是一致的。因为教材 Elementary Linear Algebra 中已经有大量的习题，所以本讲义不设置习题，只在行文中会穿插一些小的练习。学生在学习过程中还是需要通过一定的习题量来巩固学习的效果，因此还是建议学生要认真仔细阅读教材，多做教材中的习题，不能完全依赖此讲义，本讲义只是帮助同学们阅读原教材的一点辅助而已，同学们切记！

1. 线性方程组和矩阵

方程的概念大家应该从中学，甚而是小学就接触过了，刚开始接触的是一元一次方程，二元一次方程（组），接着到初高中的一元二次方程等等。现实世界中事物的很多关系，通过专业人员的探索、分析、约化、归纳，总是可以写出相对应的方程（组），因此解方程（组）总是重要的，而且它也是数学领域中很多学科的核心课题。比如说，一元五次及五次以上方程的根式解问题（Galosi理论，引导了群论、数论的发展），关于多项式方程的研究（代数几何）以及在物理，工程中应用广泛的微分方程等等。本章涉猎的是线性方程组，对它的研究我们引出线性代数里面最重要的概念之一：矩阵。

1.1 线性方程组简介 回顾高中所学知识，我们知道方程

$$ax + by = c \ \ (\text{其中 } a, b \text{不都为 } 0)$$

代表了平面上的一条直线，方程

$$ax + by + cz = d \ \ (\text{其中 } a, b, c \text{不都为 } 0)$$

代表了（三维）空间中的一个平面。一般地，我们定义 n 个变量 (variable) $x_1, x_2, \ldots, x_n$ 的线性方程(linear equation)如下：

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

其中 $a_1, a_2, \ldots, a_n$ 不全为 0. 一种特别且重要的情况是 $b = 0$,此时上述方程即为

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0,$$

我们这样的方程为关于变量 $x_1, x_2, \ldots, x_n$的齐次线性方程 (homogenous linear equation).

一般地，线性方程组可以写成下面的形式：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.1}$$

如果数$s_1, s_2, \ldots s_n$分别代入未知元$x_1, x_2, \ldots x_n$所在的位置，使得方程成立的话，我们就称

$$x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$$

是方程的一个解 (solution). 为了方便，有时也将这样的解记成带序的 n 元组 (ordered n-tupe), 写成$(s_1, s_2, \ldots, s_n)$.

从几何的角度看，考虑二个变量的情形，方程组

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases} \tag{1.2}$$

的解即为两条直线的交点。而两条直线的关系，以及对应方程组解的情况有如下三种：(1)平行(无解)；(2)相交但不平行（有一个解）；(3) 重合(无穷多个解)。

考虑三个变量的情况时，方程的解对应着平面的交点。但平面之间的关系略微复杂一些，但教材中的 Figure 1.1.2 将具体图像画了出来，带给了我们直观的视觉。对于更多个变量的情况，我们已经没有更好的几何直观了（至少对绝大多数普通人是这样）。

方程组称为是相容的(consistent)，如果它含有至少一个解。称为是不相容的(inconsistent), 如果方程组没有解。但通过教材中的一些例子以及二个变量和三个变量情形下的几何直观，我们似乎发现了一条规律，方程的解只有如下三种情况：(1) 无解；(2) 有一个解；(3) 无穷多个解。具体的原因我们留待后面严谨说明(同学们亦可自己先行思考)。

我们在解一些具体的线性方程组的时候，会有这样的感觉，我们总是在对未知元前面的系数做一些加加减减，乘法和除法，化到比较简单的形式后，再做代入（本质上就是后文要论述的高斯消元法）。因此对于给定的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.3}$$

我们将系数写成下面的长方形的数阵形式

$$\left(\begin{array}{cccc:c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right),$$

这称为线性方程组(1.3) 的增广矩阵 (augmented matrix). 在解线性方程组的时候，我们有三种初等的操作：

(1) 在某个方程上乘以非零常数。

(2) 交换两个方程的位置。

(3) 将某个方程的常数倍加到另外一个方程上。

很容易注意到这三种初等操作都是可逆的操作，因此经过这三种初等操作得到的方程和原来的方程的解是一致的。这三种初等操作可以自然地过渡到对相应的增广矩阵的操作，我们称为初等行变换 (elementary row operations), 即：

(1) 增广矩阵的某一行乘以非零常数。

(2) 交换增广矩阵的两行。

(3) 将增广矩阵某一行的常数倍加到另外一行上。

有了这些初等行变换，我们可以比较方便地介绍高斯消元法。

1.2 高斯消元法 (Gauss Elimination)

高斯消元法是比较规整，程序化地解线性方程组的方法。主要想法是通过前面引入的初等行变化将线性方程组对应的增广矩阵转换成比较好求解的形式，即行阶梯型 (row echelon form). 如果一个矩阵满足下面的条件 (1)-(3):

(1) 如果某一行不全为 0, 那么其第一个非零元为 1, 这样的 1 称为首 1(leading 1).

(2) 如果某些行全为 0, 则它们都位于矩阵的底部。

(3) 如果相连的两行都不全为 0, 则在下面一行出现的首 1 位于上一行首 1 的右侧。

我们就称此矩阵有行阶梯型 (row echelon form). 更多地，如果还满足条件：

(4) 如果某一列包含首 1, 则此列的其余元素都为 0.

我们就称矩阵具有简约行阶梯型(reduced row echelon form)。具体的例子可参见教材中§1.2的例 1和例 2.

一旦线性方程组的增广矩阵具有简约行阶梯型，那么线性方程组的求解是很容易的。只简单举一例，比如增广矩阵是

$$\left(\begin{array}{ccc:c} 1 & 0 & 2 & 5 \\ 0 & 1 & -3 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

相应的方程的解为 $x = 5 - 2z$, $y = 4 + 3z$. 这里的 x, y 是首变量 (leading variables), z 是自由变量 (free variables), z 取任意的值都确定了 x, y 的值，从而最终确定了方程的一个解。

下面我们论述如何将给定的增广矩阵转化成简约行阶梯型的步骤，即著名的高斯-约当消元法。我们考虑这样的例子来帮助我们理解，增广矩阵是

$$\left(\begin{array}{ccccc:c} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{array}\right).$$

步骤一：定位最左边非零元的列。例子中，上述矩阵中的最左列不全为零。

步骤二：对于刚刚步骤一中选定的列，如果其第一行的元素为 0, 我们总可以交换第一行和其他某一行使得第一行的元素不为 0. 在例子中，我们交换矩阵的第一行和第二行，有

$$\left(\begin{array}{ccccc:c} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{array}\right).$$

步骤三：我们对于第一行乘以某个常数，使得第一个非零元变成 1，即得到了首 1. 在例子中，我们将第一行乘以 $\frac{1}{2}$, 得到：

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{array}\right).$$

步骤四：将第一行的某个倍数加到其他行，使得第一行的首 1下面的元素都变成 0. 在例子中，我们将第一行乘以 -2加到第三行上，得到：

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{array}\right).$$

步骤五：现在，我们保持第一行不动，我们对剩下的子矩阵进行从步骤一开始的操作。在考虑的例子中，首先定位最左边的非零列，将第二行乘以$-\frac{1}{2}$, 得到首 1, 矩阵变成

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{array}\right).$$

第二行的 -5 倍加到第三行，矩阵变成

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{array}\right).$$

从而前面两行已经处理好了，我们保持不动，考虑第三行，只需要第三行乘以 2, 得到一个首 1, 矩阵变成

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right).$$

到此为止，我们得到了一个行阶梯型的增广矩阵。

步骤六：我们从最下面的非零行开始，乘以适当的倍数加到它上面的行上，使得每一个首 1上面的元素都为 0. 在例子中，将第三行的 $\frac{7}{2}$ 加到第二行，将第三行的 -6 倍加到第一行上，得到矩阵

$$\left(\begin{array}{ccccc:c} 1 & 2 & -5 & 3 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right).$$

将第二行的 5 倍加到第一行上，得到矩阵

$$\left(\begin{array}{ccccc:c} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right).$$

现在，我们回头来看一个线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.4}$$

经过高斯－约当消元法，可以变成如下的形式：

$$
(1.5)\qquad \begin{cases}
x_{k_1} \qquad\qquad\qquad + \sum(\ ) = b'_1 \\
\qquad x_{k_2} \qquad\qquad + \sum(\ ) = b'_2 \\
\qquad\qquad \cdots \quad \cdots \quad \cdots \quad \cdots \\
\qquad\qquad\quad x_{k_r} \quad + \sum(\ ) = b'_r \\
\qquad\qquad\qquad\qquad\quad 0 = b'_{r+1} \\
\qquad\qquad\qquad\qquad\qquad \cdots\cdots \\
\qquad\qquad\qquad\qquad\quad 0 = b'_m
\end{cases}
$$

根据简约的行阶梯型，我们很自然地有如下分析：

(1) 如果$b'_{r+1},\dots,b'_m$ 不全为零，则方程组无解；

(2) 如果$b'_{r+1},\dots,b'_m$ 全为零，且此时$r=n$, 方程只有一个解。

(3) 如果$b'_{r+1},\dots,b'_m$ 全为零，且此时$r<n$, 此时方程组有自由变量，方程组有无穷多个解。

在线性方程组(1.4)中，当$b_1=b_2=\cdots=b_m=0$ 时，我们得到一类特殊且重要的线性方程组，称为齐次线性方程组(homogenous linear systems). 省略到后面全为0的行，我们也可以得到它的简约的行阶梯型：

$$
(1.6)\qquad \begin{cases}
x_{k_1} \qquad\qquad\qquad + \sum(\ ) = 0 \\
\qquad x_{k_2} \qquad\qquad + \sum(\ ) = 0 \\
\qquad\qquad \cdots \quad \cdots \quad \cdots \quad \cdots \\
\qquad\qquad\quad x_{k_r} \quad + \sum(\ ) = 0
\end{cases}
$$

此时$x_1=0,x_2=0,\dots,x_n=0$总是方程的一个解，称为平凡解(trivial solution). 而且根据之前的分析，我们知道此时，方程组的解只有两种情况：

(1)当$r=n$时，方程组只有一个解，即平凡解；

(2)当$r<n$ 时，方程组有$n-r$ 个自由变量，因此方程组有无穷多个解，这时候非零的解我们称为非平凡解(nontrivial solutions).

根据(2), 我们得到下述推论：在齐次线性方程组中，如果未知元的个数比方程的个数多的话，那么这个齐次线性方程组总是有无穷多个解。

注: 在前面的论述中，我们都是用高斯-约当消元法将增广矩阵化成简约行阶梯型。但当用计算机处理庞大的线性方程组时，高斯消元法更有效率，即将增广矩阵化成行阶梯型，先一步步地从下面的方程开始解起，逐步代入上面的方程中，最终得到方程组的解。具体的算法和例子参见教材中§1.2节的Gaussian Elimination and Back-Substitution, 这里不再赘述。

最后，我们不加证明地提出下面三个事实：

(1) 每一个矩阵都有唯一的简约行阶梯型。

(2) 行阶梯型并不是唯一的。不同的初等行变换可能得到不同的行阶梯型。

(3) 虽然行阶梯型并不是唯一的，但是它们零行的个数相同，且首1都处于相同的位置。

1.3 矩阵和矩阵运算

矩阵(matrix)就是长方形的数阵，其中的数称为矩阵的项(entry). 一般地，一个$m \times n$的矩阵可以写成

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

也可以简记为$(a_{ij})_{m\times n}$ 或者(a_{ij}), 前一种记法强调使用情景下的行列数。矩阵 A的第i行和第j列记为$(A)_{ij}$, 在上述矩阵中，我们有$(A)_{ij} = a_{ij}$. 当$m = 1$ 或者$n = 1$时，我们得到一些特殊的矩阵, 行向量 (row vector) 记成

$$\mathbf{a} = (a_1, a_2, \dots, a_n)$$

以及列向量 (column vector) 记成

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}.$$

还有一类非常重要的情形，当$m = n$ 时，A 是一个n阶的方阵 (square matrix of order n),

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

此时，$a_{11}, a_{22}, \dots, a_{nn}$ 形成 A的主对角线 (main diagonal)元素。

矩阵的加法：对于两个矩阵$A = (a_{ij})_{m\times n}$和$B = (b_{ij})_{m\times n}$, 定义矩阵的加法如下：

$$(A + B)_{ij} = (A)_{ij} + (B)_{ij} = a_{ij} + b_{ij}.$$

这里注意，定义矩阵的加法时，两个矩阵必须具有相同的行数和列数，即具有同样的尺寸。同样地，可以定义矩阵的减法：

$$(A - B)_{ij} = (A)_{ij} - (B)_{ij} = a_{ij} - b_{ij}.$$

矩阵的纯量乘积：对于一个矩阵 $A = (a_{ij})$ 以及纯量 c. 我们定义矩阵的纯量乘法 (scalar multiple) 为：

$$(cA)_{ij} = c(A)_{ij} = ca_{ij}.$$

矩阵的乘法：对于矩阵 $A=(a_{ij})_{m\times r}$ 和矩阵 $B=(b_{ij})_{r\times n}$, 我们定义矩阵的乘法 AB 如下：

$$(AB)_{ij}=\sum_{k=1}^{r}a_{ik}b_{kj}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{ir}b_{rj}.$$

这里必须强调的是，只有当A的列数和B的行数相等时，才可以定义上述的乘法。刚接触矩阵乘法的初学者可能觉得这种乘法定义不太自然，但是在以后有了很多的准备后，我们会详细解释为什么要这么定义矩阵的乘法。另外，这样定义的乘法确实是合理的，有着其广泛应用的，初学者先接受如此定义的乘法。

练习：(1) 对于矩阵$A_{m\times r}$, $B_{r\times s}$ 和$C_{s\times n}$，验证 $(AB)C=A(BC)$.

(2) 对于矩阵$A_{m\times r}$，$B_{m\times r}$ 和矩阵 $C_{r\times n}$, 验证 $(A+B)C=AC+BC$.

分块矩阵(partitioned matrices)：我们可以在矩阵的行之间插入一些水平线，在列之间插入一些竖直线，将矩阵划分成一些小矩阵。即对于矩阵$A=(a_{ij})_{m\times n}$, 我们可以有它的分块的形式：

$$\left(\begin{array}{c:c:c:c} A_{11} & A_{12} & \dots & A_{1s} \\ \hdashline A_{21} & A_{22} & \dots & A_{2s} \\ \hdashline \dots & \dots & \dots & \dots \\ \hdashline A_{r1} & A_{r2} & \dots & A_{rs} \end{array}\right),$$

这里的A_{kl} $(1\le k\le r, 1\le l\le s)$ 都是A的子矩阵(submatrix). 比较特殊的两种分块形式：一种是将矩阵 A 写成

$$A=\begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dots \\ \mathbf{r}_m \end{pmatrix},$$

其中 $\mathbf{r}_k$ $(1\le k\le m)$是矩阵的行向量。另外一种是将矩阵 A 写成

$$A=\begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{pmatrix},$$

其中 $\mathbf{c}_l$ $(1\le l\le n)$是矩阵的列向量。有了这种分块矩阵的定义，我们可以将矩阵$A_{m\times r}$ 和$B_{r\times n}$的乘法写成

$$AB=A\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{pmatrix}=\begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{pmatrix},$$

以及

$$AB=\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_m \end{pmatrix}B=\begin{pmatrix} \mathbf{a}_1B \\ \mathbf{a}_2B \\ \dots \\ \mathbf{a}_mB \end{pmatrix}.$$

思考：对于更加一般的分块矩阵，我们如何合理定义它们的乘法？

矩阵的线性组合(linear combination):对于相同型（一样多的行数和列数）的矩阵$A_1, A_2, \ldots, A_r$. 设$c_1, c_2, \ldots, c_r$是纯量，则

$$c_1 A_1 + c_2 A_2 + \cdots + c_r A_r$$

称为矩阵$A_1, A_2, \ldots, A_r$的线性组合，其中$c_1, c_2, \ldots, c_r$ 是系数(coefficient). 设

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \text{以及列向量 } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

于是此时$A\mathbf{x}$可以定义，其结果为

$$A\mathbf{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{pmatrix},$$

也就是说$A\mathbf{x}$ 可以写成矩阵A列向量的线性组合，其系数是$\mathbf{x}$中的项。

注：如果对分块矩阵的乘法有所思考和感觉的话，我们可以直接有

$$A\mathbf{x} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

对于之前考虑的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.7}$$

在有了矩阵的乘法的概念之后，我们可以将上面的方程组写成：

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix}.$$

如果将上面的矩阵分别记成$A, \mathbf{x}, \mathbf{b}$ 的话，我们直接有关于矩阵的等式$A\mathbf{x} = \mathbf{b}$. 这里的矩阵A是系数矩阵(coefficient matrix), 和之前一样，增广矩阵(augmented matrix) 可以记成

$$\left(\begin{array}{c|c} A & \mathbf{b} \end{array}\right) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right),$$

矩阵的转置(transpose): 对于矩阵A, 将矩阵进行翻转，将原矩阵A的行变为列，列变为行，得到新的矩阵，记为A^T, 称为矩阵A的转置。严格地讲，设A 是一个$m \times n$的矩阵，则A^T是一个$n \times m$ 的矩阵，满足：$(A^T)_{ij} = (A)_{ji}$.

练习：证明：(1) $(A^T)^T = A$. (2) $(A+B)^T = A^T + B^T$. (3) $(\lambda A)^T = \lambda A^T$. (4)$(AB)^T = B^T A^T$.

矩阵的迹(trace): 设矩阵$A = (a_{ij})_{n\times n}$ 是一个方阵，矩阵A的迹是A的主对角线上的元素之和，记成 $\mathrm{tr}(A)$, 即有 $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

练习：设有两个矩阵，分别记成 $A_{m\times n}$ 和$B_{n\times m}$, 于是AB和BA 都是方阵，从而可以定义它们的迹。证明：$\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

1.4 矩阵的基本性质和矩阵的逆

矩阵的加法和乘法满足下面的性质：

(1) （加法交换律）$A + B = B + A$.

(2) （加法结合律）$(A + B) + C = A + (B + C)$.

(3) （乘法结合律）$(AB)C = A(BC)$.

(4) （矩阵乘法的分配律）$(A + B)C = AC + BC$, $A(B + C) = AB + AC$.

(5) （纯量乘法的分配律）$\lambda(A + B) = \lambda A + \lambda B$, $(\lambda + \mu)A = \lambda A + \mu A$.

更多的性质可以参照教材§1.4中的定理1.4.1, 上述性质的证明的思路是简单的，就是比较等式两边矩阵的第i行和第j列的项是否相等，可以参见教材中的证明，这里不再赘述。

重点注意：矩阵的乘法不满足交换律！主要可能出现如下三点原因：

(1) 矩阵$A_{m\times r}$ 和$B_{r\times n}$，乘法AB可以定义，但是BA不可以定义。

(2) 矩阵$A_{m\times n}$ 和$B_{n\times m}$，乘法AB和BA都是可以定义的，但是AB和BA的型不一样，因而不可以比较是否相等。

(3) 矩阵$A_{n\times n}$ 和$B_{n\times n}$ 都是n 阶方阵，但是$AB \neq BA$. 举一个简单的例子：

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad 和 \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

容易计算得到：

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad 以及 \quad BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

两者是不相等的。

零矩阵 (zero matrix)：对于矩阵$A=(a_{ij})_{m\times n}$，如果每一项$a_{ij}=0$, 则我们称 A为零矩阵，记成 0 或者 $0_{m\times n}$. 零矩阵具有很多简单易证的性质：(1)$A+0=0+A=A$; (2) $A-A=A+(-A)=0$; (3) $0A=0$; (4) 对于纯量乘法，如果 $cA=0$, 则$c=0$或者$A=0$.
注意：(1) 对于一般的矩阵乘法$AB=AC$且$A\neq 0$, 我们并不能得到$B=C$. （参见教材中§1.4的例3.）
(2) 如果$AB=0$, 我们并不能得到$A=0$ 或者$B=0$. （参见教材中§1.4的例4.）

单位矩阵 (identity matrix)：对于n 阶方阵$A=(a_{ij})_{n\times n}$，如果它的主对角线元素全为1, 而其余的元素全为 0, 即有$a_{kk}=1$ $(1\leq k\leq n)$ 以及$a_{ij}=0$ $(i\neq j)$. 我们称这样的矩阵为单位矩阵，记为 I_n. 我们很容易验证：对于$m\times n$的矩阵 A, 有：

$$AI_n=A \ \text{ 以及 } \ I_mA=A.$$

定理 1.4.3: 对于一个n阶方阵A，记R是其简约行阶梯型，则R有零行或者R是单位矩阵I_n.
证明：设方阵$R=(r_{ij})_{n\times n}$ 是A的简约行阶梯型，则其最后一行要么为零行，要么非零，如果有零行则命题得证。如果最后一行非零，必然有首1. 观察简约行阶梯型的形式，我们知道下方的首1 必然在上方首1 的右边，而R 是n 阶方阵，因此首1 必然在主对角线上，于是R 就是单位矩阵I_n，证毕。 □

矩阵的逆(inverse): 设A是一个 n 阶方阵，如果存在 n 阶方阵 B 使得 $AB=BA=I_n$. 我们就称 A 是可逆的(invertible) 或者非奇异的(nonsingular), 这时称 B 是 A 的逆(inverse)矩阵。如果满足条件 $AB=BA=I_n$ 的矩阵 B 不存在，我们就称 A 是奇异的(singular). 如果方阵A 是方阵B的逆，则显然B也是A的逆，于是称它们互为逆矩阵。下面的定理告诉我们，矩阵如果有逆，那么它的逆是唯一的。
定理 1.4.4: 如果B和C 都是A 的逆矩阵，则$B=C$.
证明：因为$AB=I$, 在矩阵两边左乘C, 有$CAB=C$. 但注意到C 是A 的逆矩阵，于是我们有$CA=I$, 从而得到$B=C$. □

根据此定理，我们可以将可逆方阵A的逆记成A^{-1}. 于是有：

$$AA^{-1}=A^{-1}A=I$$

矩阵的逆非常重要，一个很显然的观察是，对于写成矩阵形式$A\mathbf{x}=\mathbf{b}$ 的线性方程组，如果A 是方阵，且是可逆的，那么在方程两边左乘矩阵A^{-1}, 我们就可以得到方程的解$\mathbf{x}=A^{-1}\mathbf{b}$，而且据此知方程的解是唯一的。

对于2 阶的方阵，我们求矩阵的逆，下面的定理可以作为一个很好的练习。
定理 1.4.5: 2 阶的方阵

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

可逆当且仅当$ad-bc\neq 0$, 此时A的逆矩阵是

$$A^{-1}=\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

注: 这里的 $ad-bc$ 是 2 阶方阵的行列式 (determinant), 记为 $\det(A)=ad-bc$. 上述定理告诉我们 A 可逆当且仅当行列式不为 0. 关于行列式的一般理论，我们留待第二章学习。

定理 1.4.6: 设A和B是两个同型的可逆方阵，则AB也是可逆的，且知道$(AB)^{-1}=B^{-1}A^{-1}$.

证明：根据定义，我们验证

$$(AB)(B^{-1}A^{-1})=(B^{-1}A^{-1})(AB)=I$$

即可。由矩阵乘法的结合律，有

$$(AB)(B^{-1}A^{-1})=A(BB^{-1})A^{-1}=AIA^{-1}=AA^{-1}=I.$$

类似地，可知$(B^{-1}A^{-1})(AB)=I$. 定理得证。 □

下面我们来定义矩阵的幂(powers of matrix): 对于方阵A, 我们定义$A^0=I$. 对于正整数n, 我们定义

$$A^n=AA\dots A\ \ (n\text{个方阵}A\text{相乘}),$$

如果方阵A是可逆的，我们可以定义A的负整数幂次：

$$A^{-n}=(A^{-1})^n=A^{-1}A^{-1}\dots A^{-1}\ \ (n\text{个方阵}A^{-1}\text{相乘}).$$

根据上述定义，我们很容易验证

$$A^rA^s=A^{r+s}\ \ \text{以及}\ \ (A^r)^s=A^{rs}.$$

下面的定理根据矩阵逆的定义，也是比较容易得到的，证明可以作为一个练习。

定理 1.4.7: 设A是一个可逆方阵，n 是一个正整数，则

(a) A^{-1} 是可逆方阵，且$(A^{-1})^{-1}=A$.

(b) A^n 是可逆方阵，且$(A^n)^{-1}=A^{-n}=(A^{-1})^n$.

(c) 对于非零纯量λ, λA 是可逆方阵，且$(\lambda A)^{-1}=\lambda^{-1}A^{-1}$.

矩阵多项式 (matrix polynomial): 给定一个多项式

$$p(x)=a_0+a_1x+a_2x^2+\cdots+a_mx^m.$$

对于一个方阵A, 我们可以考虑关于 A的矩阵多项式

$$p(A)=a_0I+a_1A+a_2A^2+\cdots+a_mA^m,$$

即将变量x替换成矩阵A, 将常数项a_0替换成a_0I.

矩阵的转置和矩阵的逆有如下的关系：

定理 1.4.9: 设A是可逆方阵，则A^T也是可逆的，且有$(A^T)^{-1} = (A^{-1})^T$.

证明：根据矩阵逆的定义，我们只要验证

$$A^T(A^{-1})^T = (A^{-1})^T A^T = I.$$

而 $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$ 以及 $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$, 定理得证。 □

1.5 初等矩阵以及求矩阵逆的方法

在§1.1中我们介绍了对矩阵 A的三种初等行变换：

(1) 矩阵的某一行乘以非零常数c.

(2) 交换矩阵的 r_1 和 r_2两行。

(3) 将矩阵的r_1行的 c 倍加到r_2 行上。

如果矩阵B可以由矩阵A的某一种初等行变换得到，那么A也可以由B的初等行变换得到，原因是上面所述三种初等行变换都是可逆的操作：

(1) 在矩阵的同一行乘以常数$\frac{1}{c}$.

(2) 再次交换矩阵的 r_1 和 r_2两行。

(3) 将矩阵的r_1行的 $-c$ 倍加到r_2 行上。

注意到这个现象，我们称矩阵 A 和矩阵 B是行等价的（row equivalent），如果 A 可以通过若干初等行变换变成 B (同样地，矩阵B也可以通过若干初等行变换变成A).

初等矩阵(elementary matrix): 一个n阶方阵被称为初等矩阵，如果它是由单位矩阵I_n经过一个初等行变换所得。我们可以写出初等矩阵的具体形式：

(1)某一行乘以常数c所得的初等矩阵

$$\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \cdots & & & \\ & & & c & & \\ & & & & \cdots & \\ & & & & & 1 \end{pmatrix}.$$

(2)交换某两行所得的初等矩阵

$$\begin{pmatrix} 1 & & & & & & \\ & \ldots & & & & & \\ & & & & 1 & & \\ & & & \ldots & & & \\ & & 1 & & & & \\ & & & & & \ldots & \\ & & & & & & 1 \end{pmatrix}.$$

(3)某一行的c倍加到另外一行上所得的初等矩阵

$$\begin{pmatrix} 1 & & & & & & \\ & \ldots & & & & & \\ & & 1 & & c & & \\ & & & \ldots & & & \\ & & & & 1 & & \\ & & & & & \ldots & \\ & & & & & & 1 \end{pmatrix}.$$

我们观察到很重要的一点：对于矩阵$A_{m\times n}$ 做初等行变换，即左乘相应的初等矩阵即可。另外，我们在前文中也提到三种行变换都是可逆的操作，因此对应的初等矩阵也是可逆的，我们可以具体讲三种初等矩阵的逆的形式写出来。

(1)

$$\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ldots & & & \\ & & & c & & \\ & & & & \ldots & \\ & & & & & 1 \end{pmatrix} \text{的逆矩阵是} \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ldots & & & \\ & & & \frac{1}{c} & & \\ & & & & \ldots & \\ & & & & & 1 \end{pmatrix}.$$

(2)

$$\begin{pmatrix} 1 & & & & & & \\ & \ldots & & & & & \\ & & & & 1 & & \\ & & & \ldots & & & \\ & & 1 & & & & \\ & & & & & \ldots & \\ & & & & & & 1 \end{pmatrix} \text{的逆矩阵是它自己} \begin{pmatrix} 1 & & & & & & \\ & \ldots & & & & & \\ & & & & 1 & & \\ & & & \ldots & & & \\ & & 1 & & & & \\ & & & & & \ldots & \\ & & & & & & 1 \end{pmatrix}.$$

(3)

$$\begin{pmatrix} 1 & & & & & & \\ & \cdots & & & & & \\ & & 1 & & c & & \\ & & & \cdots & & & \\ & & & & 1 & & \\ & & & & & \cdots & \\ & & & & & & 1 \end{pmatrix} \text{的逆矩阵是} \begin{pmatrix} 1 & & & & & & \\ & \cdots & & & & & \\ & & 1 & & -c & & \\ & & & \cdots & & & \\ & & & & 1 & & \\ & & & & & \cdots & \\ & & & & & & 1 \end{pmatrix}.$$

综上，我们得到

定理 1.5.2: 初等矩阵总是可逆的，它的逆也是初等矩阵。

接下来我们所述的定理很重要，它将我们之前的很多概念，包括齐次线性方程组，矩阵，简约行阶梯型，初等矩阵，矩阵的逆等等概念联系在一起。

定理 1.5.3: 设A是一个n阶方阵，那么下面的命题是等价的：

(a) A 是可逆方阵。

(b) 齐次线性方程组$A\mathbf{x}=0$只有平凡解。

(c) 方阵A的简约行阶梯型是I_n.

(d) 方阵A可以写成一些初等矩阵的乘积。

证明：我们证明$(a)\Rightarrow(b)\Rightarrow(c)\Rightarrow(d)\Rightarrow(a)$, 这样就形成了命题的一个循环，于是命题彼此之间就是等价的。

$(a)\Rightarrow(b)$ 当A可逆时，在方程$A\mathbf{x}=0$两边同时左乘矩阵A^{-1},即得$\mathbf{x}=0$.

$(b)\Rightarrow(c)$ 考虑方程$A\mathbf{x}=0$的增广矩阵

$$\left(\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 \end{array}\right),$$

因为方程只有平凡解，因此经过高斯－约当消元法操作后，最终矩阵将变成

$$\left(\begin{array}{cccc:c} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \cdots & & \\ & & & 1 & 0 \end{array}\right),$$

这也就意味着A的简约行阶梯型只能是I_n.

$(c) \Rightarrow (d)$ 方阵A的简约行阶梯型是I_n, 也就是说A经过一些初等行变换之后，可以变成单位矩阵I_n. 而我们知道做初等行变换的结果和左乘相应的初等矩阵的结果是一样的。这也就意味着存在一些初等矩阵$E_1, E_2, \ldots, E_k$使得

$$E_k \ldots E_2 E_1 A = I_n.$$

而初等矩阵都是可逆的，所以我们按照顺序依次左乘$E_k^{-1}, \ldots, E_2^{-1}, E_1^{-1}$, 得到：

$$A = E_1^{-1} E_2^{-1} \ldots E_k^{-1} I_n = E_1^{-1} E_2^{-1} \ldots E_k^{-1},$$

我们就将矩阵A写成了一些初等矩阵的乘积。

$(d) \Rightarrow (a)$ 可逆矩阵的乘积还是可逆的，又因为初等矩阵是可逆的，所以当A可以写成一些初等矩阵的乘积的时候，我们就知道A是可逆的。 □

上述定理给我们提供了一种算法来求解矩阵的逆。设矩阵A是可逆的，则它经过一些初等行变换，即左乘矩阵$E_1, E_2, \ldots, E_k$变成单位矩阵I_n. 此时，矩阵A的逆可以写成

$$A^{-1} = E_k \ldots E_2 E_1 I_n,$$

这意味着A^{-1}是由单位矩阵I_n 经过同样顺序的初等行变换得到的。因此我们给出一种算法来求解给定矩阵A 的逆。考虑一种新的矩阵

$$\left(\begin{array}{c:c} A & I_n \end{array} \right),$$

如果我们能够通过一些初等行变换使得左侧矩阵变成单位矩阵，则自然地右侧矩阵就变成了矩阵A^{-1},即有

$$\left(\begin{array}{c:c} I_n & A^{-1} \end{array} \right).$$

教材中§1.5中的例4详细演示了此种算法，而且这种算法还可以帮助我们判定一个矩阵到底是不是可逆的，见教材中§1.5中的例5.

1.6 线性方程组和矩阵的逆

在前面讲高斯消元法的时候，我们已经提到过一个线性方程组$A\mathbf{x} = \mathbf{b}$ 的解只有三种情况：(1)无解；(2) 一个解；(3) 无穷多个解。这里我们另外一种证明。

证明：方程组无解，或者有一个解的情况，我们都有很多例子知道了。因此我们只需要说明当解的个数大于一个的时候，方程组有无穷多个解即可。假设$\mathbf{x}_1, \mathbf{x}_2$ 是方程组的两个解，我们令$\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2$. 这时

$$A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = 0$$

从而对于任意的常数λ,

$$A(\mathbf{x}_1 + \lambda\mathbf{x}_0) = A\mathbf{x}_1 + \lambda A\mathbf{x}_0 = A\mathbf{x}_1 = \mathbf{b}.$$

于是对于任意的λ, $x_1 + \lambda\mathbf{x}_0$)都是方程的解，而$\lambda$的选取有无穷多个，因此我们知道方程组$A\mathbf{x} = \mathbf{b}$ 有无穷多个解。

在前面我们定义矩阵A的逆的时候，我们需要找到矩阵B使得

$$AB = I \quad \text{和} \quad BA = I$$

两个条件同时满足。实际上下面的定理告诉我们，只需要两个条件中的任意其中一个，我们就可以定义矩阵的逆。

定理 1.6.3: 设A是一个方阵，则

(a) 如果方阵B 满足$BA = I$, 则A可逆，此时$B = A^{-1}$.

(b) 如果方阵B 满足$AB = I$, 则A可逆，此时$B = A^{-1}$.

证明：(a) 要证明A可逆，我们只需要说明方程$A\mathbf{x} = 0$只有平凡解。现在设$\mathbf{x}_0$是方程的一个解，即$A\mathbf{x}_0 = 0$. 方程两边同时乘以矩阵B, 得到$BA\mathbf{x}_0 = I\mathbf{x}_0 = \mathbf{x}_0 = 0$, 这就说明方程只有零解，于是$A$可逆。从而$B = A^{-1}$.

(b) 根据(a)的结果，我们知道B可逆，且此时$A = B^{-1}$. 因为可逆矩阵的逆仍然是可逆的，从而A可逆，于是$B = A^{-1}$. □

等价性定理：

定理 1.6.4: 设A是一个n阶方阵，那么下面的命题是等价的：

(a) A 是可逆方阵。

(b) 齐次线性方程组$A\mathbf{x} = 0$只有平凡解。

(c) 方阵A的简约行阶梯型是I_n.

(d) 方阵A可以写成一些初等矩阵的乘积。

(e) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$都是相容的(consistent), 即方程总是有解。

(f) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$只有一个解。

证明：我们在之前定理 1.5.3 中已经证明了 $(a), (b), (c), (d)$ 这四个条件是等价的。现在我们证明 $(a) \Rightarrow (f) \Rightarrow (e) \Rightarrow (a)$.

$(a) \Rightarrow (f)$ 现在假设A可逆，那么对于方程$A\mathbf{x} = \mathbf{b}$, 它的解都是$\mathbf{x} = A^{-1}\mathbf{b}$, 只有一个。

$(f) \Rightarrow (e)$ 显然。因为对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$ 有一个解，那么方程总是相容的。

$(e) \Rightarrow (a)$ 我们选取特殊的列向量$\mathbf{b}$, 我们令$\mathbf{b}_i$ 是第i行为1,其余地方为0的列向量。根据条件(e), 方程组$A\mathbf{x} = \mathbf{b}_i$ 都是相容的，记它的一个解为$\mathbf{x}_i$. 我们令方阵C 为

$$C = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{pmatrix},$$

于是根据矩阵分块的乘法我们有

$$AC = \begin{pmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \ldots & A\mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{pmatrix} = I_n$$

而这就说明了A是可逆矩阵。 □

之前我们提到过，当矩阵A 和B都可逆的时候，AB 都是可逆的。下面的定理告诉我们其逆命题也成立。

定理 1.6.5: 设A和B是同型的方阵。如果AB是可逆的，则A和B都是可逆的。

证明：假设A是奇异的(不可逆)，那么根据上述等价性定理，存在列向量$\mathbf{b}$, 使得方程$A\mathbf{y}=\mathbf{b}$不相容，即方程没有解。此时我们知道$AB\mathbf{x}=\mathbf{b}$没有解，这与AB可逆是矛盾的。

现在假设A是可逆的，但是B是奇异的。根据上述等价性定理，存在列向量$\mathbf{b}$, 使得$B\mathbf{x}=\mathbf{b}$ 没有解，因为A可逆，这等价于方程 $AB\mathbf{x}=A\mathbf{b}$ 没有解，这与AB可逆是矛盾的。从而最终我们得到了定理的证明。 □

1.7 对角矩阵，三角矩阵和对称矩阵

对角矩阵：一个方阵，如果除了对角线的元素外，其余的元素都为0,则称此方阵是对角矩阵。一般地，我们可以一个n 阶对角方阵写成

$$D=\begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

对角矩阵可逆当且仅当其主对角线的元素全不为0, 此时，它的逆是

$$D^{-1}=\begin{pmatrix} \frac{1}{d_1} & 0 & \dots & 0 \\ 0 & \frac{1}{d_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{d_n} \end{pmatrix}.$$

另外，对角矩阵的幂次也是容易求得的，对于正整数k, 我们有

$$D^k=\begin{pmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_n^k \end{pmatrix}.$$

三角矩阵(triangular matrix)：如果一个方阵的主对角线上面的元素都为0,则称此方阵是下三角矩阵 (lower triangular matrix)。类似地，如果一个方阵的主对角线下面的元素都为0,则称此方阵是上三角矩阵 (upper triangular matrix)。上三角矩阵和下三角矩阵统称为三角矩阵。我们可以更严格地用记号来表达，对于方阵$A=(a_{ij})_{n\times n}$. 如果对于$i<j$, 都有$a_{ij}=0$, 则矩阵A是下三角矩阵；如果对于$i>j$, 都有$a_{ij}=0$, 则矩阵A 是上三角矩阵。

定理 1.7.1:

(a) 上三角矩阵的转置是下三角矩阵，下三角矩阵的转置是上三角矩阵。

(b) 上三角矩阵的乘积是上三角矩阵，下三角矩阵的乘积是下三角矩阵。

(c) 三角矩阵可逆当且仅当主对角线上的元素全不为零。

(d) 可逆上三角矩阵的逆还是上三角矩阵，可逆下三角矩阵的逆还是下三角矩阵。

证明：(a) 和(b) 的证明是比较容易的，留待练习。(c) 和(d)的证明我们留待下一章引入行列式的工具后，再来给出证明。 □

对称矩阵(symmetric matrix)：一个方阵A称为是对称的，如果$A = A^T$. 具体地，对于矩阵$A = (a_{ij})_{n\times n}$, A 对称当且仅当有$a_{ij} = a_{ji}$.

当A和B都是同尺寸的对称矩阵，且λ是纯量时，则下面的性质都是容易验证的：

(1) A^T是对称的。

(2) $A + B$ 和$A - B$ 都是对称的。

(3) λA 是对称的。

一般而言，两个对称矩阵的乘积不一定还是对称矩阵。下面的定理告诉我们

定理 1.7.3: 两个对称矩阵的乘积还是对称矩阵当且仅当这两个矩阵彼此交换。

证明：设A和B 是两个对称矩阵，于是$(AB)^T = B^TA^T = BA$. 所以$(AB)^T = AB$ 当且仅当$AB = BA$, 命题得证。 □

一般而言，一个对称的矩阵不一定是可逆的，但如果它可逆，则它的逆矩阵也是对称的。即有

定理 1.7.4: 设A是一个可逆对称矩阵，则A^{-1}也是对称的。

证明：由定理1.4.9, 我们知道$(A^{-1})^T = (A^T)^{-1}$, 又注意到A本身是对称的，于是有

$$(A^{-1})^T = (A^T)^{-1} = A^{-1},$$

这说明A^{-1}是对称的。 □

设 A 是一个$m \times n$的矩阵，则A^T是一个$n \times m$的矩阵。从而AA^T是一个m阶方阵，A^TA是一个n 阶方阵，而且

$$(AA^T)^T = (A^T)^TA^T = AA^T \quad \text{和} \quad (A^TA)^T = A^T(A^T)^T = A^TA$$

成立。这说明AA^T是一个对称的m阶方阵，A^TA 是一个对称的n 阶方阵。

定理 1.7.5: 设A是一个可逆方阵，则AA^T和A^TA也是可逆的。

证明：因为A可逆，由定理1.4.9知A^T也是可逆的。因为可逆矩阵的乘积仍然是可逆的，因此AA^T和A^TA也是可逆的。 □

斜对称矩阵(skew-symmetric matrix)：一个方阵A称为是斜对称的，如果$A^T = -A$. 具体地，对于矩阵$A = (a_{ij})_{n\times n}$, A斜对称当且仅当有$a_{ij} = -a_{ji}$ 且$a_{ii} = 0$. 类似于对称矩阵，我们知道：

(a) 如果一个方阵A是斜对称的，则A^T, $A + B$, $A - B$, λA, A^{-1}也都是斜对称的。

(b)任意一个方阵可以写成对称方阵和斜对称方阵之和。(考虑$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$)

2. 行列式

上一章我们引入了一个重要的概念：矩阵。这一章里我们介绍另外一个非常重要的概念：行列式。行列式是方阵上面的一类特殊的函数，它可以帮助我们更加透彻地了解矩阵。行列式有很多彼此等价的定义，从几何上讲，它可以看作行向量张成的几何体的有向体积，感兴趣的读者可以参看席南华院士编著的《基础代数》第四章行列式的部分。而在本章中我们将从代数的角度来递归定义行列式。

2.1 余子式展开定义行列式

上一章在定理1.4.5中，我们已经提到过对于2 阶的方阵

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

我们的行列式就定义为

$$\det(A) = ad - bc \quad \text{或者} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

方阵A可逆当且仅当$\det(A) \neq 0$, 此时A的逆矩阵是

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

练习：对于上述方阵A, 我们在平面坐标系上，将其行向量记为$\mathbf{v}_1 = (a, b)$ 以及$\mathbf{v}_2 = (c, d)$. 那么由v_1 和v_2做成的平行四边形的面积是多少？行列式的符号和两个向量的定向又有什么关系？

为了定义行列式，我们先来定义子式 (minor) 和余子式 (cofactor)的概念。设$A = (a_{ij})$ 是一个方阵。元素a_{ij} 的子式，记为M_{ij}, 是将A的第i行和第j列去掉之后留下方阵的的行列式。元素a_{ij}的余子式C_{ij} 定义为 $(-1)^{i+j} M_{ij}$.

定义(行列式)：设A是一个n阶方阵，我们定义行列式为(按第i行余子式展开)

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in},$$

或者(按第j列余子式展开)

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

在上述定义中，我们不加证明地用到了这样一个事实：按不同的行和列进行余子式展开时，得到的结果都是一样的，只有这样我们才可以按照余子式展开来定义行列式。而且，我们注意到上述定义行列式的方法也给了我们一种递归计算行列式的有效办法。

定理 2.1.2: 设A是一个$n \times n$ 的三角矩阵，则$\det A$ 是其主对角线元素的乘积。

证明：设三角矩阵A的形式是

$$A = \left(\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{array} \right).$$

我们按照其第一列展开求行列式，得到$\det A = a_{11}C_{11}$, 其中C_{11}即为矩阵A去掉第一行和第一列后剩余方阵

$$\left(\begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{array} \right).$$

的行列式。容易根据归纳法，知$C_{11} = a_{22}a_{33}\dots a_{nn}$. 最后，我们就得到了

$$\det A = a_{11}a_{22}\dots a_{nn}.$$

□

对于 2 阶方阵和 3 阶方阵，我们通常需要记住其行列式，教材中Figure 2.1.1给了我们一种记忆的方法。对于 2 阶方阵，我们有：

$$\left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| = a_{11}a_{22} - a_{12}a_{21} \text{ (”捺对角线”元乘积 – ”撇对角线”元乘积)}.$$

对于 3 阶方阵，我们有：

$$\left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

这里的第一项可以看成是”捺主对角线”元乘积，第二项和第三项分别是”捺次对角线”元乘积，第四项是”撇主对角线”元乘积，第五项和第六项分别是”撇次对角线”元乘积。”捺”记为正，”撇”记为负。于是可以想成：”捺对角线”元乘积– ” 撇对角线”元乘积.

2.2 通过行约化来计算行列式

定理 2.2.1: 设 A 是一个 n 阶方阵，如果 A 的某一行或者某一列为0, 则$\det A = 0$.

证明：因为行列式对不同的行（或者不同的列）展开的结果是一样的。于是我们对于全为0的那一行或者那一列展开就得到$\det A = 0$.

我们比较A的行列式和A^T 的行列式，在计算A的行列式时，对A行展开计算，等价于对A^T的列展开计算。从而我们得到下面的定理。

定理 2.2.2: 设 A 是一个 n 阶方阵，则$\det A = \det A^T$.

从上面的定理我们也可以看出，矩阵行列式关于行的性质可以有相应的关于列的版本的叙述。

下面的定理说明了三种基本的行（列）变换后，行列式是如何变化的。

定理 2.2.3: 设A 是n阶方阵，

(a) 如果方阵B 是A的某一行（或者某一列）乘上纯量λ所得，则$\det B = \lambda \det A$.

(b) 如果方阵B 是A的两行（或者两列）交换所得，则$\det B = -\det A$.

(c) 如果方阵B 是A的某一行的倍数加到另一行上（或者某一列的倍数加到另一列上），则$\det B = \det A$.

证明：我们只对行的操作来证明定理，列的操作是类似的。设方阵A 是

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

(a) 我们不妨假设B是由方阵A的第一行乘上纯量λ所得，即

$$B = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

我们对方阵B的第一行做行展开，有

$$\det B = \lambda a_{11} C_{11} + \lambda a_{12} C_{12} + \cdots + \lambda a_{1n} C_{1n},$$

其中C_{ij}是B的余子式，它同样是A的相应的余子式，从而我们不难得到$\det B = \lambda \det A$.

(b) 我们不妨假设交换的是相邻的两行，因为任意的交换两行都可以通过交换相邻两行的操作得到（读者可以思考如此假设对我们证明的结论的合理性）。更多地，我们可以假设我们交换的是第一行和第二行，于是矩阵B成为

$$B = \begin{pmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

我们对A的第一行展开得到行列式为：

$$\det A = a_{11} C_{11} + a_{12} C_{12} + \cdots + a_{1n} C_{1n},$$

其中$C_{1j} = (-1)^{1+j} M_{ij}$, M_{1j}是元素a_{1j}对应的子式。而我们对B的第二行展开得到行列式为

$$\det B = a_{11} C'_{11} + a_{12} C'_{12} + \cdots + a_{1n} C'_{1n},$$

不难发现，这里的$C'_{1j} = (-1)^{2+j} M_{1j}$. 从而我们得到$\det B = -\det A$.
(c) 由(b) 我们发现如果矩阵的两行相等，那么交换两行之后，矩阵不变，但是行列式变成其相反数，于是在此情况下，必然要求矩阵的行列式为0.

不妨设B是由A的第i行加到第1行所得，于是

$$B = \begin{pmatrix} a_{11} + \lambda a_{i1} & a_{12} + \lambda a_{i2} & \dots & a_{1n} + \lambda a_{in} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

对矩阵B的第一行展开，得到

$$\det B = (a_{11} + \lambda a_{i1})C_{11} + (a_{12} + \lambda a_{i2})C_{12} + \cdots + (a_{1n} + \lambda a_{in})C_{1n}$$

注意到 $a_{i1}C_{11} + a_{i2}C_{12} + \cdots + a_{in}C_{1n}$ 是将矩阵A 的第1行替换成第i 行后的矩阵所得的行列式，根据我们之前所论述的，此行列是为0. 于是

$$\det B = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \det A.$$

□

对应于三种初等行变换，我们有相应的初等矩阵，它们是对单位矩阵做初等行变换所得，于是它们的行列式由下面的定理刻画。

定理 2.2.4: 设E是 n 阶初等矩阵，
(a) 如果E 由I_n的某一行乘上纯量λ 所得，则$\det E = \lambda$.
(b) 如果E 由I_n 交换两行所得，则$\det E = 1$.
(c) 如果E 由I_n 的某一行加到另一行上，则$\det E = 1$.
定理 2.2.5: 如果方阵A有两行或者两列是成比例的，则$\det A = 0$. （留作练习）

根据定理2.2.3知道，对于一个方阵，每做一步初等变换，则相应的行列式如何变化，我们是知道的。通过初等变换，我们可以将方阵变换成阶梯型，而方阵的阶梯型必然是一个上三角矩阵，而上三角矩阵的行列式为其对角线元素的乘积。从而我们给出了一种求行列式的算法。

2.3 行列式的性质，Cramer 法则

设A和B是n阶方阵，λ是一个纯量。我们考虑$\det A$, $\det B$ 和

$$\det(\lambda A), \quad \det(A + B), \quad \det(AB)$$

之间的关系。

首先考虑$\det(\lambda A)$, 矩阵λA 是方阵A的每一行都乘以纯量λ, 于是不难发现

$$\det(\lambda A) = \lambda^n \det A.$$

一般而言，$\det(A+B) \neq \det A + \det B$, 举一个简单的例子

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ 和 } B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

于是，$\det A = \det B = 0$, 但是

$$A+B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ 且有 } \det(A+B) = 1.$$

对于一些特殊的情况，我们有下面的定理。

定理 2.3.1: 设矩阵A, B和C是第r行不同，其余行相同的矩阵，而且矩阵C的第r行是A的第r行和B的第r行之和，则我们有

$$\det C = \det A + \det B.$$

证明：我们只需要对C的第r行进行展开，就可以得到上述定理。 □

设A和B是两个n阶方阵，下面我们考虑$\det(AB)$. 我们知道矩阵的乘法和行列式的定义都是较为复杂的，但是我们却有下面的非常简洁且重要的等式

$$\det(AB) = \det A \det B.$$

引理 2.3.2: 设B是一个n阶方阵，而E是n阶初等矩阵，于是我们有

$$\det(EB) = \det E \det B.$$

证明：我们的初等变换有三种，相应的初等矩阵也有三种，所以我们分下面三种情况讨论：

情形一：设E是由I_n的某一行乘以纯量λ所得的初等矩阵, 那么不难发现矩阵EB就是由B在相同的行上乘以纯量λ. 从而由定理2.2.3 (a), 我们知道$\det(EB) = \lambda \det B$. 注意到由2.2.4 (a), 我们知道$\det E = \lambda$, 于是我们有$\det(EB) = \det E \det B$.

情形二：设E是由I_n的某两行交换所得的初等矩阵，那么不难发现矩阵EB就是由B交换相同的两行得到。从而由定理2.2.3 (b), 我们知道$\det(EB) = -\det B$. 注意到由2.2.4 (b), 我们知道$\det E = -1$, 于是我们有$\det(EB) = \det E \det B$.

情形三：设E是由I_n的第r行加到第s 行上所得的初等矩阵，那么不难发现矩阵EB就是由B的第r行加到第s行上得到的。从而由定理2.2.3 (c), 我们知道$\det(EB) = \det B$. 注意到由2.2.4 (c), 我们知道$\det E = 1$, 于是我们有$\det(EB) = \det E \det B$. □

根据上述引理，我们给出一个矩阵可逆的等价描述。

定理 2.3.3: 方阵A可逆当且仅当$\det A \neq 0$.

证明：设方阵A通过一些初等行变换变成简约的行阶梯型R，相应的初等行变换对应的矩阵是$E_1, E_2, \ldots, E_r$. 从而$R = E_r \ldots E_2 E_1 A$. 根据上面的引理，我们知道

$$\det(R) = \det(E_r) \ldots \det(E_2) \det(E_1) \det(A).$$

注意到初等矩阵的行列式不是 0, 即 $\det(E_i) \neq 0, \forall i$. 于是 $\det R$ 和 $\det E$ 同时为 0 或者同时不为 0. 根据定理1.6.4, 我们知道A可逆当且仅当$R = I_n$, 从而命题得证。 □

定理 2.3.4: 如果A和B是两个同型的方阵，则$\det(AB) = \det A \det B$.

证明：我们分成A是否可逆两种情况来考虑。当A不可逆的时候，定理1.6.5告诉我们此时AB也是不可逆的。于是我们有$\det(AB) = 0$以及$\det A = 0$, 自然地有$\det(AB) = \det A \det B$.

现在假设矩阵A是可逆的，由定理1.6.4知方阵A 可以写成一系列初等矩阵的乘积，即有$A = E_1E_2 \dots E_r$. 从而，$AB = E_1E_2 \dots E_rB$. 我们用引理2.3.2可以得到

$$\det(AB) = \det(E_1)\det(E_2)\dots\det(E_r)\det(B).$$

再次用引理2.3.2, 我们有

$$\det(E_1)\det(E_2)\dots\det(E_r) = \det(E_1E_2\dots E_r) = \det A.$$

于是$\det(AB) = \det A \det B$, 定理得证。 □

根据上述定理，我们容易得到

定理 2.3.5: 设方阵A可逆，则$\det(A^{-1}) = \frac{1}{\det(A)}$.

证明：由$A^{-1}A = I$ 知 $\det(A^{-1})\det(A) = 1$. 因为A可逆，从而$\det A \neq 0$. 于是我们有

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

接下来，我们通过伴随矩阵来给出矩阵的逆的形式。

伴随矩阵（定义）：设A是n阶方阵，C_{ij} 是a_{ij}的余子式，则矩阵

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}$$

称为矩阵A的余子式矩阵。它的转置称为A的伴随矩阵，我们记为 $\operatorname{adj}(A)$. □

定理 2.3.6: 设A是可逆矩阵，则

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A).$$

证明：我们只要证明$A\operatorname{adj}(A) = \det(A)I$ 即可。首先

$$A\mathrm{adj}(A) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} C_{11} & C_{21} & \dots & C_{j1} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{j2} & \dots & C_{n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ C_{1n} & C_{2n} & \dots & C_{jn} & \dots & C_{nn} \end{pmatrix},$$

则矩阵$A\mathrm{adj}(A)$的第i行和第j列的元素是

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}.$$

当$i = j$时，上面的式子恰为$\det A$. 当$i \neq j$ 时，我们将A的第j行替换成第i行的元素，得到新的矩阵A',此时A'的第 i 行和第 j 行相等，所以$\det A' = 0$. 但是我们对A'的第j行展开求行列式，会有

$$\det A' = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = 0.$$

从而，我们得到

$$A\mathrm{adj}(A) = \begin{pmatrix} \det(A) & 0 & \dots & 0 \\ 0 & \det(A) & \dots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \dots & \det(A) \end{pmatrix} = \det(A)I.$$

当A可逆时，$\det A \neq 0$, 此时$A\left(\frac{1}{\det(A)}\mathrm{adj}(A)\right) = I$. 于是，我们有

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A).$$

□

定理 2.3.7 (Cramer 法则): 设 $A\mathbf{x} = \mathbf{b}$ 是 n 个未知元的线性方程组且$\det A \neq 0$, 则方程组有唯一的解，它们可以表达为

$$x_1 = \frac{\det(A_1)}{\det(A)},\ x_2 = \frac{\det(A_2)}{\det(A)},\ \dots,\ x_n = \frac{\det(A_n)}{\det(A)}.$$

其中A_j是将A的第j列用列向量$\mathbf{b}$替换后得到的矩阵。

证明：如果$\det A \neq 0$, 则矩阵A 可逆。于是方程 $A\mathbf{x} = \mathbf{b}$ 有唯一的解$\mathbf{x} = A^{-1}\mathbf{b}$. 根据定理2.3.6, 我们知道

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)}\mathrm{adj}(A)\mathbf{b} = \frac{1}{\det(A)}\begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \dots & \dots & \dots & \dots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

通过计算，我们得到

$$\mathbf{x} = \frac{1}{\det(A)}\begin{pmatrix} b_1C_{11} + b_2C_{21} + \cdots + b_nC_{n1} \\ b_1C_{12} + b_2C_{22} + \cdots + b_nC_{n2} \\ \dots \\ b_1C_{1n} + b_2C_{2n} + \cdots + b_nC_{nn} \end{pmatrix},$$

于是

$$x_j = \frac{b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}}{\det(A)}.$$

令矩阵A_j是将A的第j 列用列向量$\mathbf{b}$ 替换后得到的矩阵，通过对矩阵A_j的第j列展开来求其行列式，我们容易知道

$$\det(A_j) = b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj},$$

于是有$x_j = \dfrac{\det(A_j)}{\det(A)}$. □

之前在定理1.7.1中，我们还没有给出下列命题的证明。

(c) 三角矩阵可逆当且仅当主对角线上的元素全不为零。

(d) 可逆上三角矩阵的逆还是上三角矩阵，可逆下三角矩阵的逆还是下三角矩阵。

证明：(c) 设A是上三角矩阵，其对角线元为$a_{11}, a_{22}, \dots, a_{nn}$. 所以矩阵$A$可逆当且仅当其行列式

$$\det(A) = a_{11}a_{22}\dots a_{nn}$$

不等于0, 而这等价于主对角线上的元全不为0.

(d) 我们证明当方阵A是上三角矩阵时，其逆也是上三角矩阵（下三角矩阵的情况是类似的）。由定理2.3.6 知

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A),$$

我们证明伴随矩阵adj(A)是上三角矩阵，注意到

$$\mathrm{adj}(A) = \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \dots & \dots & \dots & \dots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix},$$

因此我们要证明当$i < j$时，$C_{ij} = 0$.

记矩阵B_{ij} 是矩阵A去掉第i行和第j列后得到的矩阵，则$C_{ij} = (-1)^{i+j}\det(B_{ij})$. 因为$A$是上三角矩阵，则容易看出$B_{ij}$也是上三角矩阵，$B_{ij}$的第$i$行是原来$A$的第$i+1$行。又因为$i < j$, 所以$B_{ij}$的第$i$ 行从左数起，至少有i个0. 这说明B_{ij}的主对角线有0. 从而$\det(B_{ij}) = 0$. 从而我们就证明了当$i < j$ 时，$C_{ij} = 0$. □

3. 欧几里得线性空间

3.1 向量

在高中，我们学过平面空间以及立体空间的向量 (vector)，向量就是有长度 (length) 和方向 (direction) 的量，在平面和立体空间中，我们又称呼这样的向量是几何向量 (geometric vector). 一个向量由它的始点(initial point)和终点(terminal point) 确定。两个向量，如果它们的长度和方向都一样，我们就说它们是等价的，我们就把它们看成是同一个向量，虽然可能它们处于不同的位置。一个向量的始点和终点一样，这样的向量就是零向量(zero vector), 记为$\mathbf{0}$.

向量的加法：平行四边形法则或者三角形法则（高中知识）。向量的加法满足结合律：

$$(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w}).$$

向量的减法：对于向量$\mathbf{v}$, 向量$-\mathbf{v}$与向量$\mathbf{v}$ 的长度相等，方向相反。因此

$$\mathbf{u}-\mathbf{v} = \mathbf{u}+(-\mathbf{v}),$$

于是利用向量的加法可以定义向量的减法。

纯量乘法：设$\mathbf{v}$是一个几何向量，设λ是非零纯量，我们定义纯量乘法$\lambda\mathbf{v}$ 是这样的向量：其长度是$\mathbf{v}$的长度的$|\lambda|$倍，当λ 为正数（负数）时，方向和$\mathbf{v}$的方向相同（相反）。如果$\mathbf{v}$为$\mathbf{0}$或者$\lambda = 0$, 则$\lambda\mathbf{v} = \mathbf{0}$.

向量的坐标化：对于一个平面或者立体的坐标系，将向量$\mathbf{v}$的始点置于原点，那么向量$\mathbf{v}$ 就由其终点唯一确定，因此我们可以记$\mathbf{v} = (v_1, v_2)$ （二维情况）或者$\mathbf{v} = (v_1, v_2, v_3)$ （三维情况）。如果一个向量的始点不在原点上，以平面空间为例，我们记其始点$P_1 = (x_1, y_1)$, 终点是$P_2 = (x_2, y_2)$. 于是向量

$$\overrightarrow{P_1P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1} = (x_2 - x_1, y_2 - y_1).$$

上面在平面和立体空间定义的向量及其性质可以推广到一般的高维空间。但是在高维的空间里，我们没有很好的几何直观，但是向量的坐标化给我们这种推广带来了可能。

$n-$空间：我们记 $\mathbb{R}$ 是实数。对于正整数n, 带序的n元组(ordered n-tuple)是一组实数序列$(v_1, v_2, \ldots, v_n)$. 所有的带序的n元组形成的集合称为n-空间，我们记为$\mathbb{R}^n$.

你可以将$(v_1, v_2, \ldots, v_n)$想象成将$n-$空间中的一个点或者始点在原点的，终点在$(v_1, v_2, \ldots, v_n)$的向量，这样的向量我们记为$\mathbf{v} = (v_1, v_2, \ldots, v_n)$.

两个向量$\mathbf{v}=(v_1,v_2,\ldots,v_n)$ 和$\mathbf{w}=(w_1,w_2,\ldots,w_n)$ 称为等价(或者相等)，如果$\forall i, v_i=w_i$. 我们也记为$\mathbf{v}=\mathbf{w}$.

受到平面空间和立体空间向量运算的启发，我们可以定义一般的n-空间中的向量之间的运算。对于n- 空间中的两个向量$\mathbf{v}=(v_1,v_2,\ldots,v_n)$ 和$\mathbf{w}=(w_1,w_2,\ldots,w_n)$, λ是一个纯量，我们定义

$$\mathbf{v}+\mathbf{w}=(v_1+w_1,v_2+w_2,\ldots,v_n+w_n),\quad \lambda\mathbf{v}=(\lambda v_1,\lambda v_2,\ldots,\lambda v_n).$$

$$-\mathbf{v}=(-v_1,-v_2,\ldots,-v_n),\quad \mathbf{w}-\mathbf{v}=\mathbf{w}+(-\mathbf{v})=(w_1-v_1,w_2-v_2,\ldots,w_n-v_n).$$

根据定义，下列定理都是容易验证的。

定理 3.1.1: 设$\mathbf{u},\mathbf{v}$ 和$\mathbf{w}$ 是$\mathbb{R}^n$中的向量，设λ和μ分别是纯量，则

(a) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

(b) $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(c) $\mathbf{u}+\mathbf{0}=\mathbf{0}+\mathbf{u}=\mathbf{u}$.

(d) $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.

(e) $\lambda(\mathbf{u}+\mathbf{v})=\lambda\mathbf{u}+\lambda\mathbf{v}$.

(f) $(\lambda+\mu)\mathbf{u}=\lambda\mathbf{u}+\mu\mathbf{u}$.

(g) $\lambda(\mu\mathbf{u})=(\lambda\mu)\mathbf{u}$.

(h) $1\mathbf{u}=\mathbf{u}$.

上述的定理很基本也很重要，我们以后在介绍一般的线性空间时，会使用此定理作为我们定义的基础。作为练习，可以很容易证明下面的定理。

定理 3.1.2: 设$\mathbf{v}$是$\mathbb{R}^n$ 中的向量，λ是一个纯量，则

(a) $0\mathbf{v}=\mathbf{0}$; (b) $\lambda\mathbf{0}=\mathbf{0}$; (c) $(-1)\mathbf{v}=-\mathbf{v}$.

接下来，我们介绍一个重要的概念：线性组合(linear combination). 设$\mathbf{w}$是$\mathbb{R}^n$中的一个向量，w 被称为是向量$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r$ 的线性组合，如果$\mathbf{w}$可以写成如下的形式：

$$\mathbf{w}=\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_r\mathbf{v}_r,$$

其中$\lambda_1,\lambda_2,\ldots,\lambda_r$是纯量，被称为是线性组合的系数(coefficient).

3.2 $\mathbb{R}^n$中的范数，点乘和距离

在平面空间$\mathbb{R}^2$ 中，我们可以计算向量$\mathbf{v}=(v_1,v_2)$的长度，称为$\mathbf{v}$的范数(norm)，记成$\|\mathbf{v}\|$. 根据勾股定理，知

$$\|\mathbf{v}\|=\sqrt{{v_1}^2+{v_2}^2}.$$

在立体空间$\mathbb{R}^3$ 中，类似的分析知道$\mathbf{v}=(v_1,v_2,v_3)$ 的范数是

$$\|\mathbf{v}\|=\sqrt{{v_1}^2+{v_2}^2+{v_3}^2}.$$

收到这样的启发，我们定义一般的空间$\mathbb{R}^n$ 中向量$\mathbf{v}=(v_1,v_2,\ldots,v_n)$的范数$\|\mathbf{v}\|$为

$$\|\mathbf{v}\|=\sqrt{{v_1}^2+{v_2}^2+{v_3}^2+\cdots+{v_n}^2}.$$

根据范数的定义，我们很容易验证下面的定理

定理 3.2.1: 设$\mathbf{v}$是$\mathbb{R}^n$中的向量，λ是一个纯量，则

(a) $\|\mathbf{v}\|\geq 0$.

(b) $\|\mathbf{v}\|=0$ 当且仅当$\mathbf{v}=0$.

(c) $\|\lambda\mathbf{v}\|=|\lambda|\|\mathbf{v}\|$

单位向量：范数为1的向量称为单位向量。对于一个向量$\mathbf{v}$, 我们总可以构造与之方向相同的单位向量，只要令$\mathbf{u}=\dfrac{1}{\|\mathbf{v}\|}\mathbf{v}$, 它与$\mathbf{v}$的方向一样，且

$$\|\mathbf{u}\|=\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\|=\frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\|=1.$$

很容易看出来，$\mathbb{R}^n$ 中有这样一组标准的单位向量(standard unit vectors)

$$\mathbf{e}_1=(1,0,\ldots,0),\ \mathbf{e}_2=(0,1,\ldots,0),\ \ldots,\ \mathbf{e}_n=(0,0,\ldots,1)$$

使得对于任意的$\mathbb{R}^n$ 中的向量$\mathbf{v}=(v_1,v_2,\ldots,v_n)$ 都有

$$\mathbf{v}=(v_1,v_2,\ldots,v_n)=v_1\mathbf{e}_1+v_2\mathbf{e}_2+\cdots+v_n\mathbf{e}_n.$$

接下来我们定义$\mathbb{R}^n$中的距离。首先看平面空间，设$P_1=(x_1,y_1)$和$P_2=(x_2,y_2)$是平面空间中的两个点，于是P_1和P_2的距离为

$$d=\|\overrightarrow{P_1P_2}\|=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}.$$

类似地，在立体空间中，设两个点分别为$P_1=(x_1,y_1,z_1)$和$P_2=(x_2,y_2,z_2)$, 于是P_1和P_2的距离为

$$d=\|\overrightarrow{P_1P_2}\|=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}.$$

受这些启发，我们可以定义$\mathbb{R}^n$中两点的距离，设$\mathbf{u}=(u_1,u_2,\ldots,u_n)$和$\mathbf{v}=(v_1,v_2,\ldots,v_n)$, 定义这两点的距离(distance)

$$d(\mathbf{u},\mathbf{v})=\|\mathbf{u}-\mathbf{v}\|=\sqrt{(u_1-v_1)^2+(u_2-v_2)^2+\cdots+(u_n-v_n)^2}.$$

向量的点乘(dot product)：首先我们还是来看平面空间和立体空间的情况。对于$\mathbb{R}^2$或者$\mathbb{R}^3$中的两个向量$\mathbf{u}$和$\mathbf{v}$, 根据几何直观，将两向量的试点放在一起，我们可以考虑其夹角θ, 于是夹角θ满足$0\leq\theta\leq\pi$. 我们定义$\mathbf{u}$和$\mathbf{v}$的点乘（也称为欧式内积Euclidean inner product）为

$$\mathbf{u}\cdot\mathbf{v}=\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

当$\mathbf{u}=0$或者$\mathbf{v}=0$时，我们定义$\mathbf{u}\cdot\mathbf{v}=0$.

根据定义，我们容易知道

$$\cos\theta=\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|},$$

于是有，当$\mathbf{u}\cdot\mathbf{v}>0$ 时，θ是锐角；当$\mathbf{u}\cdot\mathbf{v}<0$ 时，θ是钝角；当$\mathbf{u}\cdot\mathbf{v}=0$ 时，θ是直角。

根据三角形的余弦定理，我们知道$\cos\theta$有下面的表达形式：

$$\cos\theta=\frac{\|\mathbf{u}\|^2+\|\mathbf{v}\|^2-\|\mathbf{v}-\mathbf{u}\|^2}{2\|\mathbf{u}\|\|\mathbf{v}\|}.$$

从而我们得到了

$$\mathbf{u}\cdot\mathbf{v}=\frac{1}{2}(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2-\|\mathbf{v}-\mathbf{u}\|^2).$$

代入

$$\|\mathbf{u}\|^2={u_1}^2+{u_2}^2+{u_3}^2,\quad \|\mathbf{v}\|^2={v_1}^2+{v_2}^2+{v_3}^2$$

以及

$$\|\mathbf{v}-\mathbf{u}\|^2=(v_1-u_1)^2+(v_2-u_2)^2+(v_3-u_3)^2.$$

化简后，我们最终得到

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+u_2v_2+u_3v_3.$$

对于平面的情况，类似有

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+u_2v_2.$$

受以上启发，我们可以定义一般空间$\mathbb{R}^n$中两个向量$\mathbf{u}=(u_1,u_2,\ldots,u_n)$ 和$\mathbf{v}=(v_1,v_2,\ldots,v_n)$ 的点乘（或者欧式内积）为

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+u_2v_2+\cdots+u_nv_n.$$

点乘有很多很好的性质，下面两条定理都是比较好证明的，可以作为一个好的练习。

定理 3.2.2: 设$\mathbf{u},\mathbf{v}$ 和$\mathbf{w}$ 是$\mathbb{R}^n$中的向量，λ 是纯量，则

(a) $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$

(b) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$

(c) $\lambda(\mathbf{u}\cdot\mathbf{v})=(\lambda\mathbf{u})\cdot\mathbf{v}$

(d) $\mathbf{v}\cdot\mathbf{v}\geq 0$ 且$\mathbf{v}\cdot\mathbf{v}=0$ 当且仅当$\mathbf{v}=0$

定理 3.2.3: 设$\mathbf{u},\mathbf{v}$ 和$\mathbf{w}$ 是$\mathbb{R}^n$ 中的向量，λ是纯量，则

(a) $\mathbf{0}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{0}=0$

(b) $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$

(c) $\mathbf{u}\cdot(\mathbf{v}-\mathbf{w})=\mathbf{u}\cdot\mathbf{v}-\mathbf{u}\cdot\mathbf{w}$

(d) $(\mathbf{u}-\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}-\mathbf{v}\cdot\mathbf{w}$

(c) $\lambda(\mathbf{u}\cdot\mathbf{v})=\mathbf{u}\cdot(\lambda\mathbf{v})$

下面我们介绍$\mathbb{R}^n$ 中的Cauthy-Schwarz不等式，在后面的章节中会有更加一般的版本。

定理 3.2.4: 设$\mathbf{u}=(u_1,u_2,\ldots,u_n)$ 和$\mathbf{v}=(v_1,v_2,\ldots,v_n)$ 是$\mathbb{R}^n$中的两个向量，则

$$|\mathbf{u}\cdot\mathbf{v}|\le\|\mathbf{u}\|\|\mathbf{v}\|$$

或者等价地有

$$|u_1v_1+u_2v_2+\ldots u_nv_n|\le(u_1^2+u_2^2+\cdots+u_n^2)^{1/2}(v_1^2+v_2^2+\cdots+v_n^2)^{1/2}.$$

证明：我们考虑$\mathbb{R}$上的函数 $f(x)=\sum\limits_{i=1}^{n}(u_ix-v_i)^2$, 则$f(x)\ge 0,\ \forall x\in\mathbb{R}$. 将$f(x)$ 展开，我们有

$$f(x)=(\sum_{i=1}^{n}u_i^2)x^2-2(\sum_{i=1}^{n}u_iv_i)x+\sum_{i=1}^{n}v_i^2,$$

根据判别式恒大于等于0, 我们就得到了定理的结论。 □

在平面几何几何中，我们知道(a)三角形的第三边总是小于另外两边之和；(b)两点之间线段最短。这两条结论我们可以推广到一般的$\mathbb{R}^n$中。

定理 3.2.5: 设$\mathbf{u},\mathbf{v}$ 和$\mathbf{w}$ 是$\mathbb{R}^n$中的向量，λ 是纯量，则有下列三角不等式成立

(a) $\|\mathbf{u}+\mathbf{v}\|\le\|\mathbf{u}\|+\|\mathbf{v}\|$

(b) $d(\mathbf{u},\mathbf{v})\le d(\mathbf{u},\mathbf{w})+d(\mathbf{w},\mathbf{v})$

证明：(a) 因为我们有

$$\|\mathbf{u}+\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})=\mathbf{u}\cdot\mathbf{u}+2(\mathbf{u}\cdot\mathbf{v})+\mathbf{v}\cdot\mathbf{v}\le\|\mathbf{u}\|^2+2|\mathbf{u}\cdot\mathbf{v}|+\|\mathbf{v}\|^2,$$

根据Cauthy-Schwarz 不等式有

$$|\mathbf{u}\cdot\mathbf{v}|\le\|\mathbf{u}\|\|\mathbf{v}\|.$$

从而我们知道

$$\|\mathbf{u}+\mathbf{v}\|^2\le\|\mathbf{u}\|^2+2\|\mathbf{u}\|\|\mathbf{v}\|+\|\mathbf{v}\|^2=(\|\mathbf{u}\|+\|\mathbf{v}\|)^2.$$

两边开根号，即得$\|\mathbf{u}+\mathbf{v}\|\le\|\mathbf{u}\|+\|\mathbf{v}\|$.

(b) $d(\mathbf{u},\mathbf{v})=\|\mathbf{u}-\mathbf{v}\|=\|(\mathbf{u}-\mathbf{w})+(\mathbf{w}-\mathbf{v})\|\le\|(\mathbf{u}-\mathbf{w})\|+\|(\mathbf{w}-\mathbf{v})\|=d(\mathbf{u},\mathbf{w})+d(\mathbf{w},\mathbf{v})$

在平行四边形中，我们知道两条对角线的平方和等于四条边的平方和。这在一般的$\mathbb{R}^n$ 中也有类似的定理。

定理 3.2.6（向量的平行四边形等式）：设$\mathbf{u}$ 和$\mathbf{v}$ 是$\mathbb{R}^n$ 中的向量，则：

$$\|\mathbf{u}+\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{v}\|^2=2(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2).$$

证明：我们直接计算，有

$$\|\mathbf{u}+\mathbf{v}\|^2+\|\mathbf{u}-\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})+(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})=2(\mathbf{u}\cdot\mathbf{u})+2(\mathbf{v}\cdot\mathbf{v})=2(\|\mathbf{u}\|^2+\|\mathbf{v}\|^2).$$

□

下面的定理告诉我们点乘可以用范数表示出来，证明留作练习。

定理 3.2.7: 设$\mathbf{u}$ 和$\mathbf{v}$ 是$\mathbb{R}^n$ 中的向量，则：

$$\mathbf{u}\cdot\mathbf{v}=\frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2-\frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2$$

其实，$\mathbb{R}^n$ 中的点乘可以看成矩阵的乘法的一种特殊情况。对于$\mathbb{R}^n$ 中的向量$\mathbf{u},\mathbf{v}$, 我们把它们看成是$n\times 1$的列向量，于是我们很自然地有

$$\mathbf{u}\cdot\mathbf{v}=\mathbf{u}^T\mathbf{v}=\mathbf{v}^T\mathbf{u}$$

对于任意的$n\times n$矩阵A, 有

$$A\mathbf{u}\cdot\mathbf{v}=\mathbf{v}^TA\mathbf{u} \text{ 以及 } \mathbf{u}\cdot A^T\mathbf{v}=(A^T\mathbf{v})^T\mathbf{u}=\mathbf{v}^TA\mathbf{u}$$

从而我们得到了$A\mathbf{u}\cdot\mathbf{v}=\mathbf{u}\cdot A^T\mathbf{v}$. 从而就有$\mathbf{u}\cdot A\mathbf{v}=A^T\mathbf{u}\cdot\mathbf{v}$

注：当把$\mathbb{R}^n$ 中的向量$\mathbf{u},\mathbf{v}$看作行向量或者列向量的时候，点乘都有相应的矩阵乘法的表示方式，具体参见教材中的Table 1.

3.3 正交性质

在上一节中我们定义了点乘，在平面和立体空间中，两个向量$\mathbf{u},\mathbf{v}$ 的夹角是直角，当且仅当$\mathbf{u}\cdot\mathbf{v}=0$. 这个概念可以推广到在一般的$\mathbb{R}^n$中，我们称两个向量$\mathbf{u},\mathbf{v}$是正交的，如果$\mathbf{u}\cdot\mathbf{v}=0$. 我们看作零向量$\mathbf{0}$和所有的向量都正交。如果$\mathbb{R}^n$的某个子集中的向量彼此之间正交，则称呼此子集是正交集。如果再加上这些向量都是单位向量的条件，我们就称呼这个子集是标准正交子集。

给定向量$\mathbf{n}$, 我们考虑满足等式$\mathbf{n}\cdot\mathbf{v}=0$的向量$\mathbf{v}$, 即所有与$\mathbf{n}$正交的向量$\mathbf{v}$. 在平面空间中，记$\mathbf{n}=(a,b)$, 于是方程$\mathbf{n}\cdot\mathbf{v}=0$是

$$ax+by=0,$$

是与向量$\mathbf{n}=(a,b)$ 垂直的直线。在立体空间中，记$\mathbf{n}=(a,b,c)$ 于是方程$\mathbf{n}\cdot\mathbf{v}=0$ 是

$$ax+by+cz=0,$$

是与向量$\mathbf{n}=(a,b,c)$ 垂直的平面。

我们来看平面上的直线$ax+by=c$, 假设(x_0,y_0) 直线上的一点，于是$ax_0+by_0=c$. 从而直线可以写成

$$a(x-x_0)+b(y-y_0)=0.$$

这就说明了此条直线可以看成过点(x_0,y_0)的与向量$\mathbf{n}=(a,b)$垂直的直线。

类似地，我们看空间上的平面$ax+by+cz=d$, 假设(x_0,y_0,z_0) 平面上的一点，于是$ax_0+by_0+cz_0=d$. 从而平面可以写成

$$a(x-x_0)+b(y-y_0)+c(z-z_0)=0.$$

这就说明了此平面可以看成过点(x_0,y_0,z_0)的与向量$\mathbf{n}=(a,b,c)$垂直的平面。

定理 3.3.2（投射定理）：设 $\mathbf{u}$ 和 $\mathbf{a}$ 是 $\mathbb{R}^n$ 中的向量且 $\mathbf{a}\neq\mathbf{0}$, 则 $\mathbf{u}$ 可以唯一地写成 $\mathbf{u}=\mathbf{w}_1+\mathbf{w}_2$, 其中 $\mathbf{w}_1$ 和 $\mathbf{a}$ 共线（方向相同或者相反），$\mathbf{w}_2$ 与 $\mathbf{a}$ 正交。

证明：我们令 $\mathbf{w}_1=\lambda\mathbf{a}$ 以及 $\mathbf{w}_2=\mathbf{u}-\lambda\mathbf{a}$. 考虑到$\mathbf{w}_2$ 与 $\mathbf{a}$ 正交，于是有

$$\mathbf{w}_2\cdot\mathbf{a}=(\mathbf{u}-\lambda\mathbf{a})\cdot\mathbf{a}=\mathbf{u}\cdot\mathbf{a}-\lambda\|\mathbf{a}\|^2=0.$$

从而知道$\lambda=\dfrac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}$, 于是我们找出了满足条件的表达式 $\mathbf{u}=\mathbf{w}_1+\mathbf{w}_2$, 而且根据上面的推导过程，我们知道这种表达式是唯一的。 □

有了上面的投射分解，我们称$\mathbf{w}_1=\mathrm{proj}_{\mathbf{a}}\mathbf{u}$ 称为$\mathbf{u}$在$\mathbf{a}$ 上的正交投射，称$\mathbf{w}_2=\mathbf{u}-\mathrm{proj}_{\mathbf{a}}\mathbf{u}$称为与$\mathbf{a}$正交的$\mathbf{u}$的分量。

勾股定理(Pythagoras定理)：在平面直角三角形中，我们有著名的勾股定理。一般地，我们有下面的定理，其证明是容易的，我们留作练习。

定理 3.3.3：设 $\mathbf{u}$ 和 $\mathbf{v}$ 是 $\mathbb{R}^n$ 中彼此正交的向量，则

$$\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+\|\mathbf{v}\|^2.$$

在本小节最后，我们来给出平面上一点到另外一条直线的距离，以及空间中一点到另外一平面的距离的公式。

定理 3.3.4：

(a) 在$\mathbb{R}^2$中，一点$P_0(x_0,y_0)$到直线$ax+by+c=0$的距离D为

$$D=\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}.$$

(b) 在$\mathbb{R}^3$中，一点$P_0(x_0,y_0,z_0)$到平面$ax+by+cz+d=0$的距离D 为

$$D=\frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}.$$

证明：我们给出(b)的证明，(a) 的证明是类似的。设$Q(x_1,y_1,z_1)$是平面$ax+by+cz+d=0$上的一点，而平面的法向量为$\mathbf{n}=(a,b,c)$, 从而D 是向量$\overrightarrow{QP_0}$ 在$\mathbf{n}$ 上正交投影的长度，即有

$$D=\|\mathrm{proj}_{\mathbf{n}}\overrightarrow{QP_0}\|=\frac{|\overrightarrow{QP_0}\cdot\mathbf{n}|}{\|\mathbf{n}\|},$$

注意到$\overrightarrow{QP_0}=(x_0-x_1,y_0-y_1,z_0-z_1)$, 于是

$$\overrightarrow{QP_0}\cdot\mathbf{n}=a(x_0-x_1)+b(y_0-y_1)+c(z_0-z_1).$$

而$\|\mathbf{n}\|=\sqrt{a^2+b^2+c^2}$, 我们得到

$$D=\frac{|a(x_0-x_1)+b(y_0-y_1)+c(z_0-z_1)|}{\sqrt{a^2+b^2+c^2}}.$$

但是$Q(x_1, y_1, z_1)$是平面$ax + by + cz + d = 0$上的一点，于是$ax_1 + by_1 + cz_1 + d = 0$. 从而有

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

3.4 线性方程的几何

首先我们来看，平面或立体空间中的一条直线可以由上面的一点以及它的方向唯一确定下来。现在假设直线L上有一点x_0, 而向量$\mathbf{v}$与直线L平行。于是我们知道直线可以由下列等式表达出来：

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}.$$

当$x_0 = 0$时，直线过原点，方程为$\mathbf{x} = t\mathbf{v}$.

设W是立体空间$\mathbb{R}^3$上的平面，其包含点x_0, 与向量$\mathbf{v}_1$和$\mathbf{v}_2$是（$\mathbf{v}_1$和$\mathbf{v}_2$ 不共线）平行的，于是平面可以由下列等式表达出来：

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2.$$

当$x_0 = 0$时，平面过原点，方程为$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$.

一般地，在空间$\mathbb{R}^n$中，方程

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}.$$

代表过点x_0, 与向量$\mathbf{v}$ 平行的直线。在空间$\mathbb{R}^n$中，设向量$\mathbf{v}_1$和$\mathbf{v}_2$ 不共线，则方程

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2.$$

代表过x_0且与向量$\mathbf{v}_1$和$\mathbf{v}_2$平行的平面。

平面几何中，我们知道两点确定一条直线。现在设$\mathbf{x}_0$和$\mathbf{x}_1$是$\mathbb{R}^n$中两个不同的点。于是过这两点的直线可以表达为

$$\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0),$$

或者等价地，有$\mathbf{x} = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$. 更多地，我们知道

$$\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)(0 \le t \le 1), \quad \mathbf{x} = (1-t)\mathbf{x}_0 + t\mathbf{x}_1(0 \le t \le 1)$$

代表着连接$\mathbf{x}_0$ 和$\mathbf{x}_1$的线段。

给定含有n 个未知元的线性方程$x_1, x_2, \ldots, x_n$

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

设向量$\mathbf{a} = (a_1, a_2, \ldots, a_n)$ 以及向量$\mathbf{x} = (x_1, x_2, \ldots, x_n)$, 于是上述方程可以写成$\mathbf{a} \cdot \mathbf{x} = b$的形式。特别地，当$b = 0$时，齐次方程可以写为$\mathbf{a} \cdot \mathbf{x} = 0$. 于是对于一般的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{3.1}$$

如果记系数形成的行向量分别为$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$, 则方程可以写成

$$\mathbf{r}_1 \cdot \mathbf{x} = 0, \quad \mathbf{r}_2 \cdot \mathbf{x} = 0, \quad \ldots, \quad \mathbf{r}_m \cdot \mathbf{x} = 0$$

这样我们就证明了下述定理：

定理 3.4.3：设A是$m \times n$ 的矩阵，则齐次线性方程组$A\mathbf{x} = \mathbf{0}$ 的解包含所有与矩阵A的m 个行向量都正交的向量。

下面，我们考虑齐次线性方程组$A\mathbf{x} = \mathbf{0}$ 与非齐次线性方程组$A\mathbf{x} = \mathbf{b}$之间的关系。

定理 3.4.4：方程$A\mathbf{x} = \mathbf{b}$的通解可以由它的一个特解加上方程$A\mathbf{x} = \mathbf{0}$的通解得到。

证明：我们记$\mathbf{x}_0$是方程$A\mathbf{x} = \mathbf{b}$的一个特解，记W是$A\mathbf{x} = \mathbf{0}$ 的解空间。我们首先说明$\mathbf{x}_0 + W$是方程$A\mathbf{x} = \mathbf{b}$ 的解。设$\mathbf{x} = \mathbf{x}_0 + \mathbf{w} \in \mathbf{x}_0 + W$, 则

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{w}) = A\mathbf{x}_0 + A\mathbf{w} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

反过来，假设$\mathbf{x}$是方程$A\mathbf{x} = \mathbf{b}$的任意一个解，我们考虑$\mathbf{x} - \mathbf{x}_0$, 于是有

$$A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

这说明$\mathbf{x} - \mathbf{x}_0 \in W$, 即有$\mathbf{x} \in \mathbf{x}_0 + W$. 命题得证。 □

3.5 向量的叉乘(cross product)

这一节我们介绍在物理和工程上应用很广泛的：向量的叉乘。

叉乘的定义：设$\mathbf{u} = (u_1, u_2, u_3)$以及$\mathbf{v} = (v_1, v_2, v_3)$ 是$\mathbb{R}^3$中的两个向量，于是叉乘$\mathbf{u} \times \mathbf{v}$定义为向量

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1),$$

也可以写成行列式的形式：

$$\mathbf{u} \times \mathbf{v} = \left(\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}, - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}, \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \right)$$

我们记$\mathbf{i}=(1,0,0)$, $\mathbf{j}=(0,1,0)$, $\mathbf{k}=(0,0,1)$, 那么很容易验证叉乘$\mathbf{u}\times\mathbf{v}$可以形式地写为

$$\mathbf{u}\times\mathbf{v}=\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}=\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}\mathbf{i}-\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}\mathbf{j}+\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}\mathbf{k}$$

注意：向量的点乘运算后是纯量，而向量的叉乘运算后仍然是向量。

定理 3.5.1：（向量点乘和叉乘的联系）设$\mathbf{u},\mathbf{v}$和$\mathbf{w}$是$\mathbb{R}^3$中的向量，则

(a) $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=0$

(b) $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})=0$

(c) $\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2$

(d) $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})=(\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{u}\cdot\mathbf{v})\mathbf{w}$

(e) $(\mathbf{u}\times\mathbf{v})\times\mathbf{w}=(\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{u}$

证明：(a) 记$\mathbf{u}=(u_1,u_2,u_3)$, $\mathbf{v}=(v_1,v_2,v_3)$, 按定义有

$$\mathbf{u}\times\mathbf{v}=(u_2v_3-u_3v_2,\ u_3v_1-u_1v_3,\ u_1v_2-u_2v_1),$$

从而我们有

$$\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=u_1(u_2v_3-u_3v_2)+u_2(u_3v_1-u_1v_3)+u_3(u_1v_2-u_2v_1)=0.$$

类似于(a), 我们可以证明(b).

(c) 我们对等式(c)的两端分别展开：

$$\|\mathbf{u}\times\mathbf{v}\|^2=(u_2v_3-u_3v_2)^2+(u_3v_1-u_1v_3)^2+(u_1v_2-u_2v_1)^2$$

以及

$$\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2=(u_1^2+u_2^2+u_3^2)(v_1^2+v_2^2+v_3^2)-(u_1v_1+u_2v_2+u_3v_3)^2$$

从而我们知道(c)成立。

(d) 因为任意的$\mathbf{u}=(u_1,u_2,u_3)$可以写成$\mathbf{u}=u_1\mathbf{i}+u_2\mathbf{j}+u_3\mathbf{k}$, 从而根据定理3.5.2的性质$(c)$, 我们只需要对$\mathbf{u}=\mathbf{i},\mathbf{j},\mathbf{k}$证明即可。我们以$\mathbf{u}=\mathbf{i}$为例来验算。首先

$$\mathbf{v}\times\mathbf{w}=(v_2w_3-v_3w_2,v_3w_1-v_1w_3,v_1w_2-v_2w_1)$$

于是

$$\mathbf{i}\times(\mathbf{v}\times\mathbf{w})=(0,v_2w_1-v_1w_2,v_3w_1-v_1w_3).$$

而我们注意到

$$(\mathbf{i}\cdot\mathbf{w})\mathbf{v}=w_1\mathbf{v}=(w_1v_1,w_1v_2,w_1v_3),\quad (\mathbf{i}\cdot\mathbf{v})\mathbf{w}=v_1\mathbf{w}=(v_1w_1,v_1w_2,v_1w_3).$$

最终得到$\mathbf{i}\times(\mathbf{v}\times\mathbf{w})=(\mathbf{i}\cdot\mathbf{w})\mathbf{v}-(\mathbf{i}\cdot\mathbf{v})\mathbf{w}$.

(e)的计算和(d)是类似的。 □

注意：由(d)和(e)的结果，我们可以知道一般来讲

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) \neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}.$$

定理 3.5.2 （向量叉乘的基本性质）：设$\mathbf{u}, \mathbf{v}$和$\mathbf{w}$是$\mathbb{R}^3$ 中的向量，λ 是纯量，则

(a) $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$

(b) $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$

(c) $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$

(d) $\lambda(\mathbf{u} \times \mathbf{v}) = (\lambda \mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (\lambda \mathbf{v})$

(e)$\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$

(f)$\mathbf{u} \times \mathbf{u} = \mathbf{0}$

证明：上述性质都可以由行列式的基本性质推导而得到。 □

我们定理3.5.1的性质(a)和(b)可以知道向量$\mathbf{u} \times \mathbf{v}$与$\mathbf{u}$和$\mathbf{v}$都是正交的，它的方向可以通过右手法则确定：设$\mathbf{u}$和$\mathbf{v}$的夹角为$\theta (\theta \leq \pi)$, 我们的右手四指按着角$\theta$, 由$\mathbf{u}$ 向$\mathbf{v}$ 旋转，拇指的方向就是$\mathbf{u} \times \mathbf{v}$的方向(参见教材中Figure 3.5.3)。确定了$\mathbf{u} \times \mathbf{v}$的方向，那么向量$\mathbf{u} \times \mathbf{v}$ 的长度呢？

定理 3.5.3：设$\mathbf{u}, \mathbf{v}$是$\mathbb{R}^3$中的向量，则$\|\mathbf{u} \times \mathbf{v}\|$ 等于$\mathbf{u}$和$\mathbf{v}$确定的平行四边形的面积。

证明：注意到$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u} \times \mathbf{v}\| \cos \theta$, 其中$\theta$是$\mathbf{u}$和$\mathbf{v}$的夹角，那么

$$\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2 \theta) = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2 \theta.$$

于是我们有$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$, 恰为$\mathbf{u}$和$\mathbf{v}$确定的平行四边形的面积。 □

对于$\mathbb{R}^3$ 中的向量$\mathbf{u}, \mathbf{v}$和$\mathbf{w}$, 我们定义它们的三元纯量积(scalar triple product)为$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$. 我们记$\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$以及$\mathbf{w} = (w_1, w_2, w_3)$, 首先有

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} \mathbf{k}$$

从而

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} u_1 - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} u_2 + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} u_3$$

恰好是$\mathbf{u}, \mathbf{v}$和$\mathbf{w}$ 作为行向量所形成矩阵的行列式

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

之前我们提到过行列式有其几何的描述，下面我们具体给出定理。

定理 3.5.4：

(a) 行列式

$$\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$$

的绝对值等于由向量$\mathbf{u} = (u_1, u_2)$和$\mathbf{v} = (v_1, v_2)$张成的平行四边形的面积。

(b) 行列式

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

的绝对值等于由向量$\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$ 和$\mathbf{w} = (w_1, w_2, w_3)$张成的平行四边体的体积。

证明：对于(a), 我们给出两种方法证明。平行四边形的面积是$\|\mathbf{u}\|\|\mathbf{v}\| \sin\theta$, 其中$\theta$是向量$\mathbf{u}$和$\mathbf{v}$的夹角。而

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{u_1 v_1 + u_2 v_2}{\sqrt{u_1^2 + u_2^2}\sqrt{v_1^2 + v_2^2}}$$

于是通过计算有

$$\sin\theta = \frac{|u_1 v_2 - u_2 v_1|}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

从而平行四边形的面积是$|u_1 v_2 - u_2 v_1|$.

另外一种证明(a)的方法是使用叉乘的几何意义。我们将$\mathbf{u}$和$\mathbf{v}$看成是$\mathbb{R}^3$中的向量$\mathbf{u} = (u_1, u_2, 0)$和$\mathbf{v} = (v_1, v_2, 0)$. 于是$\|\mathbf{u} \times \mathbf{v}\|$是$\mathbf{u}$ 和$\mathbf{v}$确定的平行四边形的面积，而

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & 0 \\ v_1 & v_2 & 0 \end{vmatrix} = \det \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix} \mathbf{k}$$

则

$$\|\mathbf{u} \times \mathbf{v}\| = |\det \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix}|$$

(b) 我们将$\mathbf{v}$和$\mathbf{w}$形成的平行四边形看成平行四面体的底面。从而底面的面积是$\|\mathbf{v} \times \mathbf{w}\|$. 下面我们来求高，因为向量$\mathbf{v} \times \mathbf{w}$与底面垂直，所有我们需要求$\mathbf{u}$在向量$\mathbf{v} \times \mathbf{w}$上的投影，则有高

$$h = \|\mathrm{proj}_{\mathbf{v} \times \mathbf{w}} \mathbf{u}\| = \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|}.$$

从而平行四面体的体积是底面乘以高，为$|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|$, 就是行列式

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

的绝对值。 □

4. 一般的线性空间

4.1 实向量空间

在前面我们介绍了$\mathbb{R}^n$中的向量所满足的一些性质。在这一节中我们将这些性质作为公理，从而介绍一般的线性空间。我们会发现我们经常遇到的很多例子都有向量空间的结构。向量空间可以定义在一般的域上，在本课程中我们更多地考虑的是定义在实数以及复数上的向量空间，此一节只考虑实向量空间。

向量空间的定义：设V是一个非空的集合，上面有两种运算：加法和纯量乘积。其中加法：对于V中任意的两个对象$\mathbf{u}$和$\mathbf{v}$, 记$\mathbf{u}+\mathbf{v}$为相应的加法。纯量乘积：对于V中对象$\mathbf{u}$ 以及纯量λ, 记$\lambda\mathbf{u}$为相应的纯量乘积。我们称V是一个向量空间，如果下列条件满足：

1. 如果$\mathbf{u}$和$\mathbf{v}$ 在V中，则$\mathbf{u}+\mathbf{v}$也在V中。
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
3. $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$.
4. 存在V中对象0（称为零向量）使得：对于$\mathbf{v}\in V$, 有$0+\mathbf{v}=\mathbf{v}+0$.
5. 对于任意的$\mathbf{v}\in V$, 存在$-\mathbf{v}\in V$ 使得$\mathbf{v}+(-\mathbf{v})=(-\mathbf{v})+\mathbf{v}=0$.
6. 设λ是纯量，$\mathbf{v}$是V中的任意对象，则$\lambda\mathbf{v}$在V中。
7. $\lambda(\mathbf{u}+\mathbf{v})=\lambda\mathbf{u}+\lambda\mathbf{v}$.
8. $(\lambda+\mu)\mathbf{v}=\lambda\mathbf{v}+\mu\mathbf{v}$.
9. $\lambda(\mu\mathbf{v})=(\lambda\mu)\mathbf{v}$.
10. $1\mathbf{v}=\mathbf{v}$.

当V是一个向量空间时，我们称V中的对象是向量（如上面所说的$\mathbf{u},\mathbf{v}$以及$\mathbf{w}$）。

我们验证一个集合V是不是向量空间的时候，首先要看是否可以定义它上面作为向量的加法以及纯量乘积。然后验证集合V对加法和纯量乘积是不是封闭的（即定义中的公理1和公理6）. 最后再验证剩余的公理。我们举下面的例子：

例1: 零向量空间。集合中只有一个元素，我们记为 $\mathbf{0}$, 其中加法和纯量乘积定义为：

$$\mathbf{0}+\mathbf{0}=\mathbf{0},\quad \lambda\mathbf{0}=\mathbf{0}$$

很容易验证此为向量空间，我们称为零向量空间。

例2: 我们在第三章研究较多的实向量空间$\mathbf{R}^n$.

例3: 所有的$m\times n$阶的矩阵，我们记成$M_{m\times n}(\mathbb{R})$.

例4: 记V是所有的从$\mathbb{R}$到$\mathbb{R}$上的函数。记f和g是函数，λ是纯量，我们可以在V上定义加法和纯量乘积为：

$$(f+g)(x)=f(x)+g(x),\quad (\lambda f)(x)=\lambda f(x).$$

可以验证，这样定义的加法和纯量乘积使V形成一个向量空间（作为练习）。

例5: 最后我们给出一个例子说明在集合$\mathbb{R}^2$上定义另外一种加法和纯量乘积，$\mathbb{R}^2$可以不是向量空间。对于$\mathbf{u} = (u_1, u_2)$以及$\mathbf{v} = (v_1, v_2)$, 定义

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2), \quad \lambda\mathbf{u} = (\lambda u_1, 0)$$

可以容易验证，从公理1到公理9都是成立的，但公理10却不满足，$\forall \mathbf{u} = (u_1, u_2) \in \mathbb{R}^2$ 满足$u_2 \neq 0$我们有

$$1\mathbf{u} = (u_1, u_2) \neq (u_1, 0).$$

定理 4.1.1：设V是向量空间，$\mathbf{v}$是其中的向量，λ 是纯量，则

(a) $0\mathbf{v} = \mathbf{0}$

(b) $\lambda\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{v} = -\mathbf{v}$

(d) 如果$\lambda\mathbf{v} = 0$, 则$\lambda = 0$或者$\mathbf{v} = \mathbf{0}$.

证明：(a) 由公理8我们知道$0\mathbf{v} + 0\mathbf{v} = (0 + 0)\mathbf{v} = 0\mathbf{v}$. 因为向量$0\mathbf{v}$有逆向量$-0\mathbf{v}$, 两边同时加上$-0\mathbf{v}$, 于是左边成为$0\mathbf{v}+(0\mathbf{v}+(-0\mathbf{v})) = 0\mathbf{v}+\mathbf{0} = 0\mathbf{v}$ 而右边成为$0\mathbf{v}+(-0\mathbf{v}) = \mathbf{0}$, 从而有$0\mathbf{v} = \mathbf{0}$.

(b) 与(a)的证明类似，我们有$\lambda\mathbf{0} + \lambda\mathbf{0} = \lambda(\mathbf{0} + \mathbf{0}) = \lambda\mathbf{0}$, 两边加上$-\lambda\mathbf{0}$, 我们有$\lambda\mathbf{0} = \mathbf{0}$.

(c) 我们只要证明$\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$即可。由公理10, 我们知道$1\mathbf{v} = \mathbf{v}$, 从而

$$\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1 + (-1))\mathbf{v} = 0\mathbf{v} = \mathbf{0}.$$

(d) 我们反设$\lambda \neq 0$以及$\mathbf{v} \neq \mathbf{0}$, 在$\lambda\mathbf{v} = 0$两边同时乘以$\lambda^{-1}$, 于是有

$$\lambda^{-1}(\lambda\mathbf{v}) = (\lambda^{-1}\lambda)\mathbf{v} = 1\mathbf{v} = \mathbf{v} = \mathbf{0}.$$

从而我们得到矛盾。 □

4.2 子空间

子空间的定义：向量空间V的子集W称为是V的子空间，如果它在V的加法和纯量乘积下是一个向量空间。

向量空间V的子集W的加法和纯量乘积由V决定，所以我们要验证W是子空间，需要验证向量空间定义中的10条公理，但是有一些公理，如公理2, 3, 7, 8, 9, 10是自然成立的，由V中向量的性质过渡而来, 因此只需要考虑公理1, 4, 5, 6. 下面的定理告诉我们其实只需要考虑公理1 和公理6, 即子集W对加法和纯量乘法封闭。

定理 4.2.1：设W是V中包含一些向量的非空子集，则W是V的子空间当且仅当下面的条件满足：

(a) 当$\mathbf{u}$和$\mathbf{v}$在W 中，则$\mathbf{u} + \mathbf{v}$在W中。

(b) 设λ是纯量，$\mathbf{u}$是W中的向量，则$\lambda\mathbf{u}$在W中。

证明：当W是V的子空间时，性质(a)和(b)自然是满足的。反过来，假设条件(a)和(b)是成立的，它们就是公理1和公理6。根据之前的分析，我们只需要验证公理4和公理5. 对于W中的任意一

个向量$\mathbf{u}$, 考虑$0\mathbf{u} = \mathbf{0}$ 以及$(-1)\mathbf{u} = -\mathbf{u}$, 由性质(b)知道，它们都在$W$中。从而我们就证明了定理。 □

线性子空间的例子：

例1: 零子空间：只包含一个元素零向量。容易验证它对加法和数乘都是封闭的。

例2: 矩阵空间的一些子空间：(a) 对称矩阵; (b) 三角矩阵; (c) 对角矩阵.

例3: 实数上的函数空间$F(-\infty,\infty)$有如下的子空间的包含关系：

$$F(-\infty,\infty) \supseteq C(-\infty,\infty) \supseteq C^1(-\infty,\infty) \supseteq C^m(-\infty,\infty) \supseteq C^\infty(-\infty,\infty) \supseteq P_n$$

这里$C(-\infty,\infty)$是连续函数，$C^m(-\infty,\infty)$ 是m次连续可导函数，$C^\infty(-\infty,\infty)$是任意阶可导函数，P_n 是次数小于等于n的多项式函数。

定理 4.2.2：设$W_1, W_2, \ldots, W_r$ 是V的子空间，则它们的交也是V的子空间。

证明：记W是$W_1, W_2, \ldots, W_r$的交，由于$W_1, W_2, \ldots, W_r$ 是V的子空间，它们都含有零向量，因此W含有零向量，对于W 在V下的加法和数乘的封闭性是易于验证的。 □

线性组合(linear combination)的定义：设$\mathbf{w}$是V中的向量，我们称$\mathbf{w}$是向量$\mathbf{v}_1$, $\mathbf{v}_2$,..., $\mathbf{v}_r$的线性组合，如果$\mathbf{w}$可以写成如下的形式

$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r,$$

这里的纯量$\lambda_1, \lambda_2, \ldots, \lambda_r$是线性组合的系数。

定理 4.2.3：设$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$是线性空间$V$的非空集合，则：

(a) S中向量的所有线性组合形成的集合W是V的子空间。

(b) 上述的W是包含S中所有向量的最小子空间。

证明：(a) 按照条件，有W中的向量都是S中向量的线性组合，于是对于任意的$\mathbf{u}, \mathbf{v} \in W$, 有

$$\mathbf{u} = \lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2 + \cdots + \lambda_r\mathbf{w}_r, \quad \mathbf{v} = \mu_1\mathbf{w}_1 + \mu_2\mathbf{w}_2 + \cdots + \mu_r\mathbf{w}_r$$

从而$\mathbf{u} + \mathbf{v} = (\lambda_1 + \mu_1)\mathbf{w}_1 + (\lambda_2 + \mu_2)\mathbf{w}_2 + \cdots + (\lambda_r + \mu_r)\mathbf{w}_r$, 仍然是$S$中向量的线性组合，这说明$W$对于加法是封闭的。类似地，我们可以证明$W$对于数乘也是封闭的。因此$W$是$V$的子空间。

(b) 设W'是包含S中所有向量的一个线性空间，因为W'对加法和数乘封闭，因此必然是包含S中向量的线性组合，从而包含线性子空间W, 这意味着W是包含S中所有向量的最小子空间。

线性张成(span): 设V是一个线性空间，S是它的一个非空集合，由S中的向量的所有线性组合形成的子空间称为由S 张成的子空间，记为$\text{span}\{S\}$. 特别地，当$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$, 则记此张成为$\text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$. □

线性张成的例子：

例1: 在线性空间R^n中，我们考虑这些向量：

$$\mathbf{e}_1 = (1,0,0,\ldots,0),\ \mathbf{e}_2 = (0,1,0,\ldots,0),\ \ldots,\ \mathbf{e}_n = (0,0,0,\ldots,1)$$

于是对于R^n中的任意向量$\mathbf{v} = (v_1, v_2, \ldots, v_n)$, 我们可以将$\mathbf{v}$写成

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n$$

是$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$的线性组合, 从而$R^n$ 可以由$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$张成。
例2: 在矩阵空间$M_{m\times n}(\mathbb{R})$中， 设$\mathbf{e}_{ij}$是第$i$行, 第$j$列为1,其余的元素为0的矩阵， 则容易验证$M_{m\times n}(\mathbb{R})$由$\mathbf{e}_{ij}(1 \le i \le m, 1 \le j \le n)$ 张成。
例3: 次数小于等于n的多项式空间P_n. 因为任意的多项式$f(x) \in P_n$都可以写成

$$f(x) = a_0 + a_1x + \cdots + a_nx^n.$$

从而P_n可以由$1, x, \ldots, x^n$张成。

定理 4.2.4：关于n个变量的齐次线性方程组$A\mathbf{x} = \mathbf{0}$的解的集合是$\mathbb{R}^n$ 的子空间。
证明：记W是$A\mathbf{x} = \mathbf{0}$的解的集合，首先零解在W中。下面我们只需要说明W在加法和数乘下封闭。设$\mathbf{x}_1$和$\mathbf{x}_2$是W中的向量，即有

$$A\mathbf{x}_1 = \mathbf{0}, \quad A\mathbf{x}_2 = \mathbf{0}$$

从而我们知道

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

则$\mathbf{x}_1 + \mathbf{x}_2 \in W$, 即$W$对向量的加法封闭。

另一方面, 我们也有

$$A(\lambda\mathbf{x}_1) = \lambda A\mathbf{x}_1 = \lambda\mathbf{0} = \mathbf{0},$$

即W对纯量乘积也是封闭的。这就说明W是$\mathbb{R}^n$ 的子空间。
定理 4.2.5：设$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ 以及$S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$是向量空间$V$的非空集合，则

$$\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = \mathrm{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$$

当且仅当任意S中的向量都可以表示成S'中向量的线性组合，且任意S'中的向量都可以表示成S中向量的线性组合。
证明：我们记$W = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ 以及$W' = \mathrm{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$. 根据定义，$W$中的向量都是$S$中向量的线性组合而得到，根据条件因为任意$S$中的向量都可以表示成$S'$ 中向量的线性组合，于是W中的向量都可以表示成S'中向量的线性组合。注意到W'是S'中所有向量的线性组合，从而我们知道必然有$W \subseteq W'$. 因为任意S'中的向量也可以表示成S中向量的线性组合，于是同理我们有$W' \subseteq W$. 从而我们最终得到$W = W'$. □

4.3 线性无关

在上一节中，我们引入了线性张成的概念。对于线性空间V的子空间$W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, W中的向量都可以写成$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$的线性组合。我们会问，在$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$中是否存在更少个数的向量张成$W$. 为此，我们引入线性无关和线性相关的概念。

线性无关(linear independence)和线性相关(linear dependence)的定义：设$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是线性空间$V$中的向量。如果方程

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r = \mathbf{0}$$

只有零解，我们称向量组$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是线性无关的，否则（即方程有非平凡的解），我们称$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是线性相关的。

(注)：如果$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$中有零向量，则此向量组一定是线性相关的。

线性无关和线性相关的例子：

例1: 在$\mathbb{R}^n$中，考虑下面标准的单位向量

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0),\ \mathbf{e}_2 = (0, 1, 0, \ldots, 0),\ \ldots,\ \mathbf{e}_n = (0, 0, 0, \ldots, 1),$$

容易验证这些向量是线性无关的。

例2: 在矩阵空间$M_{m\times n}(\mathbb{R})$中，设$\mathbf{e}_{ij}$是第$i$行, 第$j$列为1,其余的元素为0的矩阵，则$\mathbf{e}_{ij}(1 \leq i \leq m, 1 \leq j \leq n)$是线性无关的。

例3: 次数小于等于n的多项式空间P_n. 考虑向量组$1, x, x^2, \ldots, x^n$, 此向量组线性无关。

线性无关和线性相关的另外一种描述：

定理 4.3.1：设线性空间V的子集S至少包含两个向量。

(a) S中的向量线性相关当且仅当至少存在一个S中的向量可以写成剩余向量的线性组合。

(b) S中的向量线性无关当且仅当不存在S中的向量可以写成其他剩余向量的线性组合。

证明：(a) 设$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, 于是$S$中的向量线性无关，当且仅当存在不全为0的纯量$\lambda_1, \lambda_2, \ldots, \lambda_r$ 使得

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r = \mathbf{0}$$

不妨假设$\lambda_k \neq 0$, 从而我们可以得到

$$\mathbf{v}_k = -\frac{\lambda_1}{\lambda_k}\mathbf{v}_1 - \cdots - \frac{\lambda_{k-1}}{\lambda_k}\mathbf{v}_{k-1} - \frac{\lambda_{k+1}}{\lambda_k}\mathbf{v}_{k+1} - \cdots - \frac{\lambda_r}{\lambda_k}\mathbf{v}_r,$$

即$\mathbf{v}_k$可以写成其他向量的线性组合。

反过来，我们设

$$\mathbf{v}_j = \mu_1\mathbf{v}_1 + \cdots + \mu_{j-1}\mathbf{v}_{j-1} + \mu_{j+1}\mathbf{v}_{j+1} + \cdots + \mu_n\mathbf{v}_n$$

从而有

$$\mu_1\mathbf{v}_1 + \cdots + \mu_{j-1}\mathbf{v}_{j-1} - \mathbf{v}_j + \mu_{j+1}\mathbf{v}_{j+1} + \cdots + \mu_n\mathbf{v}_n = 0.$$

这就说明S是线性相关的。

通过(a)的证明，我们可以知道(b)也是正确的。 $\square$

定理 4.3.3：设$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$是$\mathbb{R}^n$中的$r$个向量。证明：当$r > n$ 时，向量组$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$是线性相关的。

证明：现在设向量$\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{in})$. 我们考虑方程

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_r \mathbf{v}_r = \mathbf{0},$$

它等价于如下的线性方程组：

$$\begin{cases} v_{11}\lambda_1 + v_{21}\lambda_2 + \cdots + v_{r1}\lambda_r = 0 \\ v_{12}\lambda_1 + v_{22}\lambda_2 + \cdots + v_{r2}\lambda_r = 0 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ v_{1n}\lambda_1 + v_{2n}\lambda_2 + \cdots + v_{rn}\lambda_r = 0 \end{cases}$$

这是一个齐次线性方程组，且$r > n$, 即未知元的个数大于方程组方程的个数，从而方程组有非平凡的解。于是我们知道此时$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$是线性相关的。 $\square$

定义函数的Wronski行列式(Wronskian): 设$\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \dots, \mathbf{f}_n = f_n(x)$ 是在 $(-\infty, \infty)$ 上 $n-1$ 阶可导的函数，称行列式

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \dots & f_n(x) \\ f_1'(x) & f_2'(x) & \dots & f_n'(x) \\ \dots & \dots & \dots & \dots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \dots & f_n^{(n-1)}(x) \end{vmatrix},$$

是$f_1(x), f_2(x), \dots, f_n(x)$的Wronski行列式。

定理 4.3.4：设函数$f_1(x), f_2(x), \dots, f_n(x)$ 是在$(-\infty, \infty)$上$n-1$阶可导的函数，如果这些函数的Wronski行列式在$(-\infty, \infty)$上不是恒等于0, 则这些函数在$C^{(n-1)}(-\infty, \infty)$ 中是线性无关的。

证明：假设函数$\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \dots, \mathbf{f}_n = f_n(x)$是线性相关的，即存在不全为0的纯量$\lambda_1, \lambda_2, \dots, \lambda_n$使得

$$\lambda_1 \mathbf{f}_1 + \lambda_2 \mathbf{f}_2 + \cdots + \lambda_n \mathbf{f}_n = \mathbf{0},$$

即对于任意的$x \in (-\infty, \infty)$, 都有

$$\lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_n f_n(x) = 0.$$

对此等式求导，可以得到下面的方程组：

$$\begin{cases} \lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_n f_n(x) = 0 \\ \lambda_1 f_1'(x) + \lambda_2 f_2'(x) + \cdots + \lambda_n f_n'(x) = 0 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ \lambda_1 f_1^{(n-1)}(x) + \lambda_2 f_2^{(n-1)}(x) + \cdots + \lambda_n f_n^{(n-1)}(x) = 0 \end{cases}$$

也可以写成

$$\begin{pmatrix} f_1(x) & f_2(x) & \dots & f_n(x) \\ f_1'(x) & f_2'(x) & \dots & f_n'(x) \\ \dots & \dots & \dots & \dots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \dots & f_n^{(n-1)}(x) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}.$$

由于$f_1(x), f_2(x), \dots, f_n(x)$的Wronski行列式不恒等于0. 从而上述方程中的解只能是

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0,$$

与我们的假设矛盾，于是可以知道函数$\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \dots, \mathbf{f}_n = f_n(x)$ 是线性无关的。 □

4.4 基底和坐标(basis and coordinates)

基底(basis)的定义：设V是线性空间，$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$是$V$的有限子集，则$S$称为$V$的基底，如果下面的条件满足：

(a) S是线性无关的。

(b) S张成V.

首先看一些基本的例子：

例1: 在线性空间R^n中，向量：

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0),\ \mathbf{e}_2 = (0, 1, 0, \dots, 0),\ \dots,\ \mathbf{e}_n = (0, 0, 0, \dots, 1)$$

是R^n的一个基底。

例2: 在矩阵空间$M_{m\times n}(\mathbb{R})$中，设$\mathbf{e}_{ij}$是第$i$行, 第$j$列为1,其余的元素为0的矩阵，则$\mathbf{e}_{ij}(1 \le i \le m, 1 \le j \le n)$是$M_{m\times n}(\mathbb{R})$的一个基底。

例3: 对于次数小于等于n的多项式空间P_n, 我们知道$S = \{1, x, \dots, x^n\}$是P_n 的一个基底。

当然我们注意到多项式空间P_∞没有有限的向量组可以张成它。如果一个空间可以由有限的向量组张成，我们就称这样的空间是有限维的(finite-dimensional)，否则称为无限维的(infinite dimensional).

给了线性空间V的基底S, 我们就可以坐标化。这是因为下面的定理保证了我们可以这样做。

定理 4.4.1：设$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$是$V$的一个基，则每一个向量$\mathbf{v}$都可以唯一地写成

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n.$$

证明：因为S张成V, 所以V中的向量都可以写成S中向量的线性组合，下面我们证明表达式的唯一性。假设$\mathbf{v}$有表达式

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n,$$

以及表达式

$$\mathbf{v} = \lambda_1'\mathbf{v}_1 + \lambda_2'\mathbf{v}_2 + \cdots + \lambda_n'\mathbf{v}_n.$$

两式相减，我们得到

$$(\lambda_1-\lambda_1')\mathbf{v}_1+(\lambda_2-\lambda_2')\mathbf{v}_2+\cdots+(\lambda_n-\lambda_n')\mathbf{v}_n=\mathbf{0}$$

由于$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$线性无关，从而$\lambda_j-\lambda_j'=0$ $(1\le j\le n)$, 即对任意的$1\le j\le n$，我们都有 $\lambda_j=\lambda_j'$, 这就证明了表达方式的唯一性。 □

有了上述定理，我们可以引入下面的概念。

设$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是$V$的一个基底，且$\mathbf{v}$有表达式

$$\mathbf{v}=\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_n\mathbf{v}_n.$$

则$\lambda_1,\lambda_2,\ldots,\lambda_n$是向量$\mathbf{v}$关于基底$S$的坐标。此时，$\mathbb{R}^n$中的向量$(\lambda_1,\lambda_2,\ldots,\lambda_n)$是$\mathbf{v}$关于$S$的坐标向量，我们记为$(\mathbf{v})_S=(\lambda_1,\lambda_2,\ldots,\lambda_n)$.

（注）：我们在坐标化的时候，往往会固定基底中向量的顺序。上述坐标化给出了从向量空间V到$\mathbb{R}^n$的一个一一对应。

4.5 维数(dimension)

首先，我们提出这一节最重要的定理：

定理 4.5.1：对于一个有限维的向量空间，它的每一个基底包含的向量个数相同。

此定理可以由下面的结果简单推导直接得到。

定理 4.5.2: 设V是一个有限维向量空间，设$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是$V$的一个基。则：

(a) 如果V的子集含有超过n个向量，则此子集是线性相关的。

(b) 如果V的子集含有的向量少于n个，则此子集无法张成空间V.

证明：(a) 设$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$是$V$的子集，且$m>n$. 我们证明$S$是线性相关的。因为$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是$V$的一个基, 从而对于任意的$k$ $(1\le k\le m)$, 有

$$\mathbf{w}_k=a_{1k}\mathbf{v}_1+a_{2k}\mathbf{v}_2+\cdots+a_{nk}\mathbf{v}_n.$$

我们考虑方程$\lambda_1\mathbf{w}_1+\lambda_2\mathbf{w}_2+\cdots+\lambda_m\mathbf{w}_m=\mathbf{0}$, 代入$\mathbf{w}_k$的表达式，有

$$\sum_{j=1}^{n}(\lambda_1a_{j1}+\lambda_2a_{j2}+\cdots+\lambda_ma_{jm})\mathbf{v}_j=\mathbf{0}.$$

因为$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是$V$的一个基，所以它们线性无关，于是

$$\begin{cases}a_{11}\lambda_1+a_{12}\lambda_2+\cdots+a_{1m}\lambda_m=0\\a_{21}\lambda_1+a_{22}\lambda_2+\cdots+a_{2m}\lambda_m=0\\\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\\a_{n1}\lambda_1+a_{n2}\lambda_2+\cdots+a_{nm}\lambda_m=0\end{cases}$$

由于$m>n$，未知元的个数大于方程的个数，从而我们知道方程有非零解，这就说明集合S是线性相关的，我们证明了(a).

(b) 现在设$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$是$V$的子集，且$m<n$. 我们证明集合$S$无法张成空间$V$. 现在假设$S$张成$V$, 则任意的$\mathbf{v}_k$都是$S$中向量的线性组合，类似于($a$)的证明方法，我们可以说明$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是线性相关的，这就得到了矛盾，从而说明了$S$无法张成空间$V$. □

根据**定理** 4.5.1, 我们可以定义有限维线性空间V的一个重要的概念，即维数，记为$\dim V$, 它是V的一个基底所含有的向量的个数。特别地，我们定义只含有零向量的空间的维数是0.

例：在之前我们反复提到的例子中，$\dim\mathbb{R}^n=n$, $\dim M_{m\times n}=mn$, $\dim P_n=n+1$.

定理 4.5.3：设S是向量空间V的非空子集。

(a)如果S是线性无关的集合，那么如果$\mathbf{v}\in V$不在空间span$\{S\}$中，我们可以知道$S\cup\{\mathbf{v}\}$仍然是线性无关的集合。

(b)设向量$\mathbf{v}\in S$可以写成S 中其他向量的线性组合，从而$S-\{\mathbf{v}\}$和S生成相同的空间。

证明：(a) 设$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$是线性无关的集合，且$\mathbf{v}\in V$不在空间span$\{S\}$中，我们证明$S'=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n,\mathbf{v}\}$也是线性无关的集合。考虑方程

$$\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_n\mathbf{v}_n+\lambda\mathbf{v}=\mathbf{0}$$

首先必有$\lambda=0$, 否则$\mathbf{v}$可以写成S 中向量的线性组合，于是方程变成

$$\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_n\mathbf{v}_n=\mathbf{0}$$

而S是线性无关的，于是我们知道$\lambda_1=\lambda_2=\cdots=\lambda_n=0$. 从而$S\cup\{\mathbf{v}\}$ 是线性无关的集合。

(b) 我们假设$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$, 且$\mathbf{v}_n$可以写成$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_{n-1}\}$的线性组合，即

$$\mathbf{v}_n=\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_{n-1}\mathbf{v}_{n-1}.$$

对于任意的$\mathbf{w}\in\mathrm{span}\{S\}$, $\mathbf{w}$可以写成

$$\mathbf{w}=\mu_1\mathbf{v}_1+\mu_2\mathbf{v}_2+\cdots+\mu_n\mathbf{v}_n$$

代入$\mathbf{v}_n$的表达式，有

$$\mathbf{w}=\mu_1\mathbf{v}_1+\mu_2\mathbf{v}_2+\cdots+\mu_{n-1}\mathbf{v}_{n-1}+\mu_n(\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_{n-1}\mathbf{v}_{n-1})$$

是$S-\{\mathbf{v}_n\}$的线性组合。这也就说明了$\mathbf{w}\in S-\{\mathbf{v}_n\}$, 于是span$\{S\}\subseteq$ span$\{S-\{\mathbf{v}\}\}$. 注意到显然的事实： span$\{S-\{\mathbf{v}\}\}\subseteq$ span$\{S\}$. 从而我们知道了$S-\{\mathbf{v}\}$和S生成相同的空间。

如何判断线性空间V中给定的一个一组向量是不是它的基，首先我们知道这组向量的个数必然等于V的维数。当向量组的元素个数等于空间的维数时，下面的定理告诉了我们一种判别的方法。

定理 4.5.4：设V是一个n-维的线性空间，且S是V的子集，恰好是n个向量。则S是V的基底，当且仅当S张成V或者S 是线性无关的。

证明：设S张成V，我们证明它是V的一个基。根据基的定义，只需要证明S是线性无关的。现在假设S线性相关，那么存在向量$\mathbf{v} \in S$是S中剩余向量的线性组合。于是根据**定理**4.5.3(b)，我们知道 $\mathrm{span}\{S - \{\mathbf{v}\}\} = \mathrm{span}\{S\} = V$. 但是根据**定理**4.5.2$(b)$, 这是不可能的。从而我们知道$S$是$V$的基底。

设S是线性无关的，我们证明它是V的一个基。根据基的定义，只需要证明S张成V. 假设S不能张成V, 则根据**定理**4.5.3(a)，我们知道存在$\mathbf{w} \in V$且$\mathbf{w} \notin S$, 使得$S \cup \{\mathbf{w}\}$仍然是线性无关的。但是根据**定理**4.5.2(a), 这是不可能的，推出矛盾。从而我们知道S是V的基底。

下面的定理告诉我们，如何从已有的向量组S得到V 的一个基。

定理 4.5.5：设S是有限维向量空间V的一个有限子集。

(a) 如果S张成V但不是V的基，则S 可以通过去掉某些合适的向量，得到V的一个基。

(b) 如果S是线性无关的，但不是V的基，则S可以通过添加一些合适的向量得到V的一个基。

证明：(a) 如果S张成V但不是V的基，则S是线性相关的，于是S中至少存在一个向量$\mathbf{v}$是S 中其余向量的线性组合。令$S' = S - \{\mathbf{v}\}$，根据**定理**4.5.3(b), S'也张成V. 如果S'线性无关，从而就知道S'是V的基。如果S'线性相关，那么我们可以继续刚才对S的操作。因为V 是有限维的，这样一步步操作下去，最终可以得到V的一个基。

(b) 如果S是线性无关的但不是V的基，根据**定理**4.5.3(b)，存在$\mathbf{w} \in V - \mathrm{span}\,\{S\}$, 使得$S' = S \cup \{\mathbf{w}\}$ 是线性无关的。如果S'能够张成V，则S 是V的基。否则我们对S'进行刚才同样的操作，因为V是有限维的，这样一步步操作下去，最终可以得到V的一个基。

定理 4.5.6：(作为练习) 设W是有限维向量空间V的子空间，则

(a) W是有限维的。

(b) $\dim W \leq \dim V$.

(c) $W = V$ 当且仅当$\dim W = \dim V$.

4.6 基底的改变

前面我们知道对于有限维线性空间V, 给定一个基$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, 则对于$\mathbf{v} \in V$, 我们知道$(\mathbf{v})_S = (c_1, c_2, \ldots, c_n)$是它关于基底$S$的坐标。于是$\mathbf{v} \to (\mathbf{v})_S$给出了从$V$到$\mathbb{R}^n$的一个一一对应，称为坐标映射。在这一节中，我们也可以将坐标写成列向量，记为

$$[\mathbf{v}]_S = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

我们自然而然会提出这样一个问题，如果换另外一组基，那么向量的坐标应该怎样变化？考虑线性空间V原来的一个基底$B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$和新的一个基底$B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_n\}$. 对于$\mathbf{v} \in V$, 我们来考虑$[\mathbf{v}]_B$和$[\mathbf{v}]_{B'}$的联系。

我们记

$$[\mathbf{v}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \quad [\mathbf{v}]_{B'} = \begin{bmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_n \end{bmatrix},$$

即我们有

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \begin{pmatrix} \mathbf{u}_1, & \mathbf{u}_2, & \ldots, & \mathbf{u}_n \end{pmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

$$\mathbf{v} = c'_1\mathbf{u}'_1 + c'_2\mathbf{u}'_2 + \cdots + c'_n\mathbf{u}'_n = \begin{pmatrix} \mathbf{u}'_1, & \mathbf{u}'_2, & \ldots, & \mathbf{u}'_n \end{pmatrix} \begin{bmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_n \end{bmatrix}$$

因为B和B'是V的两个基，从而可以用B将B' 中的向量表达出来。即对于任意的$k(1 \leq k \leq n)$, 有

$$\mathbf{u}'_k = p_{1k}\mathbf{u}_1 + p_{2k}\mathbf{u}_2 + \cdots + p_{nk}\mathbf{u}_n.$$

这也就意味着

$$\begin{pmatrix} \mathbf{u}'_1, & \mathbf{u}'_2, & \ldots, & \mathbf{u}'_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1, & \mathbf{u}_2, & \ldots, & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

将上述等式代入到$\mathbf{v}$的表达式，有

$$\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \begin{bmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_n \end{bmatrix}.$$

由于$B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$是一个基，所以上面等式两边的表达方式应该是唯一的，这就意味着$[\mathbf{v}]_B = P[\mathbf{v}]_{B'}$, 这里方阵$P$的列向量分别是

$$[\mathbf{u}'_1]_B, \ [\mathbf{u}'_2]_B, \ \ldots, \ [\mathbf{u}'_n]_B$$

称为从B'到B的转换矩阵(transition matrix), 记为$P_{B'\to B}$. 类似地，从B到B'的转换矩阵(transition matrix), 记为$P_{B\to B'}$.

我们很容易知道$P_{B'\to B}P_{B\to B'}=P_{B\to B}=I$ 以及$P_{B\to B'}P_{B'\to B}=P_{B'\to B'}=I$. 从而得到

定理 4.6.1：设P是从B'到B 的转换矩阵，则P是可逆的，P^{-1}是从B到B' 的转换矩阵。

在$\mathbb{R}^n$中，给定一组基$B=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$, 每一个$\mathbf{u}_k$ 看成是一个列向量，因此我们可以写成矩阵的形式

$$(\mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_n).$$

考虑$\mathbb{R}^n$中的标准基$S=\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$, 于是从$B$ 到S的转换矩阵为

$$P_{B\to S}=(\mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_n),$$

为一个可逆矩阵。

给定$\mathbb{R}^n$ 中两组基$B=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$ 和$B'=\{\mathbf{u}'_1,\mathbf{u}'_2,\ldots,\mathbf{u}'_n\}$, 我们将给出一个算法来求$P_{B'\to B}$. 根据之前的讨论，我们有

$$(\mathbf{u}'_1 \mid \mathbf{u}'_2 \mid \ldots \mid \mathbf{u}'_n)=(\mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_n)\ P_{B'\to B}.$$

如果我们记方阵$U=(\mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_n)$ 以及$U'=(\mathbf{u}'_1 \mid \mathbf{u}'_2 \mid \ldots \mid \mathbf{u}'_n)$, 并将它们拼凑在一起$(U \mid U')$, 两边同时作初等行变换，使得左边的矩阵$U$变成单位矩阵$I$, 则右边的$U'$就变成了矩阵$P_{B'\to B}$.

4.7 行空间，列空间和零空间

对于给定的一个矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \ldots & \ldots & \ldots & \ldots\\ a_{m1} & a_{m2} & \ldots & a_{mn}\end{pmatrix},$$

我们记$\mathbf{r}_i=(a_{i1},a_{i2},\ldots,a_{in})(1\le i\le m)$是$A$的$m$个行向量。记

$$\mathbf{c}_j=\begin{bmatrix} a_{1j}\\ a_{2j}\\ \vdots\\ a_{mj}\end{bmatrix}(1\le j\le n)$$

是A的n个列向量。

行空间，列空间和零空间的定义：对于给定的$m\times n$的矩阵A, 则由行向量生成的$\mathbb{R}^n$ 的子空间称为A 的行空间(row space). 由列向量生产的$\mathbb{R}^m$ 的子空间称为A的列空间(colume space). 线性方程组$A\mathbf{x}=\mathbf{0}$的解空间是$\mathbb{R}^n$ 的子空间，称为矩阵A的零空间(null space).

这一节，我们主要考虑下面两个问题：

问题1：线性方程组$A\mathbf{x} = \mathbf{b}$的解和矩阵A的行空间，列空间以及零空间的关系是怎样的？

问题2：矩阵A的行空间，列空间以及零空间的关系是怎样的？

定理 4.7.1：线性方程组$A\mathbf{x} = \mathbf{b}$是相容的，当且仅当$\mathbf{b}$在A的列空间中。

证明：将矩阵A写成按照列向量的分块矩阵的形式，即

$$A = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n)$$

从而$A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \dots x_n\mathbf{c}_n$, 这就说明了方程$A\mathbf{x} = \mathbf{b}$有解，当且仅当$\mathbf{b}$能写成$A$的列向量的线性组合，即$\mathbf{b}$在$A$的列空间中。

定理 3.4.4 告诉了我们下面的结论:

定理 4.7.2：设$\mathbf{x}_0$是方程$A\mathbf{x} = \mathbf{b}$的一个解，$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$是矩阵$A$的零空间的一组基，则方程$A\mathbf{x} = \mathbf{b}$的所有解都有形式

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \dots + \lambda_k\mathbf{v}_k.$$

下面的三个定理都是不难的，留给读者作为练习。

定理 4.7.3：初等行变换不改变矩阵的零空间。

定理 4.7.4：初等行变换不改变矩阵的行空间。

定理 4.7.5：如果一个矩阵R是行阶梯形式，则带有首1的行向量形成了矩阵R的行空间的一个基。带有首1的列向量形成了矩阵R的列空间的一个基。

定理 4.7.6：设A和B是行等价的矩阵，证明：

(a) 矩阵A的一些列向量是线性无关的，当且仅当矩阵B相应的列向量是线性无关的。

(b) 矩阵A的一些列向量形成A的列空间的一个基，当且仅当矩阵B相应的列向量形成B的列空间的一个基。

证明：注意到下面的事实，对于矩阵A的一些列向量$\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$, 通过行变换后，变成矩阵$B$的列向量$\{\mathbf{w}'_1, \mathbf{w}'_2, \dots, \mathbf{w}'_r\}$. 由于$\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$线性相关，即存在不全为0的常数$\lambda_1, \lambda_2, \dots, \lambda_r$, 使得

$$\lambda_1\mathbf{w}_1 + \lambda_2\mathbf{w}_2 + \dots + \lambda_r\mathbf{w}_r = \mathbf{0}.$$

我们不难发现下列等式成立(因为初等行变换不改变方程的解)

$$\lambda_1\mathbf{w}'_1 + \lambda_2\mathbf{w}'_2 + \dots + \lambda_r\mathbf{w}'_r = \mathbf{0}.$$

根据上述(a)的结论以及**定理** 4.5.4, 我们就得到了(b).

□

在本节的最后，我们考虑这样的问题：给定$\mathbb{R}^n$中的一个集合$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, 找出它的一个子集作为空间$\text{span}\{S\}$的基底，并将$S$ 中的其他向量表示成这个基的线性组合。具体的算法如下：

步骤一：由$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$作为列向量，形成矩阵$A$.

步骤二：将矩阵A通过初等行变换变成行阶梯形式R, 将R的列向量记为$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$.
步骤三：找出由R中包含首1的列向量，并对应到矩阵A中相应的列向量，这就找出了$\mathrm{span}\{S\}$ 的一个基。

接下来，我们将S中的其他向量表示成这个基的线性组合，并给出相应的操作步骤。

步骤四：因为矩阵R的形式比较简单，所以我们可以将R中不包含首1的向量可以写成其他含首1向量的线性组合。

步骤五：根据**定理** 4.7.6的证明，A中与R对应的向量可以写成相同系数关于基底的线性组合。

4.8 秩(rank), 零化度(nullity) 和基本矩阵空间(Fundamental matrix spaces)

定理 4.8.1：矩阵A的行空间和列空间具有相同的维数。

证明：设R是矩阵A的行阶梯型。则由**定理** 4.7.4 和**定理** 4.7.6, 我们知道

$$\text{矩阵}A\text{的行空间的维数} = \text{矩阵}R\text{的行空间的维数，}$$

$$\text{矩阵}A\text{的列空间的维数} = \text{矩阵}R\text{的列空间的维数。}$$

因此，我们只要证明: 矩阵R的行空间的维数 = 矩阵R的列空间的维数。 但是我们注意到矩阵R的行空间的维数等于非零行向量的个数，而矩阵R的列空间的维数等于首1的个数。而两者是相等的，于是矩阵的行空间和列空间的维数相等。 □

有了这个定理，我们可以定义下面的量。

定义：矩阵A的行空间和列空间的维数相等，我们称这个维数是矩阵A的秩(rank), 记成 $\mathrm{rank}(A)$. 矩阵A的零化空间的维数称为A的零化度(nullity), 记成 $nullity(A)$.

设矩阵A是$m \times n$大小的，我们来分析A的秩。注意到A的行向量空间是$\mathbb{R}^n$的子空间，列向量空间是$\mathbb{R}^m$的子空间。于是我们可以知道$\mathrm{rank}(A) \le \min(m, n)$.

定理 4.8.2：设矩阵A有n列，则

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n.$$

证明：将矩阵A变成简约阶梯型R, 则可以知道

$$\mathrm{rank}(A) = \mathrm{rank}(R), \quad \mathrm{nullity}(A) = \mathrm{nullity}(R).$$

对于阶梯型R, 我们知道$\mathrm{rank}(R)$ 恰好是首1向量的个数，而$\mathrm{nullity}(R)$是自由变量的个数。所以

$$\mathrm{rank}(R) + \mathrm{nullity}(R) = n.$$

因此$\mathrm{rank}(A) + \mathrm{nullity}(A) = n$.

定理 4.8.4: （等价性定理）设A是一个n阶方阵，那么下面的命题是等价的：

(a) A 是可逆方阵。

(b) 齐次线性方程组$A\mathbf{x} = 0$只有平凡解。

(c) 方阵A的简约行阶梯型是I_n.

(d) 方阵A可以写成一些初等矩阵的乘积。

(e) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$都是相容的(consistent), 即方程总是有解。

(f) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$只有一个解。

(g) $\det A \neq 0$.

(h) 方阵A的列向量线性无关。

(i) 方阵A的行向量线性无关。

(j) 方阵A的列向量生成$\mathbb{R}^n$.

(k) 方阵A的行向量生成$\mathbb{R}^n$.

(l) 方阵A的列向量是$\mathbb{R}^n$的基。

(m) 方阵A的行向量是$\mathbb{R}^n$的基。

(n) 方阵A的秩是n.

(o) 方阵A的零化度是0.

证明：之前，我们知道(a)到(g) 是等价的，由**定理** 4.8.1和**定理**4.5.4, 我们知道(h) 到(n) 是等价的。由**定理** 4.8.2知(n)和(o)是等价的。而根据定义(b)和(o)是等价的，因此上述定理得证。 □

定理 4.8.5: 设$A\mathbf{x} = \mathbf{b}$是含有m个方程和n个未知数的相容方程，如果方阵A 的秩为r, 则方程组的通解含有$n - r$个参数。

定理 4.8.6: 设A是$m \times n$ 的矩阵。

(a) 当$m > n$, 则线性方程组$A\mathbf{x} = \mathbf{b}$至少对一个向量$\mathbf{b}$是不相容的。

(b) 当$m < n$, 则对于任意的向量$\mathbf{b}$, 线性方程组$A\mathbf{x} = \mathbf{b}$ 或者不相容，或者有无穷多个解。

证明：(a) 假设$m > n$, 矩阵A的秩小于等于n, 从而A的列向量不能生成$\mathbb{R}^m$. 于是存在$\mathbf{b} \in \mathbb{R}^m$ 使得$\mathbf{b}$ 不在A的列向量空间中。这就意味着线性方程组$A\mathbf{x} = \mathbf{b}$对于这样的$\mathbf{b}$没有解。

(b) 当$m < n$时，线性方程组$A\mathbf{x} = \mathbf{b}$只有两种情况，不相容或者相容。在相容的情况下，

$$\mathrm{nullity}(A) = n - \mathrm{rank}(A).$$

而$\mathrm{rank}(A) \leq \min(m, n) = m$, 所以线性方程组有无穷多个解。

下面我们来比较矩阵A以及矩阵A^T的行空间，列空间和零化空间。很容易注意到，A^T的列空间是A的行空间，A^T的行空间是A的列空间。所以对于矩阵A, 我们考虑如下的基本空间：A的行空间，A的列空间，A的零化空间以及A^T的零化空间。下面的定理是显然的。

定理 4.8.7: 对于任意的矩阵A, 有$\mathrm{rank}(A) = \mathrm{rank}(A^T)$.

对于$m \times n$的矩阵A, 根据**定理** 4.8.2, 我们有

$$\mathrm{rank}(A^T) + \mathrm{nullity}(A^T) = m.$$

现在设$\mathrm{rank}(A) = \mathrm{rank}(A^T) = r$, 于是我们可以知道

$$\dim[\mathrm{row}(A)] = \dim[\mathrm{col}(A)] = r,\ \dim[\mathrm{null}(A)] = n - r,\ \dim[\mathrm{null}(A^T)] = m - r.$$

之前，我们提过$\mathbb{R}^n$是一个欧式空间，上面有内积，所以我们有如下的定义。设W是$\mathbb{R}^n$的子空间，那么$\mathbb{R}^n$中与W中每个向量都垂直的向量的集合称为W的正交补，记成$W^\perp$. 下面的定理的证明我们会在后面更一般的情形下给出。

定理 4.8.8: 设W是$\mathbb{R}^n$的子空间，则

(a) $W^\perp$是$\mathbb{R}^n$的子空间。

(b) $W \cap W^\perp = \{\mathbf{0}\}$.

(c) 空间$W^\perp$的正交补是W.

定理 4.8.9: 设A是$m \times n$ 的矩阵，则

(a) 矩阵A的零化空间和它的行空间在$\mathbb{R}^n$中互为正交补。

(b) 矩阵A^T的零化空间和矩阵A的列空间在$\mathbb{R}^m$中互为正交补。

证明：命题(a)即为**定理** 3.4.3, 而命题(b)是命题(a)的简单推论（利用A 和A^T的基本空间的关系）。

4.9 从$\mathbb{R}^n$到$\mathbb{R}^m$的矩阵变换

回顾映射的定义，给定两个集合A和B, 给定一种规则，对于集合A中的一个元素a, 唯一指定集合B 中的一个元素b. 我们记这样的规则为f, 并且有$b = f(a)$, 我们称b是a的像，集合A称为f的定义域，集合B称为f的值域。下面我们研究定义域和值域都是线性空间的映射。设V和W都是线性空间，我们称从V到W的映射f是变换，记为$f : V \to W$. 特别地，当$V = W$时，这样的变换f也称为V上的算子。

在这一节中，我们将考虑从$\mathbb{R}^n$到$\mathbb{R}^m$上的变换。假设$f_1, f_2, \ldots, f_m$是n个变量的实函数，即对于任意的$1 \le i \le m$, 都有$w_i = f_i(x_1, x_2, \ldots, x_m)$. 于是对于$\mathbb{R}^n$中任意的向量$(x_1, x_2, \ldots, x_n)$, 我们指定了$\mathbb{R}^m$中的向量$(w_1, w_2, \ldots, w_m)$. 我们记这样的变换$T : \mathbb{R}^n \to \mathbb{R}^m$为

$$T(x_1, x_2, \ldots, x_n) = (w_1, w_2, \ldots, w_m).$$

如果取上面的函数$f_1, f_2, \ldots, f_m$是线性的，即对于任意的$1 \le i \le m$, 都有

$$w_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n.$$

用矩阵的形式表达出来，有

$$\begin{pmatrix} w_1 \\ w_2 \\ \cdots \\ w_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix},$$

或者直接简写为$\mathbf{w} = A\mathbf{x}$. 我们可以将此等式，视为从$\mathbb{R}^n$中的列向量$\mathbf{x}$映到$\mathbb{R}^m$ 中的列向量$\mathbf{w}$, 称为矩阵变换，记为$T_A : \mathbb{R}^n \to \mathbb{R}^m$, 即通过左乘$A$作用上去。有时对于矩阵变换$T : \mathbb{R}^n \to \mathbb{R}^m$, 直接用记号$[T]$表示相应左乘的矩阵，即有$T(\mathbf{x}) = [T]\mathbf{x}$.

矩阵变换具有如下的性质，这些性质根据矩阵的乘法都是很好证明的。

定理 4.9.1: 给定矩阵A以及线性变换$T_A:\mathbb{R}^n\to\mathbb{R}^m$, 对于任意的$\mathbf{u},\mathbf{v}\in\mathbb{R}^n$以及任意的纯量$\lambda$, 我们有

(a) $T_A(\mathbf{0})=\mathbf{0}$.

(b) $T_A(\lambda\mathbf{u})=\lambda T_A(\mathbf{u})$.

(c) $T_A(\mathbf{u}+\mathbf{v})=T_A(\mathbf{u})+T_A(\mathbf{v})$.

(d) $T_A(\mathbf{u}-\mathbf{v})=T_A(\mathbf{u})-T_A(\mathbf{v})$.

定理 4.9.2: 设$T_A:\mathbb{R}^n\to\mathbb{R}^m$ 以及$T_B:\mathbb{R}^n\to\mathbb{R}^m$ 是矩阵变换，如果对于任意的$\mathbf{x}\in\mathbb{R}^n$有$T_A(x)=T_B(x)$, 则$A=B$.

证明: 由$T_A(x)=T_B(x)$得$A\mathbf{x}=B\mathbf{x}$ 对于任意的$\mathbf{x}\in\mathbb{R}^n$ 都成立，特别地我们取$\mathbb{R}^n$ 中的标准基$\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$, 则

$$A\mathbf{e}_j=B\mathbf{e}_j,(1\le j\le n).$$

我们不难看出$A\mathbf{e}_j$就是矩阵A的第j 列，$B\mathbf{e}_j$就是矩阵B的第j列。从而A和B对应的列相等，于是$A=B$.

现在给定一个矩阵变换T，我们如何找出相对应的矩阵A, 使得$T=T_A$？现在假设$\mathbb{R}^n$的一个基是$\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$, 那么$T_A(\mathbf{e}_j)=A\mathbf{e}_j$, 恰好是$A$的第$j$列。从而我们可以知道

$$A=[T_A(\mathbf{e}_1)\mid T_A(\mathbf{e}_2)\mid\ \ldots\ \mid T_A(\mathbf{e}_n)]$$

教材中介绍了$\mathbb{R}^2$和$\mathbb{R}^3$中很多线性变换的例子，如反射函子，投射函子，旋转函子，扩张和收缩等等。这里不再赘述。

4.10 矩阵变换的性质

我们记T_A是从$\mathbb{R}^n$到$\mathbb{R}^k$的线性映射，T_B是从$\mathbb{R}^k$到$\mathbb{R}^m$的线性映射。作为映射的复合$T_B\circ T_A$是从$\mathbb{R}^n$到$\mathbb{R}^m$的线性映射。于是

$$(T_B\circ T_A)(\mathbf{x})=T_B(T_A(\mathbf{x}))=B(T_A(\mathbf{x}))=B(A\mathbf{x})=(BA)(\mathbf{x}),$$

而这意味着$T_B\circ T_A=T_{BA}$.

一对一的矩阵变换(one-to-one matrix transformation): 一个矩阵变换$T_A:\mathbb{R}^n\to\mathbb{R}^m$称为是一对一的矩阵变换，如果$T_A$将$\mathbb{R}^n$中不同的向量映到$\mathbb{R}^m$中不同的向量。

定理 4.10.1: 设A是一个n阶方阵，记$T_A:\mathbb{R}^n\to\mathbb{R}^n$是相应的矩阵算子，则下列的命题等价：

(a) A是可逆方阵。

(b) T_A的值域是$\mathbb{R}^n$.

(c) T_A是一对一的映射。

证明: 我们推导$(a)\Rightarrow(b)\Rightarrow(c)\Rightarrow(a)$.

$(a)\Rightarrow(b)$ 假设A可逆，我们考虑方程$A\mathbf{x}=\mathbf{b}$, 对于任意的$\mathbf{b}$, 此方程都有解。这就意味着T_A的值域是$\mathbb{R}^n$.

$(b) \Rightarrow (c)$ 根据T_A的值域是$\mathbb{R}^n$, 我们知道$\forall b \in \mathbb{R}^n$, 存在$x \in \mathbb{R}^n$, 使得$A\mathbf{x} = \mathbf{b}$. 由**定理** 4.8.4中$(e)$和$(f)$的等价性，我们知道方程$A\mathbf{x} = \mathbf{b}$有唯一的解，这就知道$T_A$是一对一的映射。

$(c) \Rightarrow (a)$ 因为T_A是一对一的映射，则对于T_A的值域中的任意的向量$\mathbf{b}$, 我们知道存在唯一的$\mathbf{x}$, 使得$T_A(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$. 现在假设$A$不是可逆的，于是方程$A\mathbf{x} = \mathbf{0}$有非零解，而$\mathbf{0}$ 是在T_A的值域中的。因此，我们推出矛盾，这说明A是可逆的。 □

对于一对一的函子$T_A : \mathbb{R}^n \to \mathbb{R}^n$, 则根据上面的定理知道$A$是可逆的。因此函子$T_{A^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$ 是函子T_A的逆函子，这是因为

$$T_A \circ T_{A^{-1}} = T_{AA^{-1}} = T_I,\ \ T_{A^{-1}} \circ T_A = T_{A^{-1}A} = T_I.$$

接下来我们考虑这样的问题：对于一个映射$T : \mathbb{R}^n \to \mathbb{R}^m$, 什么时候$T$是一个矩阵变换？答案如下：

定理 4.10.2: 设$T : \mathbb{R}^n \to \mathbb{R}^m$ 是一个矩阵变换当且仅当下面的性质对于任意的$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ 以及纯量λ, 都有：

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.

(ii) $T(\lambda\mathbf{u}) = \lambda T(\mathbf{u})$.

证明: 如果T是一个矩阵映射，则自然地容易得出来(i)和(ii).

反过来，假设映射$T : \mathbb{R}^n \to \mathbb{R}^m$ 满足(i)和(ii). 于是对于$\mathbb{R}^n$中的列向量

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n,$$

我们有

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n).$$

如果我们记矩阵A为

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2) \mid \ \ldots \ \mid T(\mathbf{e}_n)],$$

那么即有$T(\mathbf{x}) = A\mathbf{x}$. □

满足上述定理中的条件(i) 和(ii)的映射，我们就称为线性映射。由上述定理知，对于从$\mathbb{R}^n$到$\mathbb{R}^m$的映射，线性映射和矩阵变换是等价的。即:

定理 4.10.3: 从从$\mathbb{R}^n$ 到$\mathbb{R}^m$的线性变换都是矩阵变换，反过来，任意的矩阵变换也是线性变换。

最后我们补充之前等价定理的另外几个等价论述。

定理 4.10.4: （等价性定理）设A是一个n阶方阵，那么下面的命题是等价的：

(a) A 是可逆方阵。

(b) 齐次线性方程组$A\mathbf{x} = 0$只有平凡解。

(c) 方阵A的简约行阶梯型是I_n.

(d) 方阵A可以写成一些初等矩阵的乘积。

(e) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$都是相容的(consistent), 即方程总是有解。

(f) 对于任意的$n \times 1$的列向量$\mathbf{b}$, 方程$A\mathbf{x} = \mathbf{b}$只有一个解。

(g) $\det A \neq 0$.

(h) 方阵A的列向量线性无关。

(i) 方阵A的行向量线性无关。

(j) 方阵A的列向量生成$\mathbb{R}^n$.

(k) 方阵A的行向量生成$\mathbb{R}^n$.

(l) 方阵A的列向量是$\mathbb{R}^n$的基。

(m) 方阵A的行向量是$\mathbb{R}^n$的基。

(n) 方阵A的秩是n.

(o) 方阵A的零化度是0.

(p) A的零化空间的正交补是$\mathbb{R}^n$.

(q) A的行空间的正交补是$\{\mathbf{0}\}$.

(r) T_A的值域是$\mathbb{R}^n$.

(s) T_A是一对一的映射。

4.11 $\mathbb{R}^2$上的矩阵算子对应的几何

$\mathbb{R}^2$上的矩阵算子对于电脑作图是非常重要的。我们已经有一些常见的矩阵算子如反射算子，旋转算子，伸缩算子等等。在本小节中，我们着重考虑$\mathbb{R}^2$上的一对一的矩阵算子。设T_A是一对一的算子，从而矩阵A是可逆矩阵，可以写成一些初等矩阵的乘积，即有$A = E_1 E_2 \dots E_r$, 从而

$$T_A = T_{E_1} \circ T_{E_2} \circ \cdots \circ T_{E_r}.$$

因此我们首先考虑对于一个初等矩阵，相应的矩阵算子的几何性质。

定理 4.11.1: 设E是一个初等矩阵，则$T_E : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$是下列其中一种形式：

(a) 沿着某个坐标轴的斜移。

(b) 关于直线$y = x$的反射。

(c) 沿着某个坐标轴的压缩(compression)。

(d) 沿着某个坐标轴的伸展(expansion)。

(e) 关于某个坐标轴的反射。

(f) 先沿着某个坐标轴伸缩，再关于某个坐标轴反射。

证明: 2×2的初等矩阵都有下面的形式：

$$\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}, \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}.$$

前面两种矩阵对应着沿着坐标轴的斜移。第三个矩阵是沿着直线$y = x$的反射。如果$0 < k < 1$, 后面的两种矩阵是沿着某个坐标轴的压缩(dilation)。如果$k > 1$, 后面的两种矩阵是沿着某个坐标轴的伸展(expansion)。如果$k < 0$, 则我们知道

$$\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -k & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -k \end{pmatrix}.$$

特别地，当$k=-1$时，矩阵算子就是沿着某个坐标轴的反射。如果$k \neq -1$时，矩阵算子就是先沿着某个坐标轴伸缩，再沿着某个坐标轴反射。 □

由上面的定理，我们知道

定理 4.11.2: 设$T_A : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$是左乘可逆矩阵$A$的矩阵算子，则$T_A$从几何上可以看出一些斜移，压缩，伸展以及反射按照某种合适顺序的复合。

在计算机画图中，许多图像都是由点连成线而构成的。因此，有必要考虑直线在矩阵算子下的像。

定理 4.11.3: 设$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$是左乘可逆矩阵$A$的矩阵算子，则

(a) 直线的像还是直线。

(b) 过原点的直线的像还是过原点的直线。

(c) 平行直线的像还是平行直线。

(d) 连接两点P和Q的直线的像是连接$T(P)$和$T(Q)$的直线。

(e) 三个点的像在一条直线上当且仅当三个点本身在一条直线上。

证明: (a) 设矩阵

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

是可逆矩阵，即$ad-bc \neq 0$. 设直线的参数方程是

$$(x, y) = (x_0, y_0) + t(u, v),$$

其中$(u, v) \neq (0, 0)$. 于是直线的像是

$$(x', y') = T(x, y) = T(x_0, y_0) + tT(u, v),$$

即我们有参数方程

$$(x', y') = (ax_0 + by_0, cx_0 + dy_0) + t(au + bv, cu + dv),$$

且$(au + bv, cu + dv) \neq 0$, 为一条直线。

(b) 根据(a), 取$(x_0, y_0) = (0, 0)$, 容易知道直线过原点时，其像也过原点。

(c) 两条平行的直线，其参数方程可以分别设为

$$(x, y) = (x_1, y_1) + t(u, v), \quad (x, y) = (x_2, y_2) + t(u, v)$$

根据(a), 它们的像对应的参数方程分别是

$$(x', y') = T(x_1, y_1) + tT(u, v), \quad (x', y') = T(x_2, y_2) + tT(u, v),$$

仍然是平行的。

(d) 平面上两点确定一条直线，由(a)可证明(d).

(e) 当三个点在一条直线上时，由(a)可知它们的像也在一条直线上，而根据T是可逆算子知道反过来的结论也是成立的。

5. 特征值(eigenvalue)和特征向量(eigenvector)

5.1 特征值和特征向量

定义：设A是一个n阶方阵，则$\mathbb{R}^n$中的非零向量 x 称为A的特征向量，如果存在纯量λ, 使得

$$A\mathbf{x} = \lambda \mathbf{x}.$$

此时，纯量λ是A的特征值(eigenvalue)，$\mathbf{x}$是对应到特征值λ的特征向量(eigenvector).

现在我们来计算n阶方阵A的特征值和特征向量。注意到$A\mathbf{x} = \lambda \mathbf{x} = \lambda I \mathbf{x}$, 或者等价地，

$$(\lambda I - A)\mathbf{x} = \mathbf{0}.$$

注意到λ是特征值，当且仅当上述方程有非零解，而这等价于方阵$\lambda I - A$的行列式为0.

定理 5.1.1: 设A是一个n阶方阵，λ是方阵A的特征值当且仅当

$$\det(\lambda I - A) = 0.$$

这就是矩阵A的特征式(characteristic equation).

当我们将行列式$\det(\lambda I - A)$展开后，会得到一个次数为n的多项式，称为方阵A的特征多项式。一般地，我们可以将一个n阶矩阵的特征多项式写成下面的形式

$$p(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_n.$$

对于实数上的多项式，次数为n的多项式至多有n个不同的根，于是方程

$$\lambda^n + c_1 \lambda^{n-1} + \cdots + c_n = 0$$

至多有n个不同的解，这就说明n阶方阵至多有n 个不同的特征值。注意到实数多项式可以有复数的根，所以一个实方阵是可以有复的特征值的。

对于一个三角矩阵A，它的特征多项式的形式是比较简单的。因而，很容易知道A的特征值就是A的主对角线上的元素。

定理 5.1.3: 设A是一个n阶方阵，下面的形式是等价的：

(a) λ是A的特征值。

(b) 线性方程组$(\lambda I - A)\mathbf{x} = \mathbf{0}$有非平凡的解。

(c) 存在非零的向量$\mathbf{x}$使得$A\mathbf{x} = \lambda\mathbf{x}$.

(d) λ是特征等式$\det(\lambda I - A) = \mathbf{0}$的根。

对于矩阵A的特征值是λ, 我们考虑方程

$$(\lambda I - A)\mathbf{x} = \mathbf{0},$$

因此特征向量是矩阵$\lambda I - A$的零化空间中的非零向量，我们称此零化空间是对应于特征值λ 的特征空间(eigenspace).

对于方阵A, 记λ是它的特征值，$\mathbf{x}$是对应的特征向量，则

$$A^k\mathbf{x} = A^{k-1}(A\mathbf{x}) = A^{k-1}(\lambda\mathbf{x}) = \lambda(A^{k-1}\mathbf{x}) = \cdots = \lambda^k\mathbf{x}$$

这就说明了λ^k是A^k的特征值，$\mathbf{x}$是相应的特征向量。

下面的定理揭示了矩阵可逆和特征值之间的关系。

定理 5.1.5: 一个方阵A可逆当且仅当0不是A的特征值。

证明：特征值是矩阵A的特征多项式对应方程

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

的根，因此0不是A的特征值当且仅当$c_n \neq 0$. 所以下面我们证明方阵A可逆当且仅当$c_n \neq 0$. 但是我们注意到在

$$\det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$$

中令$\lambda = 0$, 我们有$\det(-A) = (-1)^n \det A = c_n$. 因此$\det A \neq 0$ 当且仅当$c_n \neq 0$, 而这就又等价于方阵A是可逆的。而这又给我们的等价性定理增添了一个等价条件。 □

5.2 对角化

首先我们来介绍矩阵的相似：设A和B是两个方阵，则我们称B对于A (B is similar to A)是相似的，如果存在可逆矩阵P 使得$B = P^{-1}AP$. 当B 对于A是相似的时候，我们知道A对于B也是相似的，这是因为A也可以表达成$A = Q^{-1}BQ$, 这里$Q = P^{-1}$. 此时，我们就说A和B是相似的。

对于相似的两个矩阵A和B, $B = P^{-1}AP$, 它们有很好的相似不变量。下面我们列举其中的一些，关于它们的证明，我们只需要证明一个方向就可以，因为相似具有对称性。

(1) 矩阵A和矩阵$P^{-1}AP$有相同的行列式。这是因为

$$\det(P^{-1}AP) = \det(P^{-1}) \det A \det P = \frac{1}{\det P} \det A \det P = \det A.$$

(2) 矩阵A可逆当且仅当$P^{-1}AP$也是可逆的。由它们的行列式相等易知。

(3) 矩阵A和矩阵$P^{-1}AP$的秩相同。对于矩阵A, 左乘可逆矩阵等价于做一些初等行变换，右乘可逆矩阵等价于做一些初等列变换，因为做初等行变换和列变换不改变矩阵的秩，因此我们有A和矩阵$P^{-1}AP$的秩相同。

(4) 矩阵A和矩阵$P^{-1}AP$的零化度相同。对一个方阵A, 我们有

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n.$$

根据(3), 矩阵A和矩阵$P^{-1}AP$的秩相同，因此他们的零化度也相同。

(5) 矩阵A和矩阵$P^{-1}AP$的迹相同。这是因为

$$\mathrm{tr}(P^{-1}AP) = \mathrm{tr}(APP^{-1}) = \mathrm{tr}(A).$$

(6) 矩阵A和矩阵$P^{-1}AP$的特征多项式相等。这是因为方阵$P^{-1}AP$的特征多项式为

$$\det(\lambda I - P^{-1}AP) = \det(P^{-1}(\lambda I - A)P) = \det(\lambda I - A),$$

恰好是A的特征多项式。

(7) 矩阵A和矩阵$P^{-1}AP$有相同的特征值。由(6)可以推出。

(8) 设λ是矩阵A和矩阵$P^{-1}AP$的特征值，则矩阵A对应到λ的特征空间和矩阵$P^{-1}AP$对应到λ的特征空间的维数相同。设$\mathbf{x}$是A对应到λ 的一个特征向量，即$A\mathbf{x} = \lambda\mathbf{x}$. 我们很容易验证$P^{-1}\mathbf{x}$是$P^{-1}AP$对应到$\lambda$的一个特征向量，这是因为

$$P^{-1}AP(P^{-1}\mathbf{x}) = P^{-1}A\mathbf{x} = P^{-1}(\lambda\mathbf{x}) = \lambda(P^{-1}\mathbf{x}).$$

设$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是矩阵$A$对应到$\lambda$的特征空间的一组基，注意到$P^{-1}$ 是可逆方阵，于是

$$\{P^{-1}\mathbf{v}_1, P^{-1}\mathbf{v}_2, \ldots, P^{-1}\mathbf{v}_r\}$$

是矩阵$P^{-1}AP$对应到λ 的特征空间的一组基。

一个方阵A被称为可对角化的(diagonalizable)，如果它相似到一个对角矩阵。即，存在可逆矩阵P使得$P^{-1}AP$是对角矩阵。

定理 5.2.1: 设A是n阶方阵，则下面的命题等价：

(a) A是可对角化的。

(b) A有n个线性无关的特征向量。

证明：$(a) \Longrightarrow (b)$ 现在假设A是可对角化的，即存在可逆矩阵P和对角矩阵D使得$P^{-1}AP = D$, 或者等价地$AP = PD$. 现在我们记方阵P的列向量是$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, 于是

$$AP = A\,[\mathbf{p}_1 \mid \mathbf{p}_2 \mid \ \ldots\ \mid \mathbf{p}_n] = [A\mathbf{p}_1 \mid A\mathbf{p}_2 \mid \ \ldots\ \mid A\mathbf{p}_n].$$

我们记对角矩阵D的对角线上的元素是$\lambda_1, \lambda_2, \ldots, \lambda_n$, 从而

$$PD = [\lambda_1 \mathbf{p}_1 \mid \lambda_2 \mathbf{p}_2 \mid \ \ldots \ \mid \lambda_n \mathbf{p}_n].$$

从而，我们知道了$A\mathbf{p}_k = \lambda_k \mathbf{p}_k, (1 \leq k \leq n)$. 由于矩阵$P$是可逆的方阵，从而列向量$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$是线性无关的。于是我们得到了$A$的$n$个线性无关的特征向量。

$(b) \Longrightarrow (a)$ 假设A有n个线性无关的特征向量，记为$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, 且$\lambda_1, \lambda_2, \ldots, \lambda_n$是相应的特征值。我们令矩阵

$$P = [\mathbf{p}_1 \mid \mathbf{p}_2 \mid \ \ldots \ \mid \mathbf{p}_n],$$

从而矩阵P是可逆的矩阵。我们令D是对角线分别为$\lambda_1, \lambda_2, \ldots, \lambda_n$的对角矩阵。从而类似于之前的运算，我们有$AP = PD$. 从而$P^{-1}AP = D$, 这就说明矩阵$A$是可对角化的。 □

上面的定理给了我们一个对矩阵对角化操作的方法：

步骤一：确定矩阵A是否是有n个线性无关的特征向量。我们可以对每一个特征值，找出它的特征空间的一个基，将所有的特征空间的基合并成一个集合S. 如果此集合少于n个向量，则矩阵A不能被对角化。

步骤二：如果找到了n个线性无关的特征向量$S = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$, 将它们放在一起，形成矩阵$P$.

步骤三：矩阵$P^{-1}AP$是对角矩阵，它的元素是对应到特征向量$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$的特征值

$$\lambda_1, \lambda_2, \ldots, \lambda_n.$$

定理 5.2.2: 设$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$是方阵$A$的对应于不同的特征值的特征向量，则$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$是线性无关的。

证明：设$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$是方阵$A$的对应于不同的特征值的特征向量，然后我们假设$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$是线性相关的，然后退出矛盾。

首先注意到$\{\mathbf{v}_1\}$不为0, 所以$\{\mathbf{v}_1\}$是线性无关的。我们假设r是最大的正整数，使得$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$是线性无关的，根据假设，我们知道此时$1 \leq r < k$. 根据$r$的定义，我们知道$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}$是线性相关的，于是存在不全为0 的$c_1, c_2, \ldots, c_{r+1}$, 使得

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{r+1}\mathbf{v}_{r+1} = \mathbf{0}.$$

因为对于$1 \leq i \leq r+1$, 我们知道$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, 于是

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0}.$$

用第二个等式减去第一个等式的λ_{r+1}倍，得到

$$c_1(\lambda_1 - \lambda_{r+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{r+1})\mathbf{v}_2 + \cdots + c_r(\lambda_r - \lambda_{r+1})\mathbf{v}_r = \mathbf{0}.$$

而$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ 是线性无关的，因此我们知道

$$c_1(\lambda_1 - \lambda_{r+1}) = c_2(\lambda_2 - \lambda_{r+1}) = \cdots = c_r(\lambda_r - \lambda_{r+1}) = 0,$$

而$\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$彼此不相等，于是必然有

$$c_1 = c_2 = \cdots = c_r = 0,$$

于是$c_{r+1}\mathbf{v}_{r+1} = \mathbf{0}$, 则有$c_{r+1} = 0$. 这与$c_1, c_2, \ldots, c_{r+1}$不全为0矛盾，命题得证。 □

定理 5.2.3: 设方阵A有n个不同的特征值，则A是可对角化的。

证明： 设$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是对应于不同的特征值$\lambda_1, \lambda_2, \ldots, \lambda_n$ 的特征向量，则根据**定理** 5.2.2, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是线性无关的，因此由**定理** 5.2.1知，$A$是可对角化的。 □

很多时候我们需要求方阵A的幂次，但是一般而言，这是相对困难的。但是对于可对角化的方阵A, 我们是可以求出它的幂次的。现在假设A是可对角化的方阵，则存在方阵P使得

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = D$$

两边同时取平方，则有

$$(P^{-1}AP)^2 = P^{-1}APP^{-1}AP = P^{-1}A^2P = D^2.$$

一般地，对于一个整数k, 我们有$(P^{-1}AP)^k = P^{-1}A^kP = D^k$. 而

$$D^k = \begin{pmatrix} \lambda_1^k & 0 & \ldots & 0 \\ 0 & \lambda_2^k & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \lambda_n^k \end{pmatrix},$$

于是，最终我们得到

$$A^k = PD^kP^{-1} = P\begin{pmatrix} \lambda_1^k & 0 & \ldots & 0 \\ 0 & \lambda_2^k & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & \lambda_n^k \end{pmatrix}P^{-1}$$

□

对于一个方阵A, 如果它的特征值λ和相应的特征向量$\mathbf{x}$都已经知道了，那么我们很容易就可以求得A 的任一正整数此幂A^k(k是正整数)的特征值和特征向量，这是因为

$$A^k\mathbf{x} = A^{k-1}(A\mathbf{x}) = A^{k-1}(\lambda\mathbf{x}) = \lambda(A^{k-1}\mathbf{x}) = \cdots = \lambda^k\mathbf{x}.$$

现在，我们来考虑一个方阵A可对角化的充分必要条件。为此，我们先引入几何重数(geometric multiplicity)和代数重数(algebraic multiplicity) 的概念。设λ_0是方阵A的一个特征值，则相对应于λ_0的特征空间的维数被称为λ_0的几何重数。同时，我们注意到$\lambda-\lambda_0$是方阵A的特征多项式的一个因子，此因子的次数λ_0的代数重数。

下面的定理给出了A可对角化的充分必要条件，我们略去证明。

定理 5.2.5: 设A是一个方阵，则

(a) 对于A的每一个特征值，此特征值的几何重数总是小于等于代数重数的。

(b) 方阵A可对角化的充分必要条件是A 的每一个特征值的几何重数和代数重数都是相等的。

5.3 复向量空间

在上一节中，我们提到对于一个实的方阵，当考虑它的特征值的时候，有可能是复的。为此，我们有必要考虑复的向量空间。首先我们来看复向量空间$\mathbb{C}^n$. 设n是一个正整数，一个n复数组就是n个复数排成一起$(v_1, v_2, \ldots, v_n)$. 所有的复数组形成了一个复空间$\mathbb{C}^n$.

复空间$\mathbb{C}^n$的向量的加法和数乘与$\mathbb{R}^n$上的加法和数乘相似，只不过此时纯量是复数而已。对于复空间$\mathbb{C}^n$的向量$\mathbf{v} = (v_1, v_2, \ldots, v_n)$, 每一个$v_i$都是复的分量，可以写成

$$\mathbf{v} = (a_1 + ib_1, a_2 + ib_2, \ldots, a_n + ib_n) = (a_1, a_2, \ldots, a_n) + i(b_1, b_2, \ldots, b_n).$$

同时，也可以写成$\mathbf{v} = \mathrm{Re}(\mathbf{v}) + i\ \mathrm{Im}(\mathbf{v})$, 这里

$$\mathrm{Re}(\mathbf{v}) = (a_1, a_2, \ldots, a_n), \quad \mathrm{Im}(\mathbf{v}) = (b_1, b_2, \ldots, b_n).$$

我们可以考虑对向量$\mathbf{v}$取共轭，得到

$$\overline{\mathbf{v}} = (a_1 - ib_1, a_2 - ib_2, \ldots, a_n - ib_n) = (a_1, a_2, \ldots, a_n) - i(b_1, b_2, \ldots, b_n) = \mathrm{Re}(\mathbf{v}) - i\ \mathrm{Im}(\mathbf{v}).$$

对于一个矩阵A, 我们称A是一个实矩阵，如果A 里面的元素都是实数，称A是复方阵，如果A里面的元素允许是复数。对于一个复矩阵A, $\mathrm{Re}(A)$和$\mathrm{Im}(A)$分别是矩阵A的实部和虚部。矩阵$\overline{A}$是由A 的项取共轭所得。

下面的两个定理的证明是容易的，留作练习。

定理 5.3.1: 设$\mathbf{u}$和$\mathbf{v}$是$\mathbb{C}^n$中的复向量，λ是一个纯量，则

(a) $\overline{\overline{\mathbf{u}}} = \mathbf{u}$.

(b) $\overline{\lambda\mathbf{u}} = \overline{\lambda}\overline{\mathbf{u}}$.

(c) $\overline{\mathbf{u}+\mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}}$.

(d) $\overline{\mathbf{u}-\mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}}$.

定理 5.3.2:

设A是一个$m \times k$的复矩阵，B是一个$k \times n$的复矩阵，则

(a) $\overline{\overline{A}} = A$.

(b) $\overline{A^T} = (\overline{A})^T$.

(c) $\overline{AB} = \overline{A}\ \overline{B}$.

之前，我们在实空间$\mathbb{R}^n$上定义过点乘，现在我们在复空间$\mathbb{C}^n$上来定义点乘。设

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \quad \mathbf{v} = (v_1, v_2, \ldots, v_n)$$

是$\mathbb{C}^n$上的两个向量，则$\mathbf{u}$和$\mathbf{v}$的复的欧式内积(复点乘)，记为$\mathbf{u}\cdot\mathbf{v}$, 定义成

$$\mathbf{u}\cdot\mathbf{v} = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n}.$$

同时，我们定义它的欧式范数是

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}.$$

类似于之前实空间的情形，我们称$\mathbf{v}\in\mathbb{C}^n$是单位向量，如果$\|v\| = 1$. 称两个向量$\mathbf{u}$ 和$\mathbf{v}$正交，如果$\mathbf{u}\cdot\mathbf{v} = 0$.

对于$\mathbb{C}^n$中的向量$\mathbf{u}$和$\mathbf{v}$, 点乘$\mathbf{u}\cdot\mathbf{v}$可以写成

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\overline{\mathbf{v}}.$$

在实空间$\mathbb{R}^n$中，点乘是对称的，但是在复空间中却不是这样的，下面的定理根据定义是很好直接验证的。

定理 5.3.3: 设$\mathbf{u}$, $\mathbf{v}$和$\mathbf{w}$是$\mathbb{C}^n$中的向量，λ是一个纯量，则复欧式内积有下列的性质：

(a) $\mathbf{u}\cdot\mathbf{v} = \overline{\mathbf{v}\cdot\mathbf{u}}$.

(b) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v} + \mathbf{u}\cdot\mathbf{w}$

(c) $\lambda(\mathbf{u}\cdot\mathbf{v}) = (\lambda\mathbf{u})\cdot\mathbf{v}$.

(d) $\mathbf{u}\cdot(\lambda\mathbf{v}) = \overline{\lambda}(\mathbf{u}\cdot\mathbf{v})$

(e) $\mathbf{v}\cdot\mathbf{v}\geq 0$，同时有$\mathbf{v}\cdot\mathbf{v} = 0$ 当且仅当$\mathbf{v} = \mathbf{0}$.

和之前一样，对于一个复方阵，特征等式$\det(\lambda I - A) = 0$的根是方阵A的复特征值。而且，λ是方阵A 的特征值当且仅当存在$\mathbb{C}^n$中的非零向量$\mathbf{x}$ 使得$A\mathbf{x} = \lambda\mathbf{x}$, 这样的$x$也称为$A$对应到特征值$\lambda$的复特征向量，它们都是线性方程组$(\lambda I - A)\mathbf{x} = \mathbf{0}$的非零解，此方程的解空间是$\mathbb{C}^n$的一个子空间，称为$A$对应到特征值$\lambda$的特征空间。

下面的定理告诉我们，对于一个实方阵而言，它如果有复的特征值，则它的复特征值都是共轭出现的。

定理 5.3.4: 设λ是实方阵A的特征值，$\mathbf{x}$是对应的特征向量，于是$\overline{\lambda}$也是A的特征值，$\overline{\mathbf{x}}$是对应的特征向量。

证明：因为λ是实方阵A的特征值，$\mathbf{x}$是对应的特征向量，我们有

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\ \overline{\mathbf{x}}.$$

另一方面，由于A是实方阵，所以$\overline{A}=A$, 从而

$$\overline{A\mathbf{x}}=\overline{A}\ \overline{\mathbf{x}}=A\overline{\mathbf{x}},$$

则$A\overline{\mathbf{x}}=\overline{\lambda}\ \overline{\mathbf{x}}$, 这就说明$\overline{\lambda}$也是$A$的特征值，$\overline{\mathbf{x}}$是对应的特征向量。 □

对一个对称的实方阵而言，它的特征值总是实数。即有下面的定理：

定理 5.3.6: 设A是一个实的对称方阵，则A的特征值为实数。

证明：设λ是A的特征值，$\mathbf{x}$是对应的特征向量，则$A\mathbf{x}=\lambda\mathbf{x}$. 对此等式，同时左乘$\overline{\mathbf{x}}^T$, 有

$$\overline{\mathbf{x}}^TA\mathbf{x}=\overline{\mathbf{x}}^T\lambda\mathbf{x}=\lambda\overline{\mathbf{x}}^T\mathbf{x}=\lambda(\mathbf{x}\cdot\mathbf{x})=\lambda\|\mathbf{x}\|^2.$$

注意到$\mathbf{x}\neq 0$, 于是$\|\mathbf{x}\|^2>0$. 从而

$$\lambda=\frac{\overline{\mathbf{x}}^TA\mathbf{x}}{\|\mathbf{x}\|^2}.$$

接下来，我们只要说明$\overline{\mathbf{x}}^TA\mathbf{x}$是实数即可，即有

$$\overline{\overline{\mathbf{x}}^TA\mathbf{x}}=\overline{\mathbf{x}}^TA\mathbf{x}.$$

这是因为

$$\overline{\overline{\mathbf{x}}^TA\mathbf{x}}=\mathbf{x}^T\overline{A\mathbf{x}}=(\overline{A\mathbf{x}})^T\mathbf{x}=(A\overline{\mathbf{x}})^T\mathbf{x}=\overline{\mathbf{x}}^TA^T\mathbf{x}=\overline{\mathbf{x}}^TA\mathbf{x}.$$

□

下面，我们来考虑2阶实方阵的复特征值的一些性质。

定理 5.3.7: 设实方阵

$$C=\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

则其特征值是$\lambda=a\pm bi$.如果a,b不全为0, 则此矩阵可以分解成

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}=\begin{pmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{pmatrix}\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix},$$

这里的 ϕ 是从正x轴到点(a,b)与原点形成的线段的夹角。

证明：计算方阵C的特征多项式为$(\lambda-a)^2+b^2=0$, 于是特征值是$\lambda=a\pm bi$. 于是我们可以写成

$$a=|\lambda|\cos\phi,\qquad b=|\lambda|\sin\phi$$

这里的 ϕ 是从正x轴到点(a,b)与原点形成的线段的夹角。从而我们得到

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}=\begin{pmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{pmatrix}\begin{pmatrix} \frac{a}{|\lambda|} & -\frac{b}{|\lambda|} \\ \frac{b}{|\lambda|} & \frac{a}{|\lambda|} \end{pmatrix}=\begin{pmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{pmatrix}\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}.$$

根据此等式，我们知道定理中的方阵C可以看成一个旋转变换和一个伸缩变换的复合。 □

定理 5.3.8: 设A是一个2阶实方阵，且其特征值是$\lambda = a \pm bi$ $(b \neq 0)$. 如果$\mathbf{x}$是A对应到特征值$\lambda = a - bi$的特征向量，则矩阵$P = [\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})]$是可逆的且

$$A = P \begin{pmatrix} a & -b \\ b & a \end{pmatrix} P^{-1}.$$

证明：根据条件，我们有$A\mathbf{x} = \lambda\mathbf{x}$, 两边同时取共轭，我们得到$A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. 而$\lambda = a - bi$ 满足$b \neq 0$，于是$\lambda \neq \overline{\lambda}$, 从而$\{\mathbf{x}, \overline{\mathbf{x}}\}$ 线性无关，因此不难发现$\{\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})\}$ 也是线性无关的，这等价于矩阵$P = [\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})]$是可逆的。另外，由$A\mathbf{x} = (a - bi)\mathbf{x}$ 我们知道

$$A\mathrm{Re}(\mathbf{x}) = a\mathrm{Re}(\mathbf{x}) + b\mathrm{Im}(\mathbf{x}), \quad A\mathrm{Im}(\mathbf{x}) = -b\mathrm{Re}(\mathbf{x}) + a\mathrm{Im}(\mathbf{x}).$$

将上面的等式写成矩阵乘法的形式，有

$$A[\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})] = [\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})] \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

从而得到

$$A = P \begin{pmatrix} a & -b \\ b & a \end{pmatrix} P^{-1}.$$

□

结合**定理** 5.3.7和**定理** 5.3.8, 我们知道任意的一个二阶方阵可以有一个几何的解释。对于满足**定理** 5.3.8中的条件的矩阵A, 我们写成

$$A = PSR_\phi P^{-1} = P \begin{pmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} P^{-1},$$

如果我们将矩阵P看成基底$B = (\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x}))$相对于标准基底的转换矩阵，从而$A\mathbf{x}_0$可以看成按下面步骤的几何操作：

步骤一：由标准基坐标化的向量$\mathbf{x}_0$看成由基底B坐标化，得到新的坐标为$P^{-1}\mathbf{x}_0$.

步骤二：对于向量$P^{-1}\mathbf{x}_0$ 做旋转和伸缩的操作，得到$SR_\phi P^{-1}\mathbf{x}_0$.

步骤三：对步骤二中的向量再由标准基坐标化，得到向量$A\mathbf{x}_0 = PSR_\phi P^{-1}\mathbf{x}_0$.

6. 内积空间(Inner Product Spaces)

6.1 内积空间

之前，我们在$\mathbb{R}^n$中定义过点乘，并在此基础上引入了，长度，距离，角度等等概念。这一小节中，我们将会把这些概念推广到一般的实向量空间中。

在一个实向量空间V中，对于其中任意的向量$\mathbf{u}, \mathbf{v}$, 我们给一个配对$\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{R}$, 此配对称为是内积，如果满足下面的条件：

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$. （对称公理）

2. $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$. （加性公理）

3. $\langle \lambda\mathbf{u}, \mathbf{v}\rangle = \lambda\langle \mathbf{u}, \mathbf{v}\rangle$. （齐次公理）

4. $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, 同时$\langle \mathbf{v}, \mathbf{v}\rangle = 0$ 当且仅当$\mathbf{v} = \mathbf{0}$. （正性公理）

在$\mathbb{R}^n$中，对于向量

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \quad \mathbf{v} = (v_1, v_2, \ldots, v_n)$$

点乘定义的内积为

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + u_2v_2 + \ldots u_nv_n,$$

这样的内积称为欧氏内积，因此将带有此内积的实空间$\mathbb{R}^n$称为n维欧氏空间。和之前的点乘类似，我们定义内积空间中的长度和距离如下：对于实的内积空间V, 向量$\mathbf{v}$的范数，记为$\|\mathbf{v}\|$, 定义成

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}.$$

两个向量$\mathbf{u}$和$\mathbf{v}$的距离定义为

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = \sqrt{\langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle}.$$

范数为1的向量称为单位向量。和之前点乘所定义的长度，距离类似，我们有下面的定理，其证明是容易的。

定理 6.1.1: 设$\mathbf{u}$和$\mathbf{v}$是实内积空间V中的向量，λ是纯量，则

(a) $\|\mathbf{v}\| \geq 0$, 等号成立当且仅当$\mathbf{v} = \mathbf{0}$.

(b) $\|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\|$.

(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.

(d) $d(\mathbf{u}, \mathbf{v}) \geq 0$, 等号成立当且仅当$\mathbf{u} = \mathbf{v}$

有时，我们也考虑带权的内积。设$w_1, w_2, \ldots, w_n$是一些正实数，我们称之为权(weights). 设

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \quad \mathbf{v} = (v_1, v_2, \ldots, v_n)$$

是$\mathbb{R}^n$中的两个向量，我们定义带有权$w_1, w_2, \ldots, w_n$的欧氏内积为

$$\langle \mathbf{u}, \mathbf{v}\rangle = w_1u_1v_1 + w_2u_2v_2 + \cdots + w_nu_nv_n.$$

无论是欧氏内积还是带权的欧氏内积，都是更广泛的一种内积，称为矩阵内积(matrix inner product)的一些特殊情形。设A是一个n阶可逆方阵，对于$\mathbb{R}^n$中的两个列向量$\mathbf{u}$和$\mathbf{v}$, 我们可以定义内积

$$\langle \mathbf{u}, \mathbf{v}\rangle = A\mathbf{u} \cdot A\mathbf{v},$$

这样的内积称为是$\mathbb{R}^n$中由A生成的内积。注意到点乘$\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T\mathbf{u}$, 从而

$$\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{v})^T A\mathbf{u} = \mathbf{v}^T A^T A\mathbf{u}.$$

很容易验证，当$A = I$单位矩阵的时候，上述定义的是标准内积。当

$$A = \begin{pmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{w_n} \end{pmatrix}$$

定义的内积是带有权$w_1, w_2, \dots, w_n$的内积。

接下来，我们来举一些一般线性空间上内积的例子。

例1：在n阶方阵空间M_n中，对于两个方阵U和V, 我们定义内积

$$\langle U, V\rangle = \mathrm{tr}(U^T V).$$

例2：在次数小于等于n的多项式空间，对于两个多项式

$$\mathbf{p} = a_0 + a_1 x + \dots a_n x^n, \quad \mathbf{q} = b_0 + b_1 x + \dots b_n x^n,$$

我们定义标准内积

$$\langle \mathbf{p}, \mathbf{q}\rangle = a_0 b_0 + a_1 b_1 + \dots a_n b_n.$$

例3：在次数小于等于n的多项式空间，对于两个多项式

$$\mathbf{p} = p(x) = a_0 + a_1 x + \dots a_n x^n, \quad \mathbf{q} = q(x) = b_0 + b_1 x + \dots b_n x^n,$$

以及不同的n个实数$x_0, x_1, \dots, x_n$, 于是定值内积(evaluation inner product)定义为

$$\langle \mathbf{p}, \mathbf{q}\rangle = p(x_0)q(x_0) + p(x_1)q(x_1) + \dots p(x_n)q(x_n).$$

例4：对于区间$[a, b]$上的两个连续函数$\mathbf{f} = f(x)$和$\mathbf{g} = g(x)$, 我们定义内积

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_a^b f(x)g(x)dx.$$

6.2 内积空间中的角度和正交关系

定理 6.2.1: (Cauchy-Schwarz 不等式) 设$\mathbf{u}$和$\mathbf{v}$是内积空间V中的两个向量，则

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|.$$

证明：对于任意的实数t, 我们知道

$$\langle t\mathbf{u} + \mathbf{v}, t\mathbf{u} + \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle t^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle t + \langle \mathbf{v}, \mathbf{v}\rangle \ge 0$$

将此方程看成一个二次方程，则行列式恒小于等于0, 即

$$4\langle \mathbf{u}, \mathbf{v}\rangle^2 - 4\langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle \le 0.$$

从而，我们不难知道

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

□

由Cauchy-Schwarz 不等式，有

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$$

从而存在唯一的$\theta \in [0, \pi]$，使得

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|},$$

这样的θ称为向量$\mathbf{u}$和$\mathbf{v}$ 的夹角。

下面的定理的证明和**定理** 3.2.5的证明是类似的。

定理 6.2.2: 设$\mathbf{u}, \mathbf{v}$ 和$\mathbf{w}$ 是实内积空间V中的向量，则有下列三角不等式成立

(a) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

(b) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

内积空间V中的两个向量$\mathbf{u}$和$\mathbf{v}$称为是正交的，如果$\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

定理 6.2.3: 设$\mathbf{u}$和$\mathbf{v}$是内积空间中的两个正交向量，则

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

证明：向量$\mathbf{u}$和$\mathbf{v}$正交意味着$\langle \mathbf{u}, \mathbf{v} \rangle = 0$,于是

$$\|\mathbf{u} + \mathbf{v}\|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

□

设W是内积空间V的子空间，那么V中与W中每个向量都垂直的向量的集合称为W的正交补，记成$W^\perp$.

定理 6.2.4: 设W是内积空间V的子空间，则

(a) $W^\perp$是V的子空间。

(b) $W \cap W^\perp = \{\mathbf{0}\}$.

证明: (a) 我们只要说明$W^\perp$对于向量加法和纯量乘法封闭即可。对于$\mathbf{u}, \mathbf{v} \in W^\perp$ 以及纯量λ, 首先对于任意的向量$\mathbf{w}$, 我们有

$$\langle \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle = 0.$$

从而

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = 0, \quad \langle \lambda\mathbf{u}, \mathbf{w} \rangle = \lambda\langle \mathbf{u}, \mathbf{w} \rangle = 0.$$

我们就有$\mathbf{u} + \mathbf{v} \in W^\perp, \lambda\mathbf{u} \in W^\perp$.

(b) 设$\mathbf{v} \in W \cap W^\perp$, 则有$\langle \mathbf{v}, \mathbf{v} \rangle = 0$, 于是$\mathbf{v} = 0$, 这就说明$W \cap W^\perp = \{\mathbf{0}\}$. □

定理 6.2.5: 设W是有限维内积空间V的子空间，则$(W^\perp)^\perp = W$.

证明: 首先很容易注意到$W \subseteq (W^{\perp})^{\perp}$, 这是因为对于$\mathbf{w} \in W$以及$\mathbf{w}' \in W^{\perp}$, 我们有$\langle \mathbf{w}, \mathbf{w}' \rangle = 0$. 在这里，我们超前使用一下后面的定理，**定理** 6.3.3告诉我们，其实

$$V = W \oplus W^{\perp},$$

于是$\dim W^{\perp} = \dim V - \dim W$. 于是

$$\dim (W^{\perp})^{\perp} = \dim V - \dim W^{\perp} = \dim V - (\dim V - \dim W) = \dim W.$$

结合$W \subseteq (W^{\perp})^{\perp}$, 我们知道了$(W^{\perp})^{\perp} = W$. □

6.3 Gram-Schmidt正交化，QR分解

内积空间中的一个集合被称为是正交的，如果它中的向量彼此是正交的。一个正交的集合称为是正规正交的，如果其中的每个向量都是单位向量。

定理 6.3.1: 设$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是内积空间中不含零向量的正交集合，则$S$是线性无关的。

证明：假设S中的向量满足

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}.$$

对任意的i, 两边考虑关于向量$\mathbf{v}_i$的内积，则有

$$\sum_{k=1}^{n} \lambda_k \langle \mathbf{v}_k, \mathbf{v}_i \rangle = 0.$$

根据集合S的正交性，我们知道对任意的i, 有$\lambda_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0$. 注意到$\mathbf{v}_i \neq \mathbf{0}$, 于是$\langle \mathbf{v}_i, \mathbf{v}_i \rangle > 0$, 从而$\lambda_i = 0$. 这就说明了$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是线性无关的。 □

我们知道给定向量空间V中的一组基$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$后，我们知道$V$中的任意向量$\mathbf{u}$可以写成这个基下的线性组合：

$$\mathbf{u} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n,$$

其中c_i是系数。当S是正交的或者说正规正交的集合的时候，我们有下面的定理。

定理 6.3.2:

(a) 设$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是内积空间$V$上的一组正交基，$\mathbf{u}$是$V$的向量，则

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2} \mathbf{v}_n.$$

(b) 设$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是内积空间$V$上的一组正规正交基，$\mathbf{u}$是$V$的向量，则

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n \rangle \mathbf{v}_n.$$

证明：(a) 因为$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是向量空间$V$上的一组基，则

$$\mathbf{u} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n,$$

等式两边用$\mathbf{v}_i$作内积，有

$$\langle \mathbf{u}, \mathbf{v}_i \rangle = c_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle.$$

注意到$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是正交的，因此

$$\langle \mathbf{u}, \mathbf{v}_i \rangle = c_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = c_i\|\mathbf{v}_i\|^2,$$

于是，我们得到了$c_i = \frac{\langle \mathbf{u}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2}$, 命题得证。

(b) 因为S是正规正交的，所以对于任意的i, 有$\|\mathbf{v}_i\|^2 = 1$, 于是由(a)可得(b). □

在许多的问题中，正交基或者正规正交基总是给我们很多的帮助，但往往一开始给予的是一个很一般的基底，我们要想办法把这样的基底转换成正交基或者正规正交基。作为准备，首先我们介绍投射定理，这是**定理** 3.3.2的推广。

定理 6.3.3: （投射定理）设W是内积空间V中的有限维空间，则V中的任意向量$\mathbf{u}$可以唯一写成

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$$

其中$\mathbf{w}_1 \in W$, $\mathbf{w}_2 \in W^{\perp}$.

证明：因为W是V中的有限维空间，我们可以假设W有一个基底为$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$. 我们断言存在唯一的$\lambda_1, \lambda_2, \ldots, \lambda_m$ 使得

$$\langle \mathbf{u} - (\sum_{i=1}^{m} \lambda_i \mathbf{v}_i), \mathbf{v}_k \rangle = 0$$

对于任意的k $(1 \leq k \leq m)$成立。如果证明了这个断言，则令

$$\mathbf{w}_1 = \sum_{i=1}^{m} \lambda_i \mathbf{v}_i \text{ 以及 } \mathbf{w}_2 = \mathbf{u} - (\sum_{i=1}^{m} \lambda_i \mathbf{v}_i)$$

即可。下面我们证明断言成立，这等价于方程组

$$\begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_1, \mathbf{v}_m \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_2, \mathbf{v}_m \rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle \mathbf{v}_m, \mathbf{v}_1 \rangle & \langle \mathbf{v}_m, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_m, \mathbf{v}_m \rangle \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}, \mathbf{v}_1 \rangle \\ \langle \mathbf{u}, \mathbf{v}_2 \rangle \\ \cdots \\ \langle \mathbf{u}, \mathbf{v}_m \rangle \end{pmatrix}$$

有解。接着我们只要证明矩阵

$$A = \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_1, \mathbf{v}_m \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_2, \mathbf{v}_m \rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle \mathbf{v}_m, \mathbf{v}_1 \rangle & \langle \mathbf{v}_m, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_m, \mathbf{v}_m \rangle \end{pmatrix}$$

是可逆的即可，这就等价于证明其行向量是线性无关的。现在设

$$\sum_{k=1}^{m} c_k(\langle \mathbf{v}_k, \mathbf{v}_1\rangle, \langle \mathbf{v}_k, \mathbf{v}_2\rangle, \ldots, \langle \mathbf{v}_k, \mathbf{v}_m\rangle) = \mathbf{0},$$

根据内积的性质，这等价于

$$(\langle \sum_{k=1}^{m} c_k \mathbf{v}_k, \mathbf{v}_1\rangle, \langle \sum_{k=1}^{m} c_k \mathbf{v}_k, \mathbf{v}_2\rangle, \ldots, \langle \sum_{k=1}^{m} c_k \mathbf{v}_k, \mathbf{v}_m\rangle) = \mathbf{0},$$

于是我们知道$\sum_{k=1}^{m} c_k \mathbf{v}_k \in W \cap W^\perp = \{\mathbf{0}\}$. 注意到$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$是$W$的一个基地，这就要求每个$c_i = 0$, 从而矩阵$A$的行向量是线性无关的，矩阵$A$可逆。

□

此时，我们记$\mathbf{w}_1 = \mathrm{proj}_W \mathbf{u}$, 称为$\mathbf{u}$关于$W$的正交投射以及$\mathbf{w}_2 = \mathrm{proj}_{W^\perp} \mathbf{u}$, 称为$\mathbf{u}$关于$W^\perp$的正交投射。

定理 6.3.4: 设W是内积空间V的有限维空间，

(a) 如果$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是$W$的正交基，$\mathbf{u}$是$V$中的向量，于是

$$\mathrm{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

(b) 如果$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$是$W$的正规正交基，$\mathbf{u}$是$V$ 中的向量，于是

$$\mathrm{proj}_W \mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_r\rangle \mathbf{v}_r.$$

证明：(a) 有定理6.3.3知，V中的任意向量$\mathbf{u}$可以写成

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2,$$

其中$\mathbf{w}_1 \in W$, $\mathbf{w}_2 \in W^\perp$. 而根据定理6.3.2,

$$\mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

而对于任意的$k(1 \le k \le r)$, 有$\langle \mathbf{w}_2, \mathbf{v}_k\rangle = 0$, 从而最终我们得到了

$$\mathrm{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

(b) 因为对于任意的i, 有$\|\mathbf{v}_i\|^2 = 1$, 于是由(a)可得(b). □

接下来，我们介绍Gram-Schmidt正交化，下面的定理不仅告诉了我们正交基的存在性，同时也给出了算法。

定理 6.3.5: 非零的有限维内积空间都有正规正交基。

证明：设$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$是$W$的一个基，我们希望以此构造出一个正交基$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, 再做单位化就可以得到正规正交的基底。

步骤一：令$\mathbf{v}_1 = \mathbf{u}_1$.

步骤二：记W_1是由$\mathbf{v}_1$生成的空间，考虑

$$\mathbf{v}_2 = \mathbf{u}_2 - \mathrm{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1,$$

由$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$是$W$的一个基知，$\mathbf{v}_2 \neq 0$.

步骤三：记W_2是由$\mathbf{v}_1, \mathbf{v}_2$生成的空间，考虑

$$\mathbf{v}_3 = \mathbf{u}_3 - \mathrm{proj}_{W_2}\mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2.$$

按照这样的步骤一步步地操作下去，经过r步后，我们得到了向量组$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$, 因为正交向量是线性无关的，因此我们就得到了一个正交基。对这些基作单位化就得到了正规正交基。上述这些操作也就是Gram-Schmidt正交化。 □

定理 6.3.6: 设W是一个有限维的内积空间，则

(a) W中的正交向量组总可以扩展成W的正交基。

(b) W中的正规正交向量组总可以扩展成W的正规正交基。

证明：我们证明(a), 而(b)留作练习。对于W中的正交向量组$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$, 其都可以扩展成$W$的基底$S' = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s, \mathbf{v}_{s+1}, \dots, \mathbf{v}_m\}$, 在对$S'$作Gram-Schmidt正交化的时候，因为前$S$个向量已经正交了，所以保持不变，我们对后面的向量做Gram-Schmidt正交化，得到正交向量组$S'' = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s, \mathbf{v}'_{s+1}, \dots, \mathbf{v}'_m\}$. □

在本节的最后，我们来介绍QR分解。问题是这样的，设A是一个$m \times n$的矩阵，其列向量是线性无关的，矩阵Q 是对A的列向量作Gram-Schmidt正交化得到的矩阵，那么A和Q有什么关系呢？

设A的列向量是$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$, 矩阵$Q$的列向量是$\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. 于是对于任意的$k(1 \leq k \leq n)$, 我们知道了

$$\mathbf{u}_k = \langle \mathbf{u}_k, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_k, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_k, \mathbf{q}_n \rangle \mathbf{q}_n.$$

因此可以写成矩阵的形式:

$$[\mathbf{u}_1 \mid \mathbf{u}_2 \mid \dots \mid \mathbf{u}_n] = [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \dots \mid \mathbf{q}_n] \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \dots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \dots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \dots & \dots & \dots & \dots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \dots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix},$$

或者简记为$A = QR$. 但是根据Gram-Schmidt正交化，我们知道$\mathbf{q}_k$ 是和$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}$正交的，于是

$$R = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix}.$$

这样，我们就得到了下面的定理。

定理 6.3.7: （QR分解）设$m \times n$的矩阵A的列线性无关，则A可以分解成$A = QR$, 其中$m \times n$的矩阵Q是列向量相互正交的矩阵，R是n阶可逆的上三角矩阵。

6.4 最小二乘法

假设关于n个未知元的m个方程形成的线性方程组$A\mathbf{x} = \mathbf{b}$是不相容的。我们寻找一个$\mathbf{x}$尽量逼近方程的“解”，即要求$\|\mathbf{b} - A\mathbf{x}\|$在$\mathbb{R}^m$中的欧氏范数下最小，此时我们称$\mathbf{x}$是线性方程组$A\mathbf{x} = \mathbf{b}$的最小二乘解，$\mathbf{b} - A\mathbf{x}$是最小二乘误差向量，$\|\mathbf{b} - A\mathbf{x}\|$ 是最小二乘误差。

首先，我们介绍最优逼近定理。

定理 6.4.1: （最优逼近）设W是内积空间V中的有限维子空间，$\mathbf{b}$ 是V中的一个向量，则$\mathrm{proj}_W \mathbf{b}$满足对于任意的$\mathbf{w} \in W$ 且$\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$, 有

$$\|\mathbf{b} - \mathrm{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

于是$\mathrm{proj}_W \mathbf{b}$是$\mathbf{b}$ 相对于空间W在上述意义下的最优逼近。

证明：对于任意的$\mathbf{w} \in W$, 有

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \mathrm{proj}_W \mathbf{b}) + (\mathrm{proj}_W \mathbf{b} - \mathbf{w}),$$

注意到$\mathbf{b} - \mathrm{proj}_W \mathbf{b}$ 与W中的任意向量正交，自然和$\mathrm{proj}_W \mathbf{b} - \mathbf{w}$正交。从而，我们由毕达哥拉斯定理，知

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \mathrm{proj}_W \mathbf{b}\|^2 + \|\mathrm{proj}_W \mathbf{b} - \mathbf{w}\|^2.$$

从而，当$\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$ 时，有

$$\|\mathbf{b} - \mathbf{w}\|^2 > \|\mathbf{b} - \mathrm{proj}_W \mathbf{b}\|^2,$$

于是有

$$\|\mathbf{b} - \mathrm{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|.$$

□

根据上面的最优逼近定理，对于线性方程组$A\mathbf{x} = \mathbf{b}$, 记W是矩阵A的列空间，我们考虑$\mathbf{b}$ 相对于W的投射$\mathrm{proj}_W \mathbf{b}$. 同时考虑线性方程组$A\mathbf{x} = \mathrm{proj}_W \mathbf{b}$. 将上述方程组改写成

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \mathrm{proj}_W \mathbf{b}.$$

两边同时左乘矩阵A^T, 有

$$A^T(\mathbf{b}-A\mathbf{x})=A^T(\mathbf{b}-\mathrm{proj}_W\mathbf{b}).$$

注意到$\mathbf{b}-\mathrm{proj}_W\mathbf{b}$ 与A的列空间中的向量都是正交的，从而

$$A^T(\mathbf{b}-\mathrm{proj}_W\mathbf{b})=0.$$

于是我们有$A^T(\mathbf{b}-A\mathbf{x})=0$, 再此改写成

$$A^TA\mathbf{x}=A^T\mathbf{b}.$$

此方程被称为是关于线性方程组$A\mathbf{x}=\mathbf{b}$的正规方程。反过来，给了正规方程$A\mathbf{x}=\mathbf{b}$, 我们会有$A^T(\mathbf{b}-A\mathbf{x})=0$, 类似于上面的说明，我们知道$\mathbf{b}-A\mathbf{x}\in W^\perp$. 由$\mathbf{b}$关于$W$和$W^\perp$正交投射分解的唯一性知道，必然有$A\mathbf{x}=\mathrm{proj}_W\mathbf{b}$. 于是，我们也证明了下述定理。

定理 6.4.2: 对于任意的线性方程组$A\mathbf{x}=\mathbf{b}$, 相应的正规方程

$$A^TA\mathbf{x}=A^T\mathbf{b}$$

总是相容的，且它的解都是$A\mathbf{x}=\mathbf{b}$ 的最小二乘解。更多地，设W是矩阵A的列向量空间，$\mathbf{x}$是$A\mathbf{x}=\mathbf{b}$的最小二乘解，则$\mathbf{b}$在W上的正规投射有

$$\mathrm{proj}_W\mathbf{b}=A\mathbf{x}.$$

定理 6.4.3: 设A是一个$m\times n$的矩阵，则下列条件等价：

(a) A的列向量线性无关。

(b) A^TA是可逆矩阵。

证明：(a)$\Longrightarrow$(b) 设A的列向量线性无关，要证明A^TA是可逆矩阵，即证明方程组$A^TA\mathbf{x}=\mathbf{0}$只有零解。但是当$\mathbf{x}$是此方程的解时，$A\mathbf{x}$在$A^T$的零化空间中，同时也在$A$的列向量空间中。但是由**定理** 4.8.9 (b)知道，这两个空间是彼此正交的，于是我们知道$A\mathbf{x}=\mathbf{0}$, 注意到A的列向量线性无关，则$\mathbf{x}=\mathbf{0}$, 命题得证。

(b)$\Longrightarrow$(a) 假设A的列向量线性相关，于是存在非零的列向量$\mathbf{x}_0$, 使得$A\mathbf{x}_0=\mathbf{0}$, 那么$A^TA\mathbf{x}_0=\mathbf{0}$. 这说明方程$A^TA\mathbf{x}=\mathbf{0}$ 有非零解，这与A^TA是可逆矩阵矛盾。从而推出矛盾，因此A的列向量线性无关。 $\square$

结合**定理** 6.4.2 和**定理** 6.4.3, 我们得到下面的定理：

定理 6.4.4: 设A是一个$m\times n$的矩阵，且其列向量都是线性无关的，则对于$m\times 1$的矩阵$\mathbf{b}$, 则线性方程组$A\mathbf{x}=\mathbf{b}$有唯一的最小二乘解。其解为

$$\mathbf{x}=(A^TA)^{-1}A^T\mathbf{b}.$$

更多地，设W是A的列向量空间，则$\mathbf{b}$在W上的正规投射有

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{b}.$$

设W是$\mathbb{R}^m$中的子空间，则线性映射$P: \mathbb{R}^m \to W$将$\mathbf{x} \in \mathbb{R}^m$ 映到$\mathrm{proj}_W\mathbf{x} \in W$ 称为是从$\mathbb{R}^m$到W上的正交投射。该线性映射对应的矩阵是$[P] = A(A^TA)^{-1}A^T$, 这里的A是由W的一个基作为列向量构成的矩阵。

现在我们从另外一种观点来看最小二乘法。对于一个$m \times n$的矩阵A, 它的行向量空间和其零化空间是正交的，它的列向量空间和矩阵A^T的零化空间是正交的。于是对于线性方程组$A\mathbf{x} = \mathbf{b}$, 我们对向量$\mathbf{x} \in \mathbb{R}^m$以及向量$\mathbf{b} \in \mathbb{R}^n$作投射分解：

$$\mathbf{x} = \mathbf{x}_{\mathrm{row}(A)} + \mathbf{x}_{\mathrm{null}(A)}, \quad \mathbf{b} = \mathbf{b}_{\mathrm{null}(A^T)} + \mathbf{b}_{\mathrm{col}(A)},$$

这里的$\mathbf{x}_{\mathrm{row}(A)}$是$\mathbf{x}$ 相对于A的行向量空间的投射，$\mathbf{x}_{\mathrm{null}(A)}$ 是$\mathbf{x}$相对于A的零化空间的投射, $\mathbf{b}_{\mathrm{null}(A^T)}$是$\mathbf{b}$相对于$A^T$的零化空间的投射，$\mathbf{b}_{\mathrm{col}(A)}$是$\mathbf{b}$相对于$A$的列向量空间的投射。参看教材中Figure 6.4.3, 注意到$A\mathbf{x} \in \mathrm{col}(A)$, 于是$A\mathbf{x} = \mathbf{b}$的最小二乘解恰好是方程组$A\mathbf{x} = \mathbf{b}_{\mathrm{col}(A)}$的解。

7. 对角化(diagonalization)和二次型(quadratic form)

7.1 正交矩阵

一个方阵A称为正交的，如果它的逆和转置是相等的，即$A^{-1} = A^T$, 或者等价地有

$$AA^T = A^TA = I.$$

定理 7.1.1: 对于n阶方阵A, 下列条件是等价的：

(a) A是正交的。

(b) A的行向量形成$\mathbb{R}^n$关于欧氏内积的正规正交基。

(c) A的列向量形成$\mathbb{R}^n$关于欧氏内积的正规正交基。

证明：我们证明(a)和(b)是等价的，(a)和(c)的等价性是类似的。

设A的行向量是$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$. 从而不难发现

$$AA^T = \begin{pmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_1 \cdot \mathbf{r}_n \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_2 \cdot \mathbf{r}_n \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{r}_n \cdot \mathbf{r}_1 & \mathbf{r}_n \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_n \cdot \mathbf{r}_n \end{pmatrix},$$

从而A是正交的，等价于$AA^T = I$. 而这意味着$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ 是$\mathbb{R}^n$中关于欧氏内积的正规正交基。 □

定理 7.1.2: (留作练习)

(a) 正交矩阵的逆仍然是正交矩阵。

(b) 正交矩阵的乘积是正交矩阵。

(c) 对于正交矩阵，则$\det A = 1$ 或者$\det A = -1$.

定理 7.1.3: 设A是一个n阶方阵，则下面的条件等价：

(a) A是正交矩阵。

(b) 对于任意的$\mathbf{x} \in \mathbb{R}^n$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$ 成立。

(c) 对于任意的$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

证明：我们证明$(a) \Longrightarrow (b) \Longrightarrow (c) \Longrightarrow (a)$.

$(a) \Longrightarrow (b)$ 设A是正交矩阵，则有$AA^T = I$. 于是

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = \mathbf{x} \cdot A^T A\mathbf{x} = \mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2,$$

从而$\|A\mathbf{x}\| = \|\mathbf{x}\|$, 得到了(b).

$(b) \Longrightarrow (c)$ 由**定理** 3.2.7, 我们知道

$$A\mathbf{x} \cdot A\mathbf{y} = \frac{1}{4}\|A\mathbf{x} + A\mathbf{y}\|^2 - \frac{1}{4}\|A\mathbf{x} - A\mathbf{y}\|^2 = \frac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \frac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2 = \mathbf{x} \cdot \mathbf{y}.$$

$(c) \Longrightarrow (a)$ 由$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot A^T A\mathbf{y}$ 知条件(c)等价于，对于任意的$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x} \cdot (A^T A - I)\mathbf{y} = 0.$$

特别地，我们取$\mathbf{x} = (A^T A - I)\mathbf{y}$, 从而

$$(A^T A - I)\mathbf{y} \cdot (A^T A - I)\mathbf{y} = 0.$$

于是$(A^T A - I)\mathbf{y} = 0$对于任意的$\mathbf{y}$都成立，这必然要求$A^T A = I$, 于是$A$ 是正交矩阵。 □

正交基对我们而言使用起来是非常方便的。

定理 7.1.4: （留作练习）设S是n维内积空间V中的正规正交基，设

$$(\mathbf{u})_S = (u_1, u_2, \ldots, u_n), \quad (\mathbf{v})_S = (v_1, v_2, \ldots, v_n).$$

则我们有

(a) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \ldots u_n^2}$

(b) $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$.

(c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.

定理 7.1.5: 设V是有限维的内积空间。如果P是从V的一组正规正交基到另一组正规正交基的转换矩阵，则P是正交矩阵。

证明 设V的两组正规正交基分别是$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ 和$S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, 设矩阵$P = (p_{ij})_{n \times n}$是转换矩阵满足

$$\mathbf{v}_k = \sum_{i=1}^{n} p_{ik} \mathbf{u}_i, \quad \text{其中} k = 1, 2, \ldots, n$$

从而

$$\langle \mathbf{v}_k, \mathbf{v}_l \rangle = \langle \sum_{i=1}^{n} p_{ik}\mathbf{u}_i, \sum_{j=1}^{n} p_{jl}\mathbf{u}_j \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ik}p_{jl}\langle \mathbf{u}_i, \mathbf{u}_j \rangle.$$

由S和S'分别是正规正交基，于是有

$$\langle \mathbf{v}_k, \mathbf{v}_l \rangle = \delta_{kl}, \quad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij},$$

则我们可得$\sum_{i=1}^{n} p_{ik}p_{il} = \delta_{kl}$, 这就说明矩阵$P$的列向量是$\mathbb{R}^n$的正规正交基，于是知道$P$是正交矩阵，命题得证。 □

7.2 正交对角化

之前，我们定义过两个矩阵的相似，即对于矩阵A和B, 如果存在一个可逆的方阵P, 使得$P^{-1}AP = B$, 则称A和B是相似的。这一节中，我们要找矩阵P为正交矩阵。

设A和B是两个方阵，我们称A和B是正交相似，如果存在正交矩阵P, 使得$P^TAP = B$. 特别地，取B为对角矩阵D的时候，$P^TAP = D$, 此时A正交相似到对角矩阵，我们称A 是可正交对角化的。

在这一节中，我们找出A可正交对角化的条件。首先注意到，如果$P^TAP = D$, 则有$A = PDP^T$. 于是

$$A^T = (PDP^T)^T = PD^TP^T = PDP^T = A.$$

这就说明A是对称的。

定理 7.2.1: 设A是n阶方阵，则下面的条件等价：

(a) A是可正交对角化的。

(b) A有n个正规正交的特征向量。

(c) A是对称的。

证明: $(a) \Longrightarrow (b)$ 设A是可正交对角化的，则存在正交矩阵P, 使得$P^TAP = D$, 这里D是对角矩阵，主对角线元素分别是$\lambda_1, \lambda_2, \ldots, \lambda_n$. 记矩阵$P$的列向量是$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, 根据$P$正交可知$P$的列向量是正规正交的。用时由$AP = PD$ 知，$A\mathbf{p}_k = \lambda_k\mathbf{p}_k$, 于是$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$形成$A$ 的n 个正规正交的特征向量。

$(b) \Longrightarrow (a)$ 将正规正交的列向量$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$放在一起，形成一个正交矩阵$P$, 则$P^TAP$就是一个对角阵，即$A$可正交对角化。

$(a) \Longrightarrow (c)$ 设A是可正交对角化的，即有$P^TAP = D$, 于是我们知道$A = PDP^T$. 从而

$$A^T = (PDP^T)^T = PD^TP^T = PDP^T = A.$$

这就说明A是对称的。

$(c) \Longrightarrow (a)$ 略去。

下面我们研究对称矩阵的性质。

定理 7.2.2：设A是实对称矩阵，则

(a) A的特征值都是实数。

(b) 不同特征值的特征空间彼此是正交的。

证明: (a) 设λ是A的特征值，向量$\mathbf{v}=(v_1,v_2,\dots,v_n)^T$是对应的特征列向量，即有$A\mathbf{v}=\lambda\mathbf{v}$. 两边同时左乘上$\overline{\mathbf{v}^T}$. 我们有

$$\overline{\mathbf{v}^T}A\mathbf{v}=\lambda\overline{\mathbf{v}^T}\mathbf{v}.$$

注意到

$$\overline{\mathbf{v}^T}\mathbf{v}=|v_1|^2+|v_2|^2+\cdots+|v_n|^2\in\mathbb{R}$$

是大于0的实数，我们只要说明$\overline{\mathbf{v}^T}A\mathbf{v}$是实数即可。首先$\overline{\mathbf{v}^T}A\mathbf{v}$是一个复数，再考虑它的转置共轭，我们有

$$\overline{\left(\overline{\mathbf{v}^T}A\mathbf{v}\right)^T}=\overline{\mathbf{v}^T}\ \overline{A^T}\mathbf{v}=\overline{\mathbf{v}^T}A\mathbf{v}.$$

这就说明$\overline{\mathbf{v}^T}A\mathbf{v}$是一个实数，从而最终知道$A$的特征值都是实数。

(b) 设$\mathbf{v}_1$ 和$\mathbf{v}_2$分别是A对应到λ_1和λ_2的特征向量，我们证明$\mathbf{v}_1\cdot\mathbf{v}_2=0$. 考虑$A\mathbf{v}_1\cdot\mathbf{v}_2$有

$$A\mathbf{v}_1\cdot\mathbf{v}_2=\mathbf{v}_1\cdot A^T\mathbf{v}_2=\mathbf{v}_1\cdot A\mathbf{v}_2$$

同时，注意到$A\mathbf{v}_1=\lambda_1\mathbf{v}_1, A\mathbf{v}_2=\lambda_2\mathbf{v}_2$, 我们就知道

$$\lambda_1\ \mathbf{v}_1\cdot\mathbf{v}_2=\lambda_2\ \mathbf{v}_1\cdot\mathbf{v}_2.$$

因为$\lambda_1\neq\lambda_2$, 所以$\mathbf{v}_1\cdot\mathbf{v}_2=0$. □

有了这个定理，我们给出正交对角化一个对称矩阵的算法。

步骤一：找出A的每一个特征空间的基底。

步骤二：对每一个特征空间的基底用Gram-Schmidt正交化，得到每个特征空间的正规正交基底。

步骤三：由步骤二得到的正规正交基底作为列向量构成矩阵P, 于是此矩阵是正交矩阵，且$P^TAP=D$.

下面我们介绍谱分解。设A是一个对称矩阵，被方阵

$$P=[\mathbf{u}_1,\mathbf{u}_2,\dots,\mathbf{u}_n]$$

所正交对角化，且设$\lambda_1,\lambda_2,\dots,\lambda_n$是$A$的特征值，相对应的单位特征向量是$\mathbf{u}_1,\mathbf{u}_2,\dots,\mathbf{u}_n$, 于是我们知道$D=P^TAP$. 从而

$$A=PDP^T=\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n\end{pmatrix}\begin{pmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & \lambda_n\end{pmatrix}\begin{pmatrix}\mathbf{u}_1^T\\ \mathbf{u}_2^T\\ \dots\\ \mathbf{u}_n^T\end{pmatrix}$$

经过简单的计算，我们有

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T,$$

称为A的谱分解。

注意到A的谱分解中，每一个分量都具有$\lambda \mathbf{u}\mathbf{u}^T$的形式，其中$\mathbf{u}$ 是A的单位特征向量，λ是对应的特征值。因为$\mathbf{u}$是$n \times 1$的列向量，于是$\mathbf{u}\mathbf{u}^T$是n阶方阵。很容易证明矩阵$\mathbf{u}\mathbf{u}^T$是从$\mathbb{R}^n$到向量$\mathbf{u}$张成空间U的投射，这是因为对于向量$\mathbf{x}$, 可以写成$\mathbf{x} = \lambda \mathbf{u} + \mathbf{v}$, 这里$\mathbf{v} \in U^{\perp}$. 从而

$$(\mathbf{u}\mathbf{u}^T)\mathbf{x} = \lambda(\mathbf{u}\mathbf{u}^T)\mathbf{u} + (\mathbf{u}\mathbf{u}^T)\mathbf{v} = \lambda \mathbf{u}(\mathbf{u} \cdot \mathbf{u}) + \mathbf{u}(\mathbf{u} \cdot \mathbf{v}) = \lambda \mathbf{u}.$$

于是一个对称矩阵A乘上向量$\mathbf{x}$, 等价于对$\mathbf{x}$在每一个A的每一个特征向量$\mathbf{u}$上做投射，然后伸缩λ倍（λ是$\mathbf{u}$的特征值），最后再将这些分量相加。

对于一个无法被正交对角化的矩阵A, 我们有时还是能找到正交矩阵P, 使得$P^T AP$的形式相对简单。我们列举出来，但是不给证明。

定理 7.2.3：设A是一个实方阵且它的特征值都是实数，于是存在正交矩阵P使得$P^T AP$是上三角矩阵，且形式为

$$P^T AP = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \lambda_2 & \dots & * \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

这里的$\lambda_1, \lambda_2, \dots, \lambda_n$是$A$的特征值（根据重数可以重复）。

定理 7.2.4：设A是一个实方阵，存在正交矩阵P使得$P^T AP$有如下的形式：

$$P^T AP = \begin{pmatrix} * & * & * & \dots & * & * & * & * \\ * & * & * & \dots & * & * & * & * \\ 0 & * & * & \dots & * & * & * & * \\ 0 & 0 & * & \dots & * & * & * & * \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \cdots & 0 & * & * & * & * \\ 0 & 0 & 0 & \cdots & 0 & * & * & * \\ 0 & 0 & 0 & 0 & \cdots & 0 & * & * \end{pmatrix},$$

这样的矩阵是具有上Hessenberg形式的矩阵。

7.3 二次型

二次型在数论，几何上常常出现，一个$\mathbb{R}^n$ 上的二次型具有如下的形式

$$a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2 + \text{交错项}$$

其实，我们可以将二次型通过矩阵乘法的形式表示出来。设$A = (a_{ij})_{n\times n}$是一个对称的n阶方阵，$\mathbf{x}$ 是一个含有n 个变元的列向量，于是

$$Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + \sum_{i<j} 2a_{ij}x_ix_j.$$

称为是与A伴连的二次型。于是，我们可以知道二次型和对称方阵存在一个一一对应的关系。特别地，当A是对角矩阵，主对角线为$\lambda_1, \lambda_2, \ldots, \lambda_n$, 则

$$\mathbf{x}^T A\mathbf{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2.$$

关于二次型，有如下一些重要的问题：

问题一：设$\mathbf{x}^T A\mathbf{x}$是$\mathbb{R}^2$上或者$\mathbb{R}^3$的二次型，则方程$\mathbf{x}^T A\mathbf{x} = k$代表什么样的曲线或者曲面。

问题二：设$\mathbf{x}^T A\mathbf{x}$是$\mathbb{R}^n$上的二次型，如果要求$\mathbf{x}^T A\mathbf{x}$ 对于$\mathbf{x} \neq 0$都取正值，则A必须满足什么条件？

问题三：设$\mathbf{x}^T A\mathbf{x}$是$\mathbb{R}^n$上的二次型，则当$\|\mathbf{x}\| = 1$时，$\mathbf{x}^T A\mathbf{x}$ 的最大值和最小值是怎样的？

在这一节中，我们考虑问题一和问题二，下一节我们将会考虑问题三。对于二次型$\mathbf{x}^T A\mathbf{x}$, 通过变量替换，我们有时会将二次型的形式变得简单。设$\mathbf{x} = P\mathbf{y}$, 如果P是可逆的，则称为变量替换，如果P是正交的，则称为正规变量替换。对于变量替换$\mathbf{x} = P\mathbf{y}$, 我们得到

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T(P^T AP)\mathbf{y}.$$

容易验证知，$B = P^T AP$是一个对称方阵。从而$\mathbf{y}^T B\mathbf{y}$是一个以$y_1, y_2, \ldots, y_n$变量的二次型。特别地，取P是正交矩阵，且使A正交对角化，从而新的二次型是$\mathbf{y}^T D\mathbf{y}$, 其中D是以A的特征值为主对角线的方阵，这就是说

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1\mathbf{y}_1^2 + \lambda_2\mathbf{y}_2^2 + \cdots + \lambda_n\mathbf{y}_n^2.$$

于是，我们就得到了下面的定理。

定理 7.3.1：（主轴定理）设A是一个n阶对称矩阵，对于二次型$\mathbf{x}^T A\mathbf{x}$, 存在正规的变量替换，使得新的二次型$\mathbf{y}^T D\mathbf{y}$没有交错项。具体地说，设正交矩阵P使A正交对角化，作变量替换$\mathbf{x} = P\mathbf{y}$, 则二次型

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1\mathbf{y}_1^2 + \lambda_2\mathbf{y}_2^2 + \cdots + \lambda_n\mathbf{y}_n^2,$$

这里的$\lambda_1, \lambda_2, \ldots, \lambda_n$是$A$相对于$P$的列向量的特征值。

圆锥曲线是高中所接触的知识，用一个平面去截对顶的圆锥体，当平面处于不同的位置时，我们会得到椭圆曲线，双曲线和抛物线。一般地，如下的方程

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0,$$

其中a, b, c不全为 0, 表示了$\mathbb{R}^2$中的一个圆锥曲线。如果$d = e = 0$, 则没有线性项，于是方程成为

$$ax^2 + 2bxy + cy^2 + f = 0,$$

代表了中心圆锥曲线，这包含了圆，椭圆，以及双曲线，但是不包含抛物线。更多地，我们令$b = 0$, 则没有交叉项，从而方程变成

$$ax^2 + cy^2 + f = 0,$$

代表了处于标准位置的圆锥曲线，见教材中的Table 1. 如果我们将f放到右侧，令$k = -f$, 于是上面的两个方程可以写成

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = k, \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = k.$$

现在，我们考虑两个变量或者三个变量的二次型$\mathbf{x}^T A\mathbf{x}$, 这代表了一条曲线或者一个曲面。下面我们主要考虑曲线的情形，即方程

$$ax^2 + 2bxy + cy^2 + f = 0,$$

它代表了一条中心圆锥曲线。当$b = 0$时，此曲线处于标准位置，当$b \neq 0$时，它会由标准曲线发生一个旋转。为方便起见，我们将曲线

$$ax^2 + 2bxy + cy^2 = k$$

写成

$$\mathbf{x}^T A\mathbf{x} = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = k.$$

考虑一个正交变量替换$\mathbf{x} = P\mathbf{x}'$, 使得$P$能让$A$正交对角化。我们可假设

$$P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

则P对xy-轴作一个旋转变成$x'y'$, 下面我们的目的是寻找θ, 能使A正交对角化。新的方程是

$$\mathbf{x}'^T D\mathbf{x}' = \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = k,$$

这里λ_1, λ_2是A的特征值。

接下去，我们介绍正定二次型。一个二次型$\mathbf{x}^T A\mathbf{x}$称为是正定的(positive definite), 如果对于$\mathbf{x} \neq 0$, 都有$\mathbf{x}^T A\mathbf{x} > 0$; 称为是负定的(negative definite), 如果对于$\mathbf{x} \neq 0$, 都有$\mathbf{x}^T A\mathbf{x} < 0$; 称为是不定的(indefinite), 如果$\mathbf{x}^T A\mathbf{x}$既取到正值也取到负值。另外，我们还有定义：称$\mathbf{x}^T A\mathbf{x}$

半正定的(positive semidefinite), 如果对于$\mathbf{x} \neq 0$, 都有$\mathbf{x}^T A\mathbf{x} \geq 0$; 称$\mathbf{x}^T A\mathbf{x}$半负定的(negative semidefinite), 如果对于$\mathbf{x} \neq 0$, 都有$\mathbf{x}^T A\mathbf{x} \leq 0$.

定理 7.3.2：设A是一个对称矩阵，则

(a) 二次型$\mathbf{x}^T A\mathbf{x}$是正定的当且仅当A的所有特征值都是正的。

(b) 二次型$\mathbf{x}^T A\mathbf{x}$是负定的当且仅当A的所有特征值都是负的。

(c) 二次型$\mathbf{x}^T A\mathbf{x}$是不定的当且仅当A至少有一个正的特征值，至少有一个负的特征值。

证明：对于二次型$\mathbf{x}^T A\mathbf{x}$, 存在正交矩阵P使得$\mathbf{x} = P\mathbf{y}$并且满足

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

其中$\lambda_1, \lambda_2, \ldots, \lambda_n$是$A$的特征值。同时，注意到$\mathbf{x} \neq 0$当且仅当$\mathbf{y} \neq 0$, 于是二次型$\mathbf{x}^T A\mathbf{x}$在$\mathbf{x} \neq \mathbf{0}$和二次型$\mathbf{y}^T D\mathbf{y}$在$\mathbf{y} = P^T\mathbf{x} \neq 0$的有相同的值。因此不难看出二次型$\mathbf{x}^T A\mathbf{x}$是正定的当且仅当$A$的所有特征值都是正的。类似地，有二次型$\mathbf{x}^T A\mathbf{x}$是负定的当且仅当$A$的所有特征值都是负的。这就证明了(a)和(b).

对于(c), 假设A至少有一个正的特征值，至少有一个负的特征值，我们不妨设$\lambda_1 > 0$, $\lambda_2 < 0$. 于是取$y_1 = 1$, 其他的y_i 为0, 有

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} > 0.$$

同样地，取$y_2 = 1$, 其他的y_i 为0, 有

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} < 0.$$

这就证明了二次型$\mathbf{x}^T A\mathbf{x}$是不定的。反过来设二次型$\mathbf{x}^T A\mathbf{x}$是不定的，如果存在$\mathbf{x}$使得$\mathbf{x}^T A\mathbf{x} > 0$, 则存在$\mathbf{y}$使得$\mathbf{y}^T D\mathbf{y} > 0$, 于是至少有一个特征值是正的。类似地可知，也必然有一个特征值是负的。 □

现在设$\mathbf{x}^T B\mathbf{x} = k$是一个圆锥曲线的方程，且$k \neq 0$, 我们可以将两边同除$k$, 方程就变成

$$\mathbf{x}^T A\mathbf{x} = 1,$$

其中$A = \frac{1}{k}B$. 如果我们做适当的旋转，将交叉项消去掉，则方程变成

$$\lambda_1 x'^2 + \lambda_2 y'^2 = 1,$$

其中λ_1, λ_2是A的特征值。此时，圆锥曲线的种类根据λ_1和λ_2的符号有下面几种情况：

- 当$\lambda_1 > 0$且$\lambda_2 > 0$时，方程代表椭圆。
- 当$\lambda_1 < 0$且$\lambda_2 < 0$时, 方程没有图像。
- 当λ_1和λ_2符号相反时，方程代表双曲线。

即我们有

定理 7.3.3：设A是2阶对称方阵，则

(a) 如果A是正定的，则$\mathbf{x}^T A\mathbf{x} = 1$ 代表椭圆。

(b) 如果A是负定的，则$\mathbf{x}^TA\mathbf{x}=1$没有图像。

(c) 如果A是不定的，则$\mathbf{x}^TA\mathbf{x}=1$代表双曲线。

因为正定矩阵是如此地重要，我们给出下面一个判定对称矩阵是否正定的判别发则。为此，我们要引入一个概念，即主子矩阵(principal submatrix)。对于一个n 阶对称方阵A, 考虑它的前k行和前k列形成的子矩阵，称为A的第k个主子矩阵。

定理 7.3.4：一个对称矩阵A是正定的，如果它的每一个主子矩阵的行列式都是正的。

7.4 用二次型做优化

在这一节中，我们主要考虑这样的问题：当$\|\mathbf{x}\|=1$时，$\mathbf{x}^TA\mathbf{x}$ 的最大值和最小值是怎样的？

定理 7.4.1：设A是一个n阶对称方阵，其特征值排序成$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$. 于是

(a) 对于满足$\|\mathbf{x}\|=1$ 的向量，二次型$\mathbf{x}^TA\mathbf{x}$ 可以取得最大值和最小值。

(b) 在(a)中取得的最大值对应于特征值λ_1的单位特征向量。

(c) 在(a)中取得的最小值对应于特征值λ_n的单位特征向量。

证明：设A是对称方阵，则主轴定理告诉我们存在正交变量替换$\mathbf{x}=P\mathbf{y}$ 使得

$$\mathbf{x}^TA\mathbf{x}=\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2$$

其中$\lambda_1,\lambda_2,\ldots,\lambda_n$是$A$的特征值。现在我们假设$\|\mathbf{x}\|=1$且$P$ 中的列向量的排序对应到的特征值满足$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$. 因为$P$是正交矩阵，于是左乘$P$保持长度不变，于是$\|\mathbf{y}\|=\|\mathbf{x}\|=1$. 从而

$$y_1^2+y_2^2+\cdots+y_n^2=1.$$

因此，我们就可以知道

$$\lambda_n=\lambda_n(y_1^2+y_2^2+\cdots+y_n^2)\leq\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2\leq\lambda_1(y_1^2+y_2^2+\cdots+y_n^2)=\lambda_1.$$

即有$\lambda_n\leq\mathbf{x}^TA\mathbf{x}\leq\lambda_1$.

现在设$\mathbf{x}$是对应到λ_1的单位特征向量，于是

$$\mathbf{x}^TA\mathbf{x}=\mathbf{x}^T(\lambda_1\mathbf{x})=\lambda_1\mathbf{x}^T\mathbf{x}=\lambda_1\|\mathbf{x}\|^2=\lambda_1,$$

这就说明$\mathbf{x}^TA\mathbf{x}$在$\mathbf{x}$ 取对应到λ_1的单位特征向量时，可以取得最大值。类似地，我们可以知道$\mathbf{x}$取对应到λ_n的单位特征向量时，可以取得最小值。 $\square$

对于一个函数$f(x,y)$, 假设其一阶偏导数存在，于是它的极大点和极小点，如果存在，一定会存在于满足

$$f_x(x,y)=0,f_y(x,y)=0$$

的点。这些点称为f的临界点(critical point). 在临界点(x_0,y_0)处，函数f的情况可以由

$$D(x,y)=f(x,y)-f(x_0,y_0)$$

来刻画，其中(x,y)很靠近(x_0,y_0), 但是不相等。

• 如果对于任意靠近(x_0,y_0), 但是不等于(x_0,y_0)的点(x,y)，有$D(x,y)>0$, 则f在(x_0,y_0)处取相对极小值。

• 如果对于任意靠近(x_0,y_0), 但是不等于(x_0,y_0)的点(x,y)，有$D(x,y)<0$, 则f在(x_0,y_0)处取相对极大值。

• 如果对于靠近(x_0,y_0), 但是不等于(x_0,y_0)的点(x,y)，$D(x,y)>0$ 和$D(x,y)<0$的点同时存在，于是称(x_0,y_0)是f 的马鞍点。

定理 7.4.2：假设(x_0,y_0)是f的临界点，且f在(x_0,y_0) 的领域内二阶连续可导。则

(a) 如果

$$f_{xx}(x_0,y_0)f_{yy}(x_0,y_0)-f_{xy}^2(x_0,y_0)>0,\quad f_{xx}(x_0,y_0)>0,$$

则f在(x_0,y_0)处取相对极小值。

(b) 如果

$$f_{xx}(x_0,y_0)f_{yy}(x_0,y_0)-f_{xy}^2(x_0,y_0)>0,\quad f_{xx}(x_0,y_0)<0,$$

则f在(x_0,y_0)处取相对极大值。

(c) 如果

$$f_{xx}(x_0,y_0)f_{yy}(x_0,y_0)-f_{xy}^2(x_0,y_0)<0,$$

则(x_0,y_0)是f的马鞍点。

(d) 如果

$$f_{xx}(x_0,y_0)f_{yy}(x_0,y_0)-f_{xy}^2(x_0,y_0)=0,$$

则此判定法失效。

对于定理中的函数f, 我们引入Hessian矩阵

$$H(x,y)=\begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{pmatrix},$$

于是，**定理** 7.4.2 就可以改写成下面的定理。

定理 7.4.3：假设(x_0,y_0)是f的临界点，且f在(x_0,y_0)的邻域内二阶连续可导。设$H(x_0,y_0)$是f在(x_0,y_0)处的Hessian矩阵，则

(a) 如果$H(x_0,y_0)$是正定的，则f在(x_0,y_0)处取相对极小值。

(b) 如果$H(x_0,y_0)$是负定的，则f在(x_0,y_0)处取相对极大值。

(c) 如果$H(x_0,y_0)$是不定的，则(x_0,y_0)是f的马鞍点。

(d) 对于其他情况，此方法不可判定。

7.5 Hermitian矩阵，酉矩阵和正规矩阵

设A是一个复方阵，则A的共轭转置，记为A^*, 定义为$A^*=\overline{A}^T$. 注意到$(\overline{A})^T=\overline{(A^T)}$, 于是先做共轭再做转置，和先做转置再共轭不会有什么影响。

定理 7.5.1：(练习) 设λ是一个复数，A, B, C是复矩阵，则

(a) $(A^*)^* = A$.

(b) $(A+B)^* = A^* + B^*$.

(c) $(A-B)^* = A^* - B^*$.

(d) $(\lambda A)^* = \overline{\lambda} A^*$.

(e) $(AB)^* = B^* A^*$.

我们定义一个方阵A是酉矩阵，如果$A^{-1} = A^*$. 一个方阵A被称作是Hermitian的，如果$A^* = A$. 我们可以将酉矩阵看成是正交矩阵在复数域上的推广，将Hermitian矩阵看成是对称矩阵在复数域上的推广。

为了方便，我们回顾一下$\mathbb{C}^n$上的复欧氏内积，给定$\mathbf{u} = (u_1, u_2, \ldots, u_n)$和$\mathbf{v} = (v_1, v_2, \ldots, v_n)$, 复欧氏内积为

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n}$$

假定$\mathbf{u}$ 和$\mathbf{v}$是列向量，于是$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\overline{\mathbf{v}}$. 设$A$是一个复方阵，此时

$$(A\mathbf{u}) \cdot \mathbf{v} = (A\mathbf{u})^T\overline{\mathbf{v}} = \mathbf{u}^T A^T \overline{\mathbf{v}} = \mathbf{u}^T(\overline{A^*\mathbf{v}}) = \mathbf{u} \cdot (A^*\mathbf{v}).$$

从而，我们有等式$(A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (A^*\mathbf{v})$.

定理 7.5.2：Hermitian矩阵的特征值都是实数。

证明：设λ是A的特征值，向量$\mathbf{v}$是对应的特征列向量，即有$A\mathbf{v} = \lambda\mathbf{v}$. 两边同时左乘上$\mathbf{v}^*$. 我们有

$$\mathbf{v}^* A\mathbf{v} = \lambda\mathbf{v}^*\mathbf{v}.$$

注意到$\mathbf{v}^*\mathbf{v}$ 是大于0的实数，我们只要说明$\mathbf{v}^*A\mathbf{v}$是实数即可。首先$\mathbf{v}^*A\mathbf{v}$是一个复数，再考虑它的转置共轭，我们有

$$(\mathbf{v}^*A\mathbf{v})^* = \mathbf{v}^*\ A^*\ \mathbf{v} = \mathbf{v}^*A\mathbf{v}.$$

这就说明$\mathbf{v}^*A\mathbf{v}$是一个实数，从而最终知道A的特征值都是实数。 □

定理 7.5.3：设A是一个Hermitian矩阵，于是不同特征值的特征空间是正交的。

证明：设$\mathbf{v}_1$ 和$\mathbf{v}_2$分别是A对应到λ_1和λ_2的特征向量，我们证明$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. 考虑$A\mathbf{v}_1 \cdot \mathbf{v}_2$有

$$A\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot A^*\mathbf{v}_2 = \mathbf{v}_1 \cdot A\mathbf{v}_2$$

同时，注意到$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$, 我们就知道

$$\lambda_1\ \mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_2\ \mathbf{v}_1 \cdot \mathbf{v}_2.$$

因为$\lambda_1 \neq \lambda_2$, 所以$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

类似于**定理** 7.1.1和**定理** 7.1.3, 我们可以给出并证明如下定理。

定理 7.5.4：设A是一个n阶复方阵，则下面的条件是等价的。

(a) A是酉矩阵。

(b) 对于任意的$\mathbf{x} \in \mathbb{C}^n$, 有$\|A\mathbf{x}\| = \|\mathbf{x}\|$.

(c) 对于任意的$\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

(d) A的列向量形成$\mathbb{C}^n$关于欧氏内积的正规正交基。

(e) A的行向量形成$\mathbb{C}^n$关于欧氏内积的正规正交基。

证明：我们证明(a), (b), (c)三者等价，而(a), (d), (e)三者等价可以作为一个好的练习。

$(a) \Longrightarrow (b)$ 根据A是酉矩阵，从而$A^*A = I$. 于是

$$\|A\mathbf{x}\|^2 = (A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{x} \cdot (A^*A\mathbf{x}) = \mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$$

$(b) \Longrightarrow (c)$ 通过计算，我们可以验证等式

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{2}(\|\mathbf{x}+\mathbf{y}\|^2 + i\|\mathbf{x}+i\mathbf{y}\|^2 - (1+i)(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)).$$

此时，$(b) \Longrightarrow (c)$是容易的。

$(c) \Longrightarrow (a)$ 我们知道

$$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot (A^*A\mathbf{y}),$$

于是条件(c)等价于对于任意的$\mathbf{x}, \mathbf{y}$, 有等式

$$\mathbf{x} \cdot (A^*A\mathbf{y} - \mathbf{y}) = 0.$$

特别地，我们取$\mathbf{x} = A^*A\mathbf{y} - \mathbf{y}$, 从而知道对于任意的$\mathbf{y}$, 有$A^*A\mathbf{y} = \mathbf{y}$成立，这就说明$A^*A = I$, 即$A$是酉矩阵。

□

之前，对于实矩阵，我们介绍了正交对角化，对于复矩阵，我们有酉对角化。一个方阵称为可酉对角化，如果存在酉方阵P 使得$P^*AP = D$是复的对角阵。我们知道对称矩阵是可正交对角化，类似地，我们有

定理 7.5.5：任意一个Hermitian 矩阵A有一个由n个特征向量组成的正规正交基，将这个正规正交基组成矩阵P, 可以使A酉对角化。

有了这个定理，我们给出酉对角化一个Hermitian矩阵A的算法。

步骤一：找出A的每一个特征空间的基底。

步骤二：对每一个特征空间的基底用Gram-Schmidt正交化，得到每个特征空间的正规正交基底。

步骤三：由步骤二得到的向量作为列向量构成矩阵P, 于是此矩阵是酉矩阵，且可以酉对角化A.

我们还有更多的一些特殊形式的矩阵，一个实方阵A 被称为是斜对称矩阵(skew-symmetric), 如果$A^T = -A$, 容易知道斜对称矩阵的对角线元素全为0. 一个复方阵阵A被称为是斜Hermitian矩阵(skew-Hermitian), 如果$A^* = -A$, 容易知道斜Hermitian 矩阵的对角线元素要么为0,要么是虚数。

在实方阵中，对称方阵和可正交对角化方阵两者是等价的。但是在复的方阵中，Hermitian矩阵只是可酉对角化矩阵的一部分。实际上，一个复方阵可以酉对角化的充分必要条件

是$A^*A = AA^*$, 这样的矩阵被称为是正规矩阵。正规矩阵是一类非常广泛的矩阵，它包含了复数情形下的Hermitian 矩阵，斜Hermitian 矩阵，酉矩阵以及实数情形下的对称矩阵，斜对称矩阵和正交矩阵。

8. 线性变换(linear transformations)

8.1 一般的线性变换

在之前，我们定义过矩阵变换$T_A : \mathbb{R}^n \to \mathbb{R}^m$如下：

$$T_A(\mathbf{x}) = A\mathbf{x},$$

其中A是$m \times n$的矩阵。我们证明了如果T是从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换，则

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \;\; T(\lambda\mathbf{u}) = \lambda T(\mathbf{u}).$$

于是，我们以此为基础来定义线性变换。

设$T : V \to W$是从向量空间V到向量空间W上的函数，则T 被称为是从V到W的线性映射，如果下面的条件满足：

(i) 对于任意的$\mathbf{u}, \mathbf{v} \in V$, 有$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(ii) 对于任意的$\mathbf{u} \in V$以及纯量λ, 有$T(\lambda\mathbf{u}) = \lambda T(\mathbf{u})$.

特别地，当$V = W$时，我们称T是V上的线性算子。

根据上面两条公理，设$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$是$V$中的向量，$\lambda_1, \lambda_2, \ldots, \lambda_r$ 是一些纯量，于是我们有

$$T(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r) = \lambda_1 T(\mathbf{v}_1) + \lambda_2 T(\mathbf{v}_2) + \cdots + \lambda_r T(\mathbf{v}_r)$$

而且，下面的定理也是很好证明的。

定理 8.1.1：设$T : V \to W$是线性变换，则

(a) $T(\mathbf{0}) = \mathbf{0}$.

(b) 对于任意的$\mathbf{u}, \mathbf{v} \in V$, 有$T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$.

定理 8.1.2：设$T : V \to W$是线性变换且V是有限维的，有一个基$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. 如果

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_r\mathbf{v}_r,$$

则我们知道

$$T(\mathbf{v}) = \lambda_1 T(\mathbf{v}_1) + \lambda_2 T(\mathbf{v}_2) + \cdots + \lambda_r T(\mathbf{v}_r),$$

于是T由其在V上的基底取值所唯一决定。

下面我们介绍关于线性映射的两个重要概念，设$T : V \to W$是线性映射，我们称集合

$$\{\mathbf{v} \mid T(\mathbf{v}) = \mathbf{0}\}$$

是映射T的核(kernel),记为$\mathrm{Ker}(T)$. 而W 中在T的像中的向量形成T的值域，记为$R(T)$.

定理 8.1.3：设$T: V \to W$是线性变换，则

(a) T的核是V的子空间。

(b) T的值域是W的子空间。

证明：(a) 我们只要证明$\mathrm{Ker}(T)$在向量的加法和纯量乘法下是封闭的即可。对于两个向量$\mathbf{u}, \mathbf{v} \in \mathrm{Ker}(T)$, 我们有

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

从而$\mathbf{u}+\mathbf{v} \in \mathrm{Ker}(T)$. 另外，对于任意的纯量$\lambda$, 我们知道

$$T(\lambda\mathbf{u}) = \lambda T(\mathbf{u}) = \lambda\mathbf{0} = \mathbf{0}.$$

从而$\lambda\mathbf{u} \in \mathrm{Ker}(T)$. 这就说明$\mathrm{Ker}(T)$是$V$的子空间。

(b) 设$\mathbf{w}_1, \mathbf{w}_2 \in W$在$T$的值域中，于是存在$\mathbf{v}_1, \mathbf{v}_2$使得

$$T(\mathbf{v}_1) = \mathbf{w}_1, \quad T(\mathbf{v}_2) = \mathbf{w}_2.$$

从而$T(\mathbf{v}_1+\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$. 这就说明$\mathbf{w}_1 + \mathbf{w}_2 \in R(T)$. 另外对于任意的纯量$\lambda$, 我们知道

$$T(\lambda\mathbf{v}_1) = \lambda T(\mathbf{v}_1) = \lambda\mathbf{w}_1.$$

这就说明$\lambda\mathbf{w}_1 \in R(T)$. 从而我们知道了$R(T)$是$W$的子空间。 □

设$T: V \to W$是线性变换，如果T的值域是有限维的空间，那么此空间的维数称为T的秩(rank)；如果T的核也是有限维空间，则称此空间的维数是T的零化度(nullity). 线性变换T的秩记成$\mathrm{rank}(T)$, 它的零化度记成$\mathrm{nullity}(T)$.

定理 8.1.4：设$T: V \to W$是线性变换，其中V是n维空间，于是我们有

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = n.$$

证明：我们只需要证明

$$\dim(R(T)) + \dim(\mathrm{Ker}(T)) = n.$$

现在假设$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$是$\mathrm{Ker}(T)$的一个基，那么自然可以扩充成$V$的一个基

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}.$$

下面我们只要证明$\{T(\mathbf{v}_{r+1}), \ldots, T(\mathbf{v}_n)\}$是$R(T)$的一个基底即可。设$\mathbf{w} \in R(T)$, 于是存在某个$\mathbf{v} \in V$使得$\mathbf{w} = T(\mathbf{v})$. 由于

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

是空间V的一个基底，因此我们知道$\mathbf{v}=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_n\mathbf{v}_n$. 注意到$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$在$T$的核中，于是$T(\mathbf{v}_1)=T(\mathbf{v}_2)=\cdots=T(\mathbf{v}_r)=0$. 从而

$$\mathbf{w}=T(\mathbf{v})=c_{r+1}T(\mathbf{v}_{r+1})+\cdots+c_nT(\mathbf{v}_n).$$

这说明$\{T(\mathbf{v}_{r+1}),\ldots,T(\mathbf{v}_n)\}$可以张成$T$的值域$R(T)$.

接下来，我们说明$\{T(\mathbf{v}_{r+1}),\ldots,T(\mathbf{v}_n)\}$是线性无关的。设存在纯量$\lambda_{r+1},\lambda_{r+2},\ldots,\lambda_n$使得

$$\lambda_{r+1}T(\mathbf{v}_{r+1})+\cdots+\lambda_nT(\mathbf{v}_n)=\mathbf{0}.$$

由于T是线性的，从而

$$T(\lambda_{r+1}\mathbf{v}_{r+1}+\cdots+\lambda_n\mathbf{v}_n)=\mathbf{0}.$$

于是$\lambda_{r+1}\mathbf{v}_{r+1}+\cdots+\lambda_n\mathbf{v}_n$在$T$的核中，从而可以写成$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r\}$的线性组合：

$$\lambda_{r+1}\mathbf{v}_{r+1}+\cdots+\lambda_n\mathbf{v}_n=\lambda_1\mathbf{v}_1+\lambda_2\mathbf{v}_2+\cdots+\lambda_r\mathbf{v}_r.$$

可以写成

$$-\lambda_1\mathbf{v}_1-\lambda_2\mathbf{v}_2-\cdots-\lambda_r\mathbf{v}_r+\lambda_{r+1}\mathbf{v}_{r+1}+\cdots+\lambda_n\mathbf{v}_n=0.$$

但因为$\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$线性无关，所以我们知道$\lambda_1=\lambda_2=\cdots=\lambda_n=0$, 这就说明了

$$\{T(\mathbf{v}_{r+1}),T(\mathbf{v}_{r+2}),\ldots,T(\mathbf{v}_n)\}$$

是线性无关的。 □

8.2 同构映射

线性变换$T:V\to W$称为是单射，如果T把V中的向量映到W中不同的向量。(注：教材中称这样的映射是一对一的(one-to-one), 但编者觉得直接翻译成单射是合适的。) 我们称T是满射的，如果W中的向量都是T中某个向量的像（即T的值域恰好为W）. 如果T即是单射，又是满射，我们就称T是同构映射，这时称V和W作为向量空间是同构的。

定理 8.2.1：如果$T:V\to W$ 是线性变换，则下面的条件是等价的：

(a) T是单射。

(b) $\ker(T)=\{\mathbf{0}\}$.

证明：$(a)\implies(b)$ 之前我们已经知道$T(\mathbf{0})=\mathbf{0}$. 由于T是单射，则没有其他向量被T映为$\mathbf{0}$, 从而$\ker(T)=\{\mathbf{0}\}$.

$(b)\implies(a)$ 现在设$\ker(T)=\{\mathbf{0}\}$, 对于两个不同的向量$\mathbf{u},\mathbf{v}$, 我们知道$\mathbf{u}-\mathbf{v}\neq\mathbf{0}$, 从而

$$T(\mathbf{u})-T(\mathbf{v})=T(\mathbf{u}-\mathbf{v})\neq\mathbf{0}.$$

即有$T(\mathbf{u})\neq T(\mathbf{v})$, 这就说明了$T$是单射。 □

定理 8.2.2：设V是一个有限维的空间，$T:V\to V$是线性算子，则下面的命题是等价的。

(a) T是单射。

(b) $\ker(T) = \{\mathbf{0}\}$.

(c) T是满射。

证明：(a)和(b)的等价已经被证明了，下面我们证明(b)和(c)是等价的。利用**定理** 8.1.4, 我们知道

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim V,$$

从而$\operatorname{rank}(T) = \dim V$等价于$\operatorname{nullity}(T) = 0$, 意味着$\ker(T) = \{\mathbf{0}\}$. 注意到$T : V \to V$是线性算子，于是$\operatorname{rank}(T) = \dim V$意味着$T$是满射。从而(b) 和(c)是等价的。

定理 8.2.3：n维的实空间都同构到$\mathbb{R}^n$.

证明：对于n维的实空间，我们选定它的一个基$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. 对于任意的$\mathbf{v} \in V$, 我们知道

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

而且这种表示方法是唯一的。于是我们考虑映射$T : V \to \mathbb{R}^n$是

$$T(\mathbf{v}) = (c_1, c_2, \ldots, c_n).$$

很容易验证T确实是一个线性映射（留作练习）。

下面我们验证T既是单射，又是满射，即需要说明$\mathrm{Ker}(T) = \{\mathbf{0}\}$, $R(T) = \mathbb{R}^n$. 对于$\mathbf{v} \in \mathrm{Ker}(T)$, $T(\mathbf{v}) = \mathbf{0}$必然要求$\mathbf{v} = \mathbf{0}$. 这就得到了$T$单射。另一方面，对于$(c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$, 我们只需要令$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, 则自然地有

$$T(\mathbf{v}) = (c_1, c_2, \ldots, c_n).$$

于是T是满射。这就说明了T同构，即n维的实空间都同构到$\mathbb{R}^n$. □

有时候，对于两个内积空间V和W, 我们会考虑它们作为内积空间的同构，即要求$T : V \to W$是线性同构，且满足

$$\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle.$$

不难证明一个n维的实内积空间同构于欧氏空间$\mathbb{R}^n$.

8.3 变换的合成和逆变换

设$T_1 : U \to V$和$T_2 : V \to W$是两个线性变换，我们知道它们作为映射是可以合成的，记成$T_2 \circ T_1$, 具体定义为：

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u})).$$

定理 8.2.3：设$T_1 : U \to V$和$T_2 : V \to W$是两个线性变换，则$T_2 \circ T_1 : U \to W$也是线性变换。

证明：设$\mathbf{u}$和$\mathbf{v}$是两个U中的向量，c是一个纯量，于是

$$(T_2 \circ T_1)(\mathbf{u} + \mathbf{v}) = T_2(T_1(\mathbf{u} + \mathbf{v})) = T_2(T_1(\mathbf{u}) + T_1(\mathbf{v})) = (T_2 \circ T_1)(\mathbf{u}) + (T_2 \circ T_1)(\mathbf{v}),$$

且同时我们有

$$(T_2 \circ T_1)(c\mathbf{u}) = T_2(T_1(c\mathbf{u})) = T_2(cT_1(\mathbf{u})) = c(T_2 \circ T_1)(\mathbf{u}),$$

这就说明了$T_2 \circ T_1 : U \to W$也是线性变换。 □

对于一个线性变换$T : V \to W$, T的值域$R(T)$是W的子空间。当T是单射的时候，对于任意的$\mathbf{w} \in R(T)$, 存在唯一的一个$\mathbf{v}$, 使得$T(\mathbf{v}) = \mathbf{w}$. 于是，我们可以构造一个新的函数$T^{-1} : R(T) \to V$, 称为$T$的逆，而且，我们不难证明$T^{-1}$也是线性映射。

定理 8.3.2：设$T_1 : U \to V$和$T_2 : V \to W$都是单射，于是

(a) $T_2 \circ T_1$是单射。

(b) $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

证明：(a) 对于U中两个不同的向量$\mathbf{u}, \mathbf{v}$, 因为T_1是单射，于是T_1将U 中两个不同的向量$\mathbf{u}$和$\mathbf{v}$映成V中两个不同的向量$T_1(\mathbf{u})$和$T_1(\mathbf{v})$, 又因为T_2是单射，于是T_2将V 中两个不同的向量$T_1(\mathbf{u})$和$T_1(\mathbf{v})$映成W中两个不同的向量$T_2(T_1(\mathbf{u}))$ 和$T_2(T_1(\mathbf{v}))$. 从而$T_2 \circ T_1$ 将U 中两个不同的向量$\mathbf{u}$和$\mathbf{v}$映成W中两个不同的向量，这就说明$T_2 \circ T_1$是单射。

(b) 我们需要证明对于$\mathbf{w} \in R(T_2 \circ T_1)$, 有

$$(T_2 \circ T_1)^{-1}(\mathbf{w}) = T_1^{-1} \circ T_2^{-1}(\mathbf{w}).$$

记$\mathbf{u} = (T_2 \circ T_1)^{-1}(\mathbf{w})$, 于是$(T_2 \circ T_1)(\mathbf{u}) = \mathbf{w}$, 即有$T_2(T_1(\mathbf{u})) = \mathbf{w}$. 这就说明

$$\mathbf{u} = T_1^{-1}(T_2^{-1}(\mathbf{w})) = (T_1^{-1} \circ T_2^{-1})(\mathbf{w}).$$

我们就完成了定理的证明。 □

8.4 一般线性变换的矩阵

我们知道矩阵对于计算来说是方便的。但是对于一般的线性映射$T : V \to W$, 我们怎么把它和矩阵联系起来呢？我们假设空间$\dim V = n$, $\dim W = m$, 在空间V上取一组基$B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, 在空间$W$ 上取一组基$B' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$. 对于$\mathbf{x} \in V$, 我们考虑$\mathbf{x}$ 在基底B下的坐标$[x]_B$, 同样地，我们可以考虑$[T(x)]$ 在B'下的坐标，我们的目的是希望找到一个$m \times n$的矩阵A使得对于任意的$\mathbf{x} \in V$, 有等式

$$A[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}.$$

根据T是线性映射，我们只要考虑$\mathbf{x}$取成基底B即可：

$$A[\mathbf{u}_1]_B = [T(\mathbf{u}_1)]_{B'}, A[\mathbf{u}_2]_B = [T(\mathbf{u}_2)]_{B'}, \ldots, A[\mathbf{u}_n]_B = [T(\mathbf{u}_n)]_{B'}.$$

注意到$\mathbf{u}_i$是第i行为1, 其余位置是0的列向量，从而

$$A[\mathbf{u}_i]_B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \cdots \\ a_{mi} \end{pmatrix} = [T(\mathbf{u}_i)]_{B'}.$$

从而我们知道矩阵A的形式为

$$A = [[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'} \mid \dots \mid [T(\mathbf{u}_n)]_{B'}].$$

称这样的矩阵是映射T相对于基底B和B'的矩阵，记成$[T]_{B',B}$, 此矩阵自然满足：

$$[T]_{B',B}\ [\mathbf{x}]_B = [T(\mathbf{x})]_{B'}.$$

特别地，当$V = W$时，$T : V \to V$是一个线性算子，此时我们取V的一个基B, 在这个基下，T 对应的变换矩阵简记为$[T]_B$.

线性映射的复合还是线性映射，那么在基底下，对应的矩阵是怎样的？

定理 8.4.1：设$T_1 : U \to V$和$T_2 : V \to W$是两个线性变换，同时假设B, B'' 和B'分别是U, V和W的基底，则

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B}$$

定理 8.4.2：设$T : V \to V$是线性算子，B是V的基底，于是下面的条件等价:

(a) T是单射。

(b) $[T]_B$是可逆矩阵。

8.5 相似

对于线性算子$T : V \to V$, 给定V的基底B，我们就得到了一个矩阵$[T]_B$. 那么自然会问，给不同的基，对应的矩阵有什么联系？能否选取一些好的基底B，使得对应的矩阵$[T]_B$的形式更简单。这是本节考虑的问题。

设$B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$和$B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$是$V$的两个基底，于是从$B$到$B'$ 以及从B'到B的转换矩阵是

$$P_{B\to B'} = [[\mathbf{u}_1]_{B'},\ [\mathbf{u}_2]_{B'},\ \dots,\ [\mathbf{u}_n]_{B'}]$$

$$P_{B'\to B} = [[\mathbf{u}'_1]_B,\ [\mathbf{u}'_2]_B,\ \dots,\ [\mathbf{u}'_n]_B]$$

且它们互为逆矩阵。

定理 8.5.1：设B和B'是线性空间V上的两个基，设$I : V \to V$是单位映射，于是

$$P_{B\to B'} = [I]_{B',B}, \quad P_{B'\to B} = [I]_{B,B'}$$

证明：现在假设$B=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$和$B'=\{\mathbf{u}'_1,\mathbf{u}'_2,\ldots,\mathbf{u}'_n\}$是$V$的两个基。于是根据定义有

$$[I]_{B',B}=[[I(\mathbf{u}_1)]_{B'}\mid[I(\mathbf{u}_2)]_{B'}\mid\ \ldots[I(\mathbf{u}_n)]_{B'}]=P_{B\to B'},$$

类似可证$P_{B'\to B}=[I]_{B,B'}$. □

定理 8.5.2：设$T:V\to V$是有限维线性空间V上的线性算子，且B和B'是V的两个基，于是

$$[T]_{B'}=P^{-1}[T]_BP$$

其中$P=P_{B'\to B}$, $P=P_{B\to B'}$.

证明：首先$T=I\circ T\circ I$, 于是由**定理** 8.4.1, 我们知道

$$[T]_{B'}=[I]_{B',B}[T]_B[I]_{B,B'}.$$

根据**定理** 8.5.1,

$$P_{B\to B'}=[I]_{B',B},\quad P_{B'\to B}=[I]_{B,B'}$$

我们就知道了

$$[T]_{B'}=P^{-1}[T]_BP.$$

其中$P=P_{B'\to B}$, $P=P_{B\to B'}$. □

我们定义两个方阵A和B是相似的，如果存在可逆矩阵P使得$B=P^{-1}AP$. 当两个方阵相似时，它们是同一个线性算子在不同基下所对应的方阵。于是它们有很多的相似不变量，比如说，行列式，可逆性，秩，零化度，迹，特征多项式，特征值，特征空间的维数等等。教材中的Table 1具体地列举了出来。

9. 主要概念的中英文名词参照

linear equation 线性方程　homogenous linear equation 齐次线性方程

system of linear equation 线性方程组　solution of a linear system 线性方程组的解

consistent 相容　inconsistent 不相容　parameter 参数

augmented matrix 增广矩阵　elementary row operation 初等行变换

reduced row echelon form 简约行阶梯型　row echelon form 行阶梯型

leading variables 首变量　free variables 自由变量

general solution to a linear system 线性方程组的通解　Gaussian elimination 高斯消元法

Gauss-Jordan elimination 高斯约当消元法　back-substitution 回代

matrix 矩阵　entries （矩阵中的）元素　column vector 列向量

row vector 行向量　square matrix 方阵　main diagonal 主对角线

matrix operation: sum,difference, scalar multiplication 矩阵的运算：和，差，纯量乘法

linear combination 线性组合　transpose 转置　trace 迹

inverse of a matrix 矩阵的逆　invertible matrix 可逆矩阵

nonsingular matrix 可逆矩阵　singular matrix 奇异矩阵

determinant 行列式　power of a matrix 矩阵的幂次　matrix polynomial 矩阵多项式

row equivalent matrices 行初等矩阵　elementary matrix 初等矩阵

lower triangular matrix 下三角矩阵　upper triangular matrix 上三角矩阵

triangular matrix 三角矩阵　diagonal matrix 对角矩阵　symmetric matrix 对称矩阵

determinant 行列式　minor 子式　cofactor 余子式　cofactor expansion 余子式展开

adjoint of a matrix 矩阵的伴随　geometric vector 几何向量　direction 方向

length 长度　initial point 起点　terminal point 终点　zero vector 零向量

vector addition 向量的加法　vector substraction 向量的减法

parallelogram rule and triangular rule 平行四边形法则和三角形法则

collinear vectors 共线的向量　components of a vector 向量的分量　coordinates 坐标

norm of a vector 向量的范数　unit vector 单位向量　distance 距离　angle 角度

dot product 点乘　Cauchy-Schwarz inequality　Cauchy-Schwarz 不等式

triangle inequality 三角不等式　parametric equations of lines 直线的参数方程

parametric equations of planes 平面的参数方程　cross product of two vectors 向量的叉乘

determinant form of cross product 叉乘的行列式形式　scalar triple product 三元纯量积

vector space 向量空间　closure under addition 加法封闭　subspace 子空间

closure under scalar multiplication 纯量乘法封闭　zero space 零空间　span 张成

solution space 解空间　trivial solution 平凡的解　linearly independent set 线性无关的集合

linear dependent set 线性相关的集合　basis 基底　finite-dimensional 有限维

infinite-dimension 无限维　coordinate vectors 坐标向量　dimension 维数

coordinate map 坐标映射　transition matrix 转换矩阵　row space 行空间

column space 列空间　null space 零化空间　general solution 通解

particular solution 特解　rank 秩　nullity 零化度

orthogonal complement 正交补　function 函数　image 像

value 值　domain 定义域　codomain 值域　transformation 变换

operator 算子　matrix transformation 矩阵变换　matrix operator 矩阵算子

reflection operator 反射算子　projection operator 投射算子　rotation operator 旋转算子

rotation matrix 旋转矩阵　expansion and compression operator （沿某个方向）伸展和收缩

dilation and contraction （沿所有方向）伸展和收缩　shear 斜移

composition of matrix transformation 矩阵变换的合成

one-to-one transformation 一对一的变换　inverse of a matrix operator 矩阵算子的逆

eigenvector 特征向量　eigenvalue 特征值　eigenspace 特征空间

characteristic polynomial 特征多项式　characteristic equation 特征等式

similarity transformation 相似变换　similarity invariant 相似不变量

similar matrices 相似的矩阵　diagonalizable matrix 可对角化矩阵
geometric multiplicity 几何重数　algebraic multiplicity 代数重数
real part （复数的）实部　imaginary part （复数的）虚部
modulus 模长　complex conjugate 复共轭　polar form （复数的）幅角形式
complex vector space 复向量空间　real matrix 实矩阵　complex matrix 复矩阵
complex Euclidean inner product 复欧氏内积　discriminant 判别式
inner product axioms 内积的公理　orthogonal vectors 正交的向量
orthogonal set 正交的集合　orthonormal set 正规正交的集合
orthogonal projection 正交投射　Gram-Schmidt process Gram-Schmidt 正交化
QR-decomposition QR分解　least square solution 最小二乘解
least square error vector 最小二乘误差向量　least squares error 最小二乘误差
best approximation 最佳逼近　normal equation 正规方程
orthogonal matrix 正交矩阵　orthogonal operator 正交算子
orthogonally similar 正交相似　orthogonally diagonalizable matrix 可正交对角化矩阵
spectral decomposition 谱分解　quadratic form 二次型
cross product term 交错项　principal axes theorem 主轴定理
conic 圆锥曲线　ellipse 椭圆　parabola 抛物线　hyperbola双曲线
central conic 中心圆锥曲线　standard position of a central conic 标准位置的中心圆锥曲线
positive definite quadratic form 正定的二次型negative definite form 负定的二次型
indefinite quadratic form 不定的二次型　principal submatrix 主子矩阵
constrained extremum 约束极值　critical point 临界点　relative minimum 相对极小值
relative maximum 相对极大值　saddle point 马鞍点　Hessian matrix Hessian矩阵
conjugate transpose 共轭转置　unitary matrix 酉矩阵　Hermitian matrix Hermitian 矩阵
unitarily diagonalizable matrix 可酉对角化矩阵　skew-symmetric matrix 斜对称矩阵
skew-Hermitian matrix 斜Hermitian矩阵　normal matrix 正规矩阵
linear transformation 线性变换　linear operator 线性算子　kernel 核　range 像