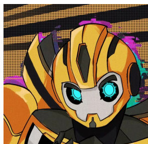
线性代数二三事

2024.7



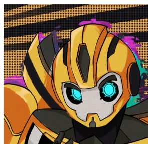
《线性代数二三事》编写人员名单

主笔 **张嘉峻** *(McGill University)* 联系方式：*zhangjohnson729@gmail.com*

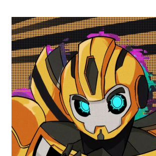
编审 **王嘉宇** *(西安交通大学)* 联系方式：*jiayuw794@gmail.com*

Jackson *(西安电子科技大学)* 联系方式：*lwj123888thu@gmail.com*

凌笙 *(武汉大学)* 联系方式：*XXXX*



目录



1 线性方程组与向量

在线性代数中，我愿把线性方程组称作是基础中的基础。把线性方程组中各个未知数的系数组合在一起，这样就构成了矩阵，同时每一个矩阵也对应了一个线性变换。所以线性代数也可以看作是一门解方程的课。在我看来任何复杂的题目在它晦涩抽象的定义背后，都有一个线性方程组在等待我们去探索，找到这个方程组，将其求解就会得到最终的结果。诚然，众里寻它（方程组）谈何容易，往往需要我们斩荆棘，破巨浪，通过对更多的概念的深入学习和了解来一点一点靠近真相。最后蓦然回首，就会发现这个我们想要的方程就在灯火阑珊处。

1.1 线性方程组的求解

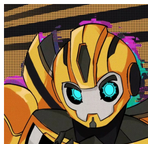
我们知道，二元一次方程组的解的几何意义是在平面直角坐标系内两个方程对应的直线的交点。我们有很多种方法去求出这个方程组的解，但是当未知数的个数逐渐增多的时候，方程组开始变得复杂且冗长，使用传统的消元方法已经无法快速地计算出答案了。因此，在线性代数二三事里面的第一节内容中，我会介绍一种独特的消元方法，使得在求解多元一次方程组时可以快速，准确地求出答案。我们不妨考虑这样的 n 元一次方程组：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{n2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

在这个方程组中 $a_{ij} : 1 \le i \le m; 1 \le j \le n; b_k : 1 \le k \le n$ 均为常数，$x_1, x_2, \cdots, x_n$ 为未知数。如果想要求解这个方程，使用我们所熟知的代入消元法会变得十分繁琐，我们不妨考虑把每个未知数前面所对应的系数进行整齐排列，然后写到一个数表中，如下图所示：

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix}$$

在这个数表中，其第 i 列正好就是原方程组中 x_i 的系数，从第一个方程开始从上至下整齐排列。这样的一个数表我们称作是矩阵，通常把这个数表用 $\mathcal{A}$ 表示，其中 $\mathcal{A}$ 共有 m 行 n 列，我们则称 $\mathcal{A}$ 为 m (行数) $\times$ n (列数）矩阵，位于该矩阵第 i 行第 j 列的元素我们写作 $\mathcal{A}_{ij}$ （注意字母的顺序，这点非常重要！）。然后，我们同样可以将剩下的未知数 $x_1, x_2, \cdots, x_n$ 写作数表的形式，我们将其写进一个由 n 行，1 列构成的数表中，从上至下分别是未知数 $x_1, x_2, \cdots$，然后我们把形如这样的 n 行 1 列数表称作 $n \times 1$ 矩阵，也可以称作



是一个向量。通常我们用字母 $\boldsymbol{x}$ 来表示由未知数组成的向量，即 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ 最后，我们运用同样的方法，把 $b_1, b_2, \cdots, b_m$ 按照相同的顺序写成一个 $m \times 1$ 向量。在当下我们只把列数为 1 的矩阵称作向量，一般用字母 $\boldsymbol{b}$ 来表示，即 $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ 然后，我们考虑形如 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 的结构，这个式子至关重要，它是奠定求解线性方程组的基础。在正式求解之前我们先提出一条法则：

设存在 $m \times n$ 矩阵 $\mathcal{A}$ 和 $k \times 1$ 向量 $\boldsymbol{x}$, 当且仅当 $k = n$ 时，运算 $\mathcal{A}\boldsymbol{x}$ 有意义，其结果为 $m \times 1$ 向量。我们不妨设

$$\mathcal{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

且 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$, 那么

$$\mathcal{A}\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \cdot \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \cdot \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \cdot \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

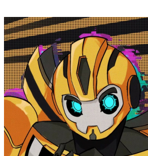
$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

同时，我们设 $\boldsymbol{x}, \boldsymbol{y}$ 为向量，c 为常数，则矩阵与向量的乘法还满足以下性质：

$$\mathcal{A}(\boldsymbol{x} + \boldsymbol{y}) = \mathcal{A}\boldsymbol{x} + \mathcal{A}\boldsymbol{y}, \;\; \mathcal{A}(c\boldsymbol{x}) = c\mathcal{A}\boldsymbol{x}$$

因此不难发现，矩阵与向量的乘法运算 $\mathcal{A}\boldsymbol{x}$ 即为把矩阵第 i 列整体当成一个向量 $\boldsymbol{v}_i$，然后用该向量去乘以原向量 $\boldsymbol{x}$ 中第 i 个位置的元素，最后加和，这也是为什么我们要强调矩阵中的列数要等于向量中的行数，否则会有无法被分配的项或元素，从而无法求解。最后得到的结果 $\boldsymbol{b}$ 则与矩阵中的行数有关。通过观察上述推理过程，与我们最开始引入的方程组做一下对比，我们是否发现二者经过运算之后完全等价?

我们回到刚才所提到的 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 的模型当中，我们可以把这个系统进一步地简化，省略 $\boldsymbol{x}$ 向量，写成 $\mathcal{A}|\boldsymbol{b}$



的形式, 像这样:

$$\left[\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} & b_2 \\ & & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} & b_m \end{array}\right]$$

这样一来，我们又得到了一个 $m \times (n+1)$ 的矩阵，我们用一条竖线来区分原方程 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 中等式的左边和右边。我们把形如这样的矩阵 $\begin{bmatrix}\mathcal{A} & \boldsymbol{b}\end{bmatrix}$ 称作是 $\mathcal{A}$ 的**增广矩阵 (Augmented Matrix)**。那么到了这一步，我们所做的全部准备工作就已经完成了，接下来我们所关心的便是如何去进行消元，从而求出原方程的解。我们在这里引入这本书里面的第一个定理：高斯消元定理。

定理 1-1 ： 高斯消元定理

在形如 $\mathcal{A}|\boldsymbol{b}$ 的矩阵中，我们可以进行下列变换，使得最后的解不受影响

① 交换任意两行

② 将一行中的所有元素全部乘以同一非零常数 k

③ 将一行中的所有元素加到另外一行相对应的元素上

我们通过一个例子来看一下高斯消元法是如何运作的:

例 1.1.1

① $\left[\begin{array}{cc|c} a & b & c \\ d & e & f \end{array}\right] \xrightarrow{\text{交换第一，二行}} \left[\begin{array}{cc|c} d & e & f \\ a & b & c \end{array}\right]$

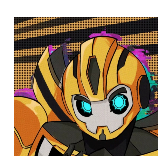
② $\left[\begin{array}{cc|c} a & b & c \\ d & e & f \end{array}\right] \xrightarrow{\text{第一行乘以常数 } k} \left[\begin{array}{cc|c} ka & kb & kc \\ d & e & f \end{array}\right]$

③ $\left[\begin{array}{cc|c} a & b & c \\ d & e & f \end{array}\right] \xrightarrow{\text{第一行对应位置的数加到第二行上}} \left[\begin{array}{cc|c} a & b & c \\ d+a & e+b & f+c \end{array}\right]$

定理 1-1 的第二，三条运算尤为重要，我们这样做的目的是尽可能地通过行与行之间的各种运算，从而产生更多的 “0”，也就是我们所谓的消元。设想经过了一些变换，某 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 被化简成为了这样的形式:

$$\left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 6 \end{array}\right] \{*\}$$

这样一来，我们便可以通过观察直接得出原方程组的解为 $\begin{cases} x_1 = 5 \\ x_2 = 6 \end{cases}$ 。因为根据我们提出的性质 1-1，对于矩阵的第一行而言其实就是方程 $1x_1 + 0x_2 = 5$; 矩阵的第二行其实就是方程 $0x_1 + 1x_2 = 6$, 因此我们便可以直接得到答案, 这种形式也是我们所想要的最终结果, 在进行消元时, 我们便可以通过定理 1-1 里面的三种运算，去尽可能地化简矩阵，得到形如（*）的式子。当然，我们需要按照一定的顺序和规则进行消元，否则最终将无法得到我们所想要的结果。在形如 $\mathcal{A}|\boldsymbol{b}$ 的式子中，消元只针对 $\mathcal{A}$ 中的元素，$\boldsymbol{b}$ 一列不能消元，因为



这一列是最终的解。**消元应该从矩阵 $\mathcal{A}$ 的第 1 列开始消元，然后按顺序依次往后。在对某一列进行消元的时候，我们只需要关注这一列的情况即可。我们的原则是尽量把第 i 列中所有除第 i 个位置的元素外的所有元素化简成 0，然后我们才能说这一列消元完毕，然后才可以进行下一列的消元**。当我们进行消元的时候，什么时候我们才知道在 $\mathcal{A}|\boldsymbol{b}$ 中原矩阵 $\mathcal{A}$ 已经化成最简形式了呢？在此之前，我们先把不能再继续消元的矩阵 $\mathcal{A}$ 称作是**最简行阶梯矩阵（Reduced Row-Echelon Form）**（也可称作行最简形矩阵）。

定义 1-1
我们把 $m \times n$ 矩阵 $\mathcal{A}$ 称作是最简行阶梯矩阵，如果其满足以下性质：
① 元素不全为零的行（非零行）的第一个非零元素所在列的下标随着行标的增大而严格增大（列标一定不小于行标）
② 元素全为零的行（如果有的话）必在矩阵的最下面几行
③ 非零行的第一个非零元素全是 1，且非零行的第一个元素 1 所在列的其余元素全为 0

为方便理解，我们直接通过几个例子来对行最简形矩阵有着更加深入的理解。

例 1.1.2
给出矩阵 $\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ 试问 $\mathcal{A}$ 是否为最简行阶梯矩阵?

解答：通过观察矩阵 $\mathcal{A}$，我们需要将定义 *1-1* 中的三条性质全部带入矩阵中进行验证。首先看第一条性质，第一行中第一个非零元素所在列标为 *1*；第二行中第一个非零元素所在列标为 *2*；第三行元素全部为 *0*。我们发现每一行中第一个非零元素所在的列标号随着行标号的增大而增大，因此满足第一条性质，同时元素全部为 *0* 的行（第三行）在矩阵的最下面，因此也满足第二条性质。在每一行中，第一个非零元素所在列的其它元素全部为 *0*；三条性质全部满足，我们称 $\mathcal{A}$ 为最简行阶梯矩阵。

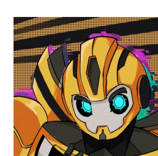
下面的几个例子同样最简行阶梯矩阵：（其中 * 代表该位置的元素可以为任何数）

$$\begin{bmatrix} 1 & * & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & * & 0 \\ 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

下面的几个例子**不是**最简行阶梯矩阵：

$$\begin{bmatrix} 1 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

读者在这里应当自行验证上述这些矩阵为什么满足（不满足）最简行阶梯矩阵



定理 1-2

任何一个矩阵 $\mathcal{A}$ 都可以经过定理 1-1 中的法则，通过一定次数的运算化简，从而得到最简行阶梯矩阵。我们将矩阵 $\mathcal{A}$ 的最简行阶梯矩阵写作 $\mathbf{RREF}(\mathcal{A})$。

接下来，我们来完成一次完整的矩阵消元：

例 1.1.3

已知矩阵 $\mathcal{A}=\begin{bmatrix}1&2&2\\1&-1&2\\2&1&-1\end{bmatrix}$，求 $\mathbf{RREF}(\mathcal{A})$。

首先，根据消元法则，我们应当从第一列开始消元。在第一列中，我们需要将位于 2，3 位置的元素消去，因此我们可以用第二行减去第一行，这样第二行第一个位置的元素就会变成 1-1=0，从而实现消元；同时我们也可用第三行减去第一行的 2 倍实现第三行第一个位置的消元。这个步骤可以写作：（我们为书写方便，一般用 R_n 来指第 n 行，C_m 来指第 m 列）

$$\begin{bmatrix}1&2&2\\1&-1&2\\2&1&-1\end{bmatrix}\xrightarrow{R_2-R_1;R_3-2R_1}\begin{bmatrix}1&2&2\\1-1&-1-2&2-2\\2-(1\times 2)&1-(2\times 2)&-1-(2\times 2)\end{bmatrix}=\begin{bmatrix}1&2&2\\0&-2&0\\0&-3&-5\end{bmatrix}$$

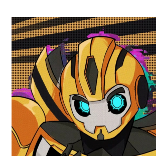
经过这一步的运算，第一列已消元完毕。此时，矩阵 $\mathcal{A}$ 是否为最简行阶梯矩阵？如果是，我们停止消元；如果不是，我们就继续进行第二列的消元，然后依次类推。直到所有列都消元完毕，或者我们提前得出最简行阶梯矩阵为止。在这里该矩阵还不为最简行阶梯矩阵（读者应自行判断，并给出依据），所以我们进行第二列的消元。在第二列中我们要消去位于位置 1，3 的元素，所以我们可以用第一行加上第二行，同时 2 倍的第三行加上 3 倍的第二行来实现对第二列的消元：

$$\begin{bmatrix}1&2&2\\0&-2&0\\0&-3&-5\end{bmatrix}\xrightarrow{R_1+R_2;2R_3+3R_2}\begin{bmatrix}1+0&2+(-2)&2+0\\0&-2&0\\2\times 0+(3\times 0)&2\times(-3)+3\times(-2)&2\times(-5)+3\times 0\end{bmatrix}=\begin{bmatrix}1&0&2\\0&-2&0\\0&0&-10\end{bmatrix}$$

此时，再进行第三列的消元：（$\mathcal{A}$ 仍不为最简行阶梯矩阵）

$$\begin{bmatrix}1&0&2\\0&-2&0\\0&0&-10\end{bmatrix}\xrightarrow{5R_1+R_3}\begin{bmatrix}5&0&0\\0&-2&0\\0&0&-10\end{bmatrix}$$

此时，矩阵 $\mathcal{A}$ 已经很像最简行阶梯矩阵了，只有一条性质还没有满足：那就是每一行中第一个非零元素为 1，所以我们再次进行化简，最终其最简行阶梯矩阵即为 $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$



在此，读者可以自行尝试求解矩阵的最简行阶梯矩阵。一个很好的素材便是例 1.1.2 中那些不满足最简行阶梯矩阵的例子。读者可以自行尝试，利用相似的方法求出这些矩阵的最简行阶梯矩阵。在最简行阶梯矩阵中，我们引入一个非常重要的概念：

定义 1-2

定义矩阵 $\mathcal{A}$ 的**秩**（***Rank***）为矩阵的最简行阶梯矩阵 ($\mathbf{RREF}(\mathcal{A})$) 中所有满足下列条件的列的数目：

在矩阵的第 i 列中，只有从上往下数位于位置 i 的元素为 1，其余元素均为 0

我们这里只是给出了一个笼统的定义，并非严格的定义。在下一节里面我们将严格定义矩阵的秩。

定理 1-3

对任意矩阵 $\mathcal{A}$ 而言，$\mathbf{rank}(\mathcal{A}) = \mathbf{rank}(\mathbf{RREF}\,(\mathcal{A}))$。

该定理的证明会在后面的章节涉及，我们当下需要知道的就是该定理的结论。这个定理告诉我们秩是矩阵本身的属性，使用高斯消元时得到的最简行阶梯矩阵与原矩阵的秩相同。但值得注意的是仅仅只有定理 1-1 中所提到的三种基本变换不会改变矩阵的秩，矩阵与矩阵相加，相乘等其他可能改变矩阵本身意义的运算同样也会改变矩阵的秩。

定义 1-3

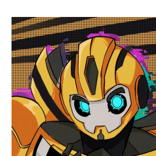
如果矩阵 $\mathcal{A}$ 的秩与 $\mathcal{A}$ 中列的数目相同，我们称矩阵 $\mathcal{A}$ 是满秩的。

满秩的概念我同样会到后续的章节中提到，现阶段我们对满秩的定义不用做过多的理解，只需直到如何求解一个矩阵的秩，以及如何判断该矩阵是否为满秩即可。

例 1.1.4

假设矩阵 $\mathcal{A} = \begin{bmatrix} 1 & 1+2 & 1+2+3 & \cdots & \cdots & 1+2+3+\cdots+2024 \\ 2 & 2+2 & 2+2+3 & \cdots & \cdots & 2+2+3+\cdots+2024 \\ 3 & 3+2 & 3+2+3 & \cdots & \cdots & 3+2+3+\cdots+2024 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 2022 & 2022+2 & 2022+2+3 & \cdots & \cdots & 2022+2+3+\cdots+2024 \end{bmatrix}$，试问 $\mathcal{A}$ 是否满秩?

在本例题中由于矩阵过于复杂，按照高斯消元法计算会变得非常繁琐，所以我们直接从定义出发。我们看到矩阵 $\mathcal{A}$ 由 *2022* 行，*2024* 列组成，根据定义，该矩阵的秩等于其行最简形矩阵中所有非零行的数目，因此无论如何矩阵 $\mathcal{A}$ 的秩无法超过 *2022*。再看满秩的概念，满秩意味着该矩阵的秩等于其矩阵的列数，然而该矩阵有 *2024* 列，大于其最大能达到的秩。所以无论如何矩阵 $\mathcal{A}$ 也无法满秩。



例 1.1.5

求解由未知数 x, y, z 构成的方程组 $\begin{cases} x+y-2z=-3 \\ 2x+y+2z=0 \\ -x-y+4z=4 \end{cases}$

运用前面所学的知识，我们应该很容易能找出在这个方程组中与 $\mathcal{A}, \boldsymbol{x}, \boldsymbol{b}$ 相对应的元素，即：

$$\mathcal{A}=\begin{bmatrix} 1 & 1 & -2 \\ 2 & 1 & 2 \\ -1 & -1 & 4 \end{bmatrix}, \quad \boldsymbol{x}=\begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \boldsymbol{b}=\begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}$$

随后，我们便知道 $\mathcal{A}$ 的增广矩阵即

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 2 & 1 & 2 & 0 \\ -1 & -1 & 4 & 4 \end{array}\right]$$

随后，我们在竖线左边部分进行消元，使用定理 *1-1* 中三条运算法则，先从第一列开始消元：

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 2 & 1 & 2 & 0 \\ -1 & -1 & 4 & 4 \end{array}\right] \xrightarrow{R_2-2R_1;\, R_3+R_1} \left[\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & -1 & 6 & 6 \\ 0 & 0 & 2 & 1 \end{array}\right]$$

注意到此时该矩阵还不为最简行阶梯矩阵，所以继续对第二列进行消元：

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & -3 \\ 0 & -1 & 6 & 6 \\ 0 & 0 & 2 & 1 \end{array}\right] \xrightarrow{R_1+R_2} \left[\begin{array}{ccc|c} 1 & 0 & 4 & 3 \\ 0 & -1 & 6 & 6 \\ 0 & 0 & 2 & 1 \end{array}\right]$$

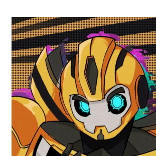
注意到此时该矩阵还不为最简行阶梯矩阵，所以继续对第三列进行消元：

$$\left[\begin{array}{ccc|c} 1 & 0 & 4 & 3 \\ 0 & -1 & 6 & 6 \\ 0 & 0 & 2 & 1 \end{array}\right] \xrightarrow{R_1-2R_3;\, R_2-3R_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 2 & 1 \end{array}\right]$$

此时所有的三列全部消元完毕（再次强调消元的时候我们只考虑对矩阵 $\mathcal{A}$ 中的列进行消元）然后我们进行如下的最后一步化简：

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 2 & 1 \end{array}\right] \xrightarrow{(-1)R_2;\, (0.5)R_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0.5 \end{array}\right]$$

至此，我们已得到 $\mathbf{RREF}(\mathcal{A})$, 所以我们可以直接写出原方程组的解：$x=1; y=-3; z=0.5$。



所以不难看出，对矩阵本身进行高斯消元和对 $\mathcal{A}|\boldsymbol{b}$ 进行消元所用到的方法完全一样，只是需要注意当我们在进行定理 1-1 中的运算时，应当把 $\boldsymbol{b}$ 中对应的元素也考虑在内，但消元仅限于 $\mathcal{A}$。

随后，我们来看一下线性方程组所产生的解的情况。我们知道一个二元一次方程组的解即为在平面内两条直线的交点，因此研究二元一次方程组里解的个数也就等同于研究两条直线的位置关系。如果两条直线平行且不重合，意味着二者没有交点，因此原方程组无解；如果两条直线完全重合，那么该直线上的每一个点都是满足条件的解，此时原方程组有无穷多组解；如果两条直线相交，那么二者在平面直角坐标系内的唯一交点即为方程组的解，此时原方程组有且仅有一个解。

因此我们发现，对于二元一次方程组而言，其解有且仅有三种情况：无解；有无穷多组解；无解。其实对于未知数更多的方程组而言，该情况同样适用。

定理 1-4

对于任意一个线性方程组而言，其解的个数只有三种情况：无解，有无穷多组解，有且仅有一个解。

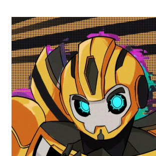
一般地，我们知道如果有 n 个未知数需要求解，我们则需要 n 个方程来进行消元和求解。假如存在 6 个未知数，却只有 5 个方程，那么这个方程组一定会有无穷多组解，或者无解。就好比一个二元一次方程组中如果我们发现只有一个方程，那么这个方程组的解集就是该直线上的所有点；同样在三元一次方程组中如果只有两个方程，那么这两个方程分别可以代表两个平面，这两个平面的交线就是满足条件的所有解，此时交线上所有的点同样是方程组的解；如果这个三元一次方程组中只有一个方程，那么首先这个方程所代表的直线上的所有点构成原方程组的解，同时我们可以对该直线进行平移，通过特定方向平移（即该方程整体乘以一个系数 k），所构成的新的直线也是原方程的解，那么所有满足条件的直线有无穷多个，这些直线构成了一个平面，平面上的所有点就是原方程组的解集。同样的的方法可以推广到未知数更多的方程组中。我们不难发现，在未知数给定的情况下，方程越少，解的“随意性”就越高，方程越多，限制条件越苛刻，解的“自由度”就越少，如果方程数过多，还可能会产生无解的情况。我们之所以要进行高斯消元，就是为了把原本复杂的方程结构得以简化，使得我们一眼就能看出原方程中哪些是“多余的”，从而大大简化我们的计算量。

假设我们有这样的一个三元一次方程组：$\begin{cases} x+2y+z=1 \\ 2x-y+2z=3 \\ x-8y+z=3 \end{cases}$，我们可能会认为这个方程组有唯一解，因为有三个方程。但其实这个方程有无穷多组解，我们不能被方程的表象所迷惑，一定要对这个方程进行简化，从而拨云见日，看出真相。经过对这个方程组的第一列消元之后，这个方程组可以写成形如下式的结构：

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -5 & 0 & 1 \\ 0 & -10 & 0 & 2 \end{array}\right]$$

我们发现第二，三行完全等价，因此原方程组中看似有三个方程，实际上对解集起作用的只有两个方程，在这种情况下，正如之前所讨论的情况，这个方程组会有无穷多组解（前提是两个平面相交）。我们继续进行消元，得到最简行阶梯矩阵如下：

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 7/5 \\ 0 & 1 & 0 & -0.2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$



那么此时，我们发现第三行全部为零，因此该矩阵的秩为 2。我们已经知道，此时原方程有无穷多组解，我们此时要做的是将其用含参数的式子把该解集表示出来。参数的选取不能随意，通过观察行最简矩阵，如果第 i 行元素全部为 0，那么第 i 列所对应的未知数我们就将其当作参数。满足这种条件的未知数我们称之为**自由变量（又称滞后变量）(Free Variables)**。其中满足第 i 列中位置 i 的元素为 1，其余元素均为零的未知数我们称之为**前导变量 (Leading Variables)**。因此在行最简矩阵中，我们称矩阵的秩为前导变量的数量，自由变量的数量我们称作是矩阵的**零化度 (Nullity)**。如果有多行元素全都为零，那么便针对每一行单独设出一个参数。如果我们发现原方程组中有 n 个未知数 m 个方程，$(m < n)$。那么为了方便运算，我们一般在矩阵中用 0 将行数与列数补齐。在本题中因为第三行元素为 0，我们将第三列对应的未知数 (z) 选为参数，然后把其他列中的未知数用含这些参数的式子表示出来。在这道题中，我们就有：$\begin{cases} x+z=7/5 \\ y=-0.2 \end{cases}$ 也就是形如 $\begin{cases} x=-z+7/5 \\ y=-0.2 \\ z=z \end{cases}$ 的形式。为了便于观察，我们一般把含有参数的项和常数分开，并且写成向量的形式。同时为了防止参数 z 和原方程中的未知数发生混淆，我们往往用其他的字母 (如 s,t,r,q 等)，因此原方程的解集如下：

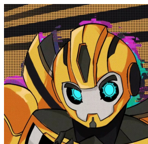
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7/5 \\ -0.2 \\ 0 \end{bmatrix} + s\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}，\text{其中 } s \text{ 为参数，} s \in \mathbb{R}$$

这样一来，原方程的解的结构就清晰可见了，它代表一条经过点 $(7/5, -1/5, 0)$ 的直线，并且该直线的方向向量为 $(-1, 0, 1)$。原方程组的解集即构成这条直线上所有的点。

例 1.1.6 求出线性方程组 $\begin{cases} x+2y+z-w=0 \\ -x+y-2z+2w=0 \end{cases}$ 的解集。

我们直接写出相应的矩阵形式，同时将第三，四行用元素 *0* 补齐，即 $\left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 0 \\ -1 & 1 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$，经过消元，我们得到其最简行阶梯矩阵为 $\left[\begin{array}{cccc|c} 1 & 0 & \frac{5}{3} & -\frac{5}{3} & 0 \\ 0 & 1 & -\frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$，此时我们将 z, w（自由变量）作为参数，分别设 $z=s, w=t$，那么我们有 $\begin{cases} x=-\frac{5}{3}s+\frac{5}{3}t \\ y=\frac{1}{3}s-\frac{1}{3}t \\ z=s \\ w=t \end{cases}$，将其写成向量的形式，因此原方程组的解集为 $\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = s\begin{bmatrix} -\frac{5}{3} \\ \frac{1}{3} \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} \frac{5}{3} \\ -\frac{1}{3} \\ 0 \\ 1 \end{bmatrix}$

那么，是不是所有的方程组都有解？我们当然知道存在无解的情况，那么无解的情况再矩阵里面是怎么表



现出来的?

例 1.1.7 求出线性方程组 $\begin{cases} x+2y+z-w=4 \\ 2x-y+2w=3 \\ 4x+3y+2z=5 \end{cases}$ 的解集。

我们直接写出相应的矩阵形式，同时将第四行用元素 0 补齐，即 $\left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 4 \\ 2 & -1 & 0 & 2 & 3 \\ 4 & 3 & 2 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$ 我们对第一列进行消元，得到 $\left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 4 \\ 0 & -5 & -2 & 4 & -5 \\ 0 & -5 & -2 & 4 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$ 此时我们注意到第二行和第三行中有形如 $-5y-2z+4w=-5$ 和 $-5y-2z+4w=-9$ 的两个平面的方程，这两个平面相互平行，因此没有交点，由此也就得出原方程组无解。

至此，我们已经学完线性方程组的求解方法，这一节是线性代数中的重中之重，所以读者应做大量的练习，让自己尽快熟能生巧。

1.1 练习

1. 设 $a \in \mathbb{R}$, 存在如下的线性方程组：$\begin{cases} x+y+3z=a \\ ax+y+5z=4 \\ x+ay+4z=a \end{cases}$ 分别找出所有满足条件的 a 的取值，使得原方程组：① 有唯一解；② 有无穷多组解；③ 无解。

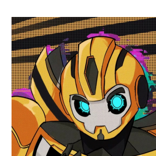
2. 设平面直角坐标系内的三条曲线：$\mathcal{C}_1: x^2+xy-y^2-1=0$, $\mathcal{C}_2: 2x^2-xy+3y^2-13=0$, $\mathcal{C}_3: x^2+3xy+2y^2=0$。试求出三条曲线的交点坐标。

3. 运用求解线性方程组的知识以及化学反应前后原子数目守恒的原则，配平下列氧化还原反应方程式：

$$\mathrm{NH_3 + CuO \longrightarrow N_2 + Cu + H_2O}$$

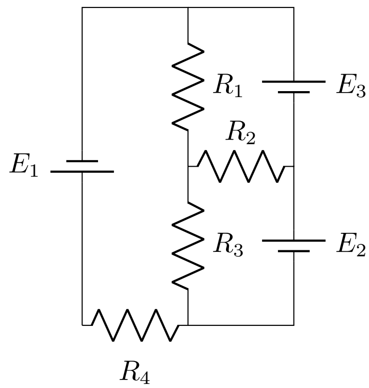
$$\mathrm{Pb(N_3)_2 + Cr(MnO_4)_2 \longrightarrow Cr_2O_3 + MnO_2 + Pb_3O_4 + H_2O}$$

4. 设矩阵 $\mathcal{A} = \begin{bmatrix} 1 & 1 & 2 & a^2 \\ 1 & 1-a & 2 & 0 \\ 2 & 2-a & 6-a & 4 \end{bmatrix}$, 其中 $a \in \mathbb{R}$。讨论 $\mathcal{A}$ 的秩的所有可能取值。

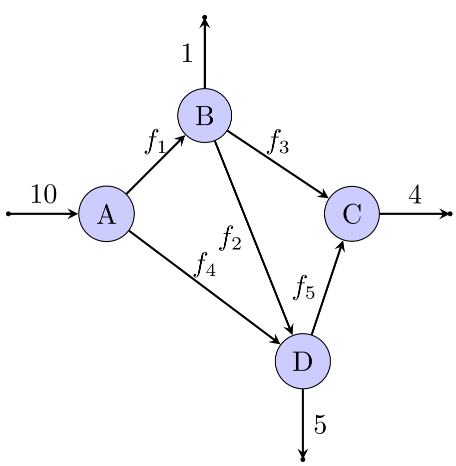


5. 在如下图所示的闭合电路中，电源 E_1, E_2, E_3 的电动势均为 $1V$ 且内阻不计, 电阻 $R_1 = R_2 = R_3 = R_4 = 1\Omega$, 不考虑导线上的分压和热损失，据此判断流经 E_1, E_2, E_3 的电流大小和方向。

(*HINTS* : ① 根据 *Kirchhoff's Laws*, 对于闭合电路的每一个节点，流入该节点的电流标量等于流出该节点的电流标量；同时任意一个闭合环路整体电势差的变化量为零。② 根据 *Ohm's Law*, 流经电阻 R 的电压 V 与其电流 I 称反比，即 $V = RI$。)

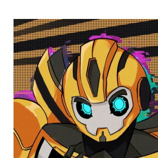


6. 在如下图所示的网格中有 A, B, C, D 四个节点通过单向流连结, 由外界流入 A 节点的流量和由 B, C, D 节点流出至外界的流量已知（流量在已在箭头上标出）。已知在此系统内部流入一节点的流量等于流出该节点的流量，在本题中，流量 $\tau \geq 0$。



(i) 据此求出流量 f_1, f_2, f_3, f_4, f_5 的所有可能取值，其方向已在图中标出。

(ii) 现规定 f_2 上的流量不得超过 3, 在此情况下可以流经 f_3,f_4 的最大流量之和是多少?



1.2 向量和向量组之间的关系

在上一节中，我们提出了形如 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 的式子，那么从这一节开始，我们将从更加抽象的角度去理解这样一个式子所代表的意义。我们提出过以下的运算法则：

设存在 $m \times n$ 矩阵 $\mathcal{A}$ 和 $k \times 1$ 向量 $\boldsymbol{x}$，当且仅当 $k = n$ 时，乘法 $\mathcal{A}\boldsymbol{x}$ 有意义，其结果为 $m \times 1$ 向量。我们不妨设

$$\mathcal{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

且 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$，那么

$$\mathcal{A}\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \cdot \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \cdot \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \cdot \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

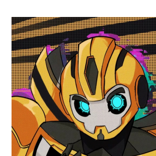
$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

我们可以发现，在上述的表达式中矩阵 $\mathcal{A}$ 的每一列都可以当做为一个向量看待，我们将这些向量称作**列向量（Column Vectors)**，这样一来，我们就有了一种全新的书写矩阵的方式。我们可以把矩阵写成是由若干列向量所构成的形式，即 $\mathcal{A} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n]$，这样一来，我们可以将 $\mathcal{A}\boldsymbol{x}$ 书写成 $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n$，其中 $\boldsymbol{v}_i$ 则对应原矩阵的第 i 列，然后 x_i 则是原方程组中的未知数 x_i。这样以来，我们就可以引入本节的第一个重要定义：

定义 1-4
设存在行数（维数）相同的向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$，如果存在另一个向量 $\boldsymbol{u}$ 和常数 $c_1, c_2, \cdots, c_n$，使得

$$\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

我们则称 $\boldsymbol{u}$ 是由 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 所构成的一个**线性组合** ***(Linear Combination)***。满足条件的所有由 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 构成的线性组合的集合我们记作 $\textbf{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n)$。



根据这个定义，我们知道在上一页所展示的运算法则中，$\boldsymbol{b}$ 则为矩阵 $\mathcal{A}$ 中列向量所构成的线性组合。我们通过观察线性组合的定义，很容易将其与线性方程组的解所联系起来，这种联系便是一个重要的定理：

定理 1-5

设矩阵 $\mathcal{A}=\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{bmatrix}$，$\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$，对于给定的向量 $\boldsymbol{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}$，线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 有解的充要条件为

$$\boldsymbol{b}\in \textbf{Span}(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n)$$

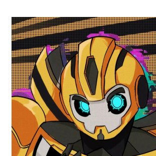
我们通过一个三元一次方程组的例子来加深我们对定理 1-5 的理解：假设一个三元一次方程组的解集为

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}=s\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}+t\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}: s,t\in\mathbb{R}$$

我们知道，这个三元一次方程组的解集代表一个经过原点的平面，平面上的点即为所有满足条件的解。因此，这个三元一次方程组其实完全可以简化成一个方程，那就是表示该平面的方程。不难求出这个平面所表示的方程为 $x-y+z=0$。根据定理 1-5，我们知道该方程有解的充要条件是 $\boldsymbol{b}\in\textbf{Span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right)$，也就是说任何在该平面以外的点都不是原方程组的解。因此我们取在平面以外的点 $\boldsymbol{u}=\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$，将其带入方程 $x-y+z=0$ 发现 $2-1+1=0$，发现原方程组无解。如果我们取平面以内的任意点 $\begin{bmatrix} s+t \\ t \\ -s \end{bmatrix}: s,t\in\mathbb{R}$ 将其带入方程 $x-y+z=0$，则发现 $(s+t)-t-s=0$ 成立。

我们也可以通过一个生活中的例子来理解线性组合，在美术学中我们经常把红色，黄色和蓝色称作是色彩的三原色。我们可以用这三种颜色通过一定比例的混合从而得到其他的颜色。比如绿色就由一份蓝色和一份黄色通过均匀混合得到的，我们便可以说绿色是由红色，蓝色和黄色三种颜色组成的一个‘线性组合’。之所以将红色，黄色和蓝色称作是三原色，是因为其中的任意一种颜色都不能由另外两种颜色混合而得到，所以我们也可以说蓝色不是由红色和黄色两种颜色组成的一个线性组合。

一种很常见的题目便是给出向量 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots\boldsymbol{v}_n$，然后我们需要求另一个向量 $\boldsymbol{u}$ 是否为这几个向量的一个线性组合，这种题型就可以参照线性方程组的解法，写成矩阵的形式 $\left[\begin{array}{cccc|c} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n & \boldsymbol{u} \end{array}\right]$。然后运用高斯消元法进行求解，如果该方程组无解，那么 $\boldsymbol{u}$ 则不是 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 所组成的一个线性组合，如果有解，那么 $\boldsymbol{u}$ 则是 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 所组成的一个线性组合。



例 1.2.1

已知向量 $\boldsymbol{v}_1=\begin{bmatrix}1\\0\\-1\end{bmatrix},\boldsymbol{v}_2=\begin{bmatrix}-1\\1\\0\end{bmatrix},\boldsymbol{v}_3=\begin{bmatrix}0\\-1\\1\end{bmatrix}$, 那么 $\boldsymbol{u}=\begin{bmatrix}1\\0\\3\end{bmatrix}$ 是否为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 的一个线性组合?

按照上一页所提到的方法，我们重点研究线性方程组 $\left[\begin{array}{ccc|c}1&-1&0&1\\0&1&-1&0\\-1&0&1&3\end{array}\right]$。我们对第一列进行消元，随后发现我们得到 $\left[\begin{array}{ccc|c}1&-1&0&1\\0&1&-1&0\\0&-1&1&4\end{array}\right]$。注意到第二行和第三行中我们得到了无解的情况，因此 $\boldsymbol{u}$ 不是 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 的一个线性组合。

例 1.2.2

已知向量 $\boldsymbol{v}_1=\begin{bmatrix}1\\0\\1\end{bmatrix},\boldsymbol{v}_2=\begin{bmatrix}1\\1\\0\end{bmatrix},\boldsymbol{v}_3=\begin{bmatrix}0\\1\\1\end{bmatrix}$, 那么 $\boldsymbol{u}=\begin{bmatrix}1\\0\\3\end{bmatrix}$ 是否为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 的一个线性组合?

我们重点研究线性方程组 $\left[\begin{array}{ccc|c}1&1&0&1\\0&1&1&0\\1&0&1&3\end{array}\right]$。我们按照消元法则进行消元，得到最后的行最简形矩阵: $\left[\begin{array}{ccc|c}1&0&0&2\\0&1&0&-1\\0&0&1&1\end{array}\right]$。因此我们不难发现，存在 $\boldsymbol{u}=2\boldsymbol{v}_1-\boldsymbol{v}_2+\boldsymbol{v}_3$ 的线性组合。

在上面的两个例子中，在 $\boldsymbol{u}$ 相同的情况下，为什么 1.2.1 中无法构成线性组合，而 1.2.2 中却可以构成线性组合？我们自然想到可能和 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 的选取有关。但是它们之间到底有着怎么样的关系？我们该如何更加快速地判断线性组合地存在与否呢？在此，我会提出一个目前最为重要的一个定义：

定义 1-5

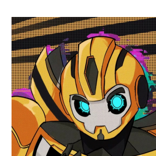
设存在维数相同的向量组 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$，我们称该向量组彼此 **线性无关** ***(Linearly Independent)*** 当且仅当线性组合

$$c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots c_n\boldsymbol{v}_n=0$$

的**唯一**解是 $c_1=c_2=\cdots=c_n=0$.

定义 1-6

我们称向量组**线性相关** ***(Linearly Dependent)***, 如果该向量组不满足彼此线性无关。



当我们了解这个定义之后，再回看例 1.2.1，在这个例子中，我们先判断 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 彼此是否线性无关。为了解决这个问题，我们自然要考虑下列方程组的解：

$$c_1\begin{bmatrix}1\\0\\-1\end{bmatrix}+c_2\begin{bmatrix}-1\\1\\0\end{bmatrix}+c_3\begin{bmatrix}0\\-1\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

然后，我们可以自然而然地把上述式子写成形如下式的一个矩阵中：

$$\left[\begin{array}{ccc|c}1&0&-1&0\\-1&1&0&0\\0&-1&1&0\end{array}\right]$$

然后，经过高斯消元，我们得到其行最简形矩阵为

$$\left[\begin{array}{ccc|c}1&0&-1&0\\0&1&-1&0\\0&0&0&0\end{array}\right]$$

那么我们可以得到其解为 $\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=s\begin{bmatrix}1\\1\\1\end{bmatrix}: s\in\mathbb{R}$。因此我们看到原方程组的解集是一条经过原点的直线，直线上所有的点都是满足条件的解，这与线性无关的概念不符。在线性无关中我们要求唯一解必须是 $c_1 = c_2 = c_3 = 0$, 但显然在 1.2.1 中我们有很多组非零解，因此 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 线性相关。同样我们可以对 1.2.2 进行验证（读者可以自行验证），得到 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 线性无关。到了这里，我们能否说如果这些向量彼此线性无关，那么就一定能构成另外一个向量的线性组合；如果这些向量彼此线性相关，那么就一定不能构成另一个向量的线性组合呢？我给出的答案是否定的，这些奥秘我们会在下一节内容中进行详细地探索，到时候便会豁然开朗。

我们再回到例 1.2.1 中，这里面 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 彼此线性无关，那么我们就假设存在如下的线性组合：

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0 \ *$$

然后我们通过移项，可以得到

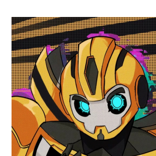
$$c_1\boldsymbol{v}_1 = -c_2\boldsymbol{v}_2 - c_3\boldsymbol{v}_3$$

因为我们知道 c_1, c_2, c_3 含有无穷多组的非零解，那么我们可以选取满足 $c_1 \neq 0$ 的一组解（如果这种解存在）。于是我们等式两边同时除以 c_1，可以得到

$$\boldsymbol{v}_1 = -\frac{c_2}{c_1}\boldsymbol{v}_2 - \frac{c_3}{c_1}\boldsymbol{v}_3$$

根据定义，我们可以推出 $\boldsymbol{v}_1 \in \mathbf{Span}(\boldsymbol{v}_2, \boldsymbol{v}_3)$。然后在（*）一式中我们同样可以把 $\boldsymbol{v}_2, \boldsymbol{v}_3$ 移到等式一侧，然后任意取一组使得 c_2, c_3 不为零的解 (如果这种解存在)，同样可以得到 $\boldsymbol{v}_2 \in \mathbf{Span}(\boldsymbol{v}_1, \boldsymbol{v}_3)$; $\boldsymbol{v}_3 \in \mathbf{Span}\,(\boldsymbol{v}_2, \boldsymbol{v}_1)$。

但是在 1.2.2 的例子中，由于 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 彼此线性无关，因此线性组合 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0$ 的唯一解是 $c_1 = c_2 = c_3 = 0$, 所以我们没有办法作上述式子里面的变形。但我们也可以尝试求解 $v_1 = a\boldsymbol{v}_2 + b\boldsymbol{v}_3$ (读者可以尝试)，运用线性组合和方程组的解的关系，我们可以得到 $v_1 \notin \mathbf{Span}(\boldsymbol{v}_2, \boldsymbol{v}_3)$; $\boldsymbol{v}_2 \notin \mathbf{Span}(\boldsymbol{v}_1, \boldsymbol{v}_3)$; $\boldsymbol{v}_3 \notin \mathbf{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2)$。



定理 1-6

设同维数的向量组 $\beta=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$，则

① 向量组 β 线性无关的充要条件是对任意的 $\boldsymbol{v}_i\in\beta$, $\boldsymbol{v}_i$ 均不能组成由剩余向量的线性组合。

② 向量组 β 线性相关的充要条件是存在 $\boldsymbol{v}_i\in\beta$, $\boldsymbol{v}_i$ 使得能组成由剩余向量的线性组合。

这个定理的证明在上一页就有所涉及，只不过上一页中我着重讨论了只有三个向量的情况，读者应尝试证明向量个数为 n 的情况，其中一种很好的方式便是利用数学归纳法。另外还有一点值得关注，为什么我在线性无关里面提到 “对任意的 $\boldsymbol{v}_i\in\beta$” ，而在线性相关里面提到” 如果存在 $\boldsymbol{v}_i\in\beta$”？这一点读者也可以进行思考，其中我举一个相似的例子以作为类比：设函数 $f(x),g(x)$。已知在区间 $[1,2]$ 上，对任意的 x，满足 $f(x)\leq g(x)$；已知在区间 $[3,4]$ 上，存在 x，使得 $f(x)\leq g(x)$。那么 $f(x),g(x)$ 在这两个区间上的大致图像应满足什么关系？

我们不妨再次回到例 1.2.1 上，我们发现向量 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 彼此线性相关，意味着其中一个或多个向量都是剩下两个向量构成的线性组合。这也告诉我们这三个向量所表达的内容和删去一个特定向量之后所剩的两个向量所表达的内容相同。因此这里面有的向量是多余的，可谓滥竽充数。因此我们不能被 “金玉其外，败絮其中” 的表象所迷惑，一定要进行深刻的分析从而把这些 “滥竽充数” 的向量抓出来。那么按照这个原理，对于一个线性方程组而言，虽然有很多个方程，但是里面也不乏滥竽充数的伪君子，很有可能其中一个方程也是另外几个方程的线性组合。所以这种情况下我们实际起作用的方程就变少了。我也说过在未知数一定的情况下，方程越少，留给解集发挥的 “自由度” 就越大，就会在解集中产生更多的参数，也因此会产生更多的元素全部为 0 的行数。根据矩阵秩的定义，产生更多的非零行会使得自由变量数目的增多，势必会使得矩阵的秩减小。如果我们存在一个完美的方程组，使得最后行最简形矩阵中非零行数的个数为 0，这也就意味着里面没有滥竽充数的方程，每一个方程都对解集起着不可磨灭的作用，这也就说明这些方程彼此不存在线性组合，意味着它们线性无关。

定理 1-7

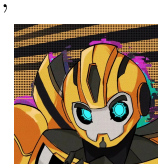
设同维数向量组 $\beta=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$ 线性无关，则由这些列向量构成的矩阵 $\mathcal{A}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ 满秩。

由定理 1-7 我们还可以引出一条推论：

定理 1-7 (a)

在任意矩阵 $\mathcal{A}$ 中，矩阵的秩等于该矩阵中彼此线性无关的列向量的最大数目。

其实我们也可以把矩阵 $\mathcal{A}$ 中的每一行当成一个向量，即**行向量 (Row Vectors)**, 这样一来矩阵也可以记作 $\mathcal{A}=\begin{bmatrix}-- & \boldsymbol{v}_1 & --\\ & \vdots & \\ -- & \boldsymbol{v}_n & --\end{bmatrix}$, 正因为有这两种不同的表达方式，我们在谈论矩阵中的向量时一定要说明是行向量还是列向量。那么二者之间有什么关系呢? 我们回想矩阵的秩的定义，在上一节里面我们只给出了一个笼统的定义，现在我们可以用本节的知识来重新定义矩阵的秩：矩阵的秩为该矩阵中线性无关的列向量的最大数目，



我们可以类比这个定义再给出一个**行秩 (Row Rank)** 的定义，我们定义矩阵的行秩等于该矩阵中线性无关的行向量的最大数目，记作 **r-Rank**。同理我们可以定义**列秩 (Column Rank)**，记作 **c-Rank**。在一个矩阵中我们知道列秩等于其线性无关的列向量的最大数，同时根据定理 1-3，我们知道这个数值同样等于该矩阵的最简行阶梯形矩阵中线性无关的列向量的最大值。我们可以由最简行阶梯形矩阵的性质得知，此时线性无关的行向量数与线性无关的列向量数相等（这一点读者可以自行通过写出一些矩阵进行验证，严格的数学证明涉及高等代数中的知识，如果条件允许在本书合适的部分会进行证明）。因此我们又的出一条重要的定理：

定理 1-8
对于任意矩阵 $\mathcal{A}$ 而言，其行秩与列秩相等。

所以我们直接用矩阵的秩来泛指行秩与列秩。

除此之外，我们还想再研究一种特殊的方程组，$\mathcal{A}\boldsymbol{x} = \mathbf{0}$, 形如这样的方程组我们将其称作**线性齐次方程组 (Linear Homogeneous Systems of Equations)**, 那么很显然，$\boldsymbol{x} = \mathbf{0}$ （零向量）为该方程组的一个解。如果该方程组的解有 $\boldsymbol{x} = \mathbf{0}$, 我们则称该解为一个**平凡解 (Trivial Solution)**, 其余的解我们称之为**非平凡解 (Non-trivial Solution)**。

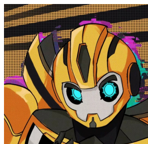
定理 1-9
若线性齐次方程组 $\mathcal{A}\boldsymbol{x} = 0$ 有唯一解 $\boldsymbol{x} = \mathbf{0}$, 则矩阵 $\mathcal{A}$ 中的列向量彼此线性无关。

例 1.2.3
假设向量组 $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ k \\ -1 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} k \\ 1 \\ 0 \end{bmatrix}$ 线性无关，求满足条件的所有 k 的取值。

我们先构造由 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 组成的线性齐次方程组 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0$, 由于这三个向量彼此线性无关，因此根据定义我们得知该方程组的唯一解为 $c_1 = 0; c_2 = 0, c_3 = 0$, 于是我们可以把原方程组写成如下所示的方程组：$\begin{cases} c_1 + kc_3 = 0 \\ kc_2 + c_3 = 0 \\ c_1 - c_2 = 0 \end{cases}$，再将其写进矩阵中，有 $\left[\begin{array}{ccc|c} 1 & 0 & k & 0 \\ 0 & k & 1 & 0 \\ 1 & -1 & 0 & 0 \end{array}\right]$ 利用高斯消元法，我们最终得到

$$\left[\begin{array}{ccc|c} \frac{1}{k} - k & 0 & 0 & 0 \\ 0 & k(1-k^2) & 0 & 0 \\ 0 & 0 & 1-k^2 & 0 \end{array}\right]$$

根据定义，我们应满足 $\frac{1}{k} - k \neq 0; k(1-k^2) \neq 0; 1-k^2 \neq 0$, 因此我们解得 $k \neq 0, 1, -1$。但需要注意的是我们解得的 $k \neq 0$ 是因为 $\frac{1}{k}$ 的限制，因此我们需要再额外验证 $k = 0$ 是否可行。最终答案为 $k \neq 1, -1$。



例 1.2.4
设向量 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ 彼此线性无关，证明向量 $\boldsymbol{u},\boldsymbol{u}+\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$ 彼此同样线性无关。

证明. 首先根据定义，我们知道

$$c_1\boldsymbol{u}+c_2\boldsymbol{v}+c_3\boldsymbol{w}=0 \quad (1)$$

的唯一解是 $c_1=c_2=c_3=0$。我们随后假设存在这样的线性齐次方程：$\lambda_1\boldsymbol{u}+\lambda_2(\boldsymbol{u}+\boldsymbol{v})+\lambda_3(\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w})=0$，将该方程整理，移项之后我们得到

$$(\lambda_1+\lambda_2+\lambda_3)\boldsymbol{u}+(\lambda_2+\lambda_3)\boldsymbol{v}+\lambda_3\boldsymbol{w}=0 \quad (2)$$

通过将（*1*），（*2*）式对比，我们发现二者结构在除了向量之前的系数不同之外其余均等价。又因为 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ 彼此线性无关，根据定义，因此在（*1*），（*2*）式中无论向量前面的系数如何选取，最终这些系数只能取零。所以我们有这样的线性齐次方程组：

$$\begin{cases}\lambda_1+\lambda_2+\lambda_3=0\\ \lambda_2+\lambda_3=0\\ \lambda_3=0\end{cases} \quad \text{解得} \quad \begin{cases}\lambda_1=0\\ \lambda_2=0\\ \lambda_3=0\end{cases}$$

因此，线性齐次方程 $\lambda_1\boldsymbol{u}+\lambda_2(\boldsymbol{u}+\boldsymbol{v})+\lambda_3(\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w})=0$ 有唯一平凡解，因此 $\boldsymbol{u},\boldsymbol{u}+\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$ 彼此线性无关。

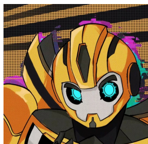
■

如果读者有兴趣，可以自行尝试证明 $1,1+x,1+x+x^2,1+x+x^2+x^3,\cdots,1+x+\cdots+x^n$ 彼此线性无关，用到的方法和例 1.2.3 完全相同，只不过向量的个数发生了变化。使用数学归纳法同样也是一个很好的证明途径。

1.2 练习

---

1. 根据线性无关的定义，零向量自身是否线性无关？非零向量自身是否线性无关？



2.（单项选择题） 假设向量组 $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ 线性无关，则下列选项中线性无关的向量组为

（A）向量组 $\boldsymbol{x},\boldsymbol{x}-\boldsymbol{y},\boldsymbol{x}-\boldsymbol{y}-\boldsymbol{z}$

（B）向量组 $\boldsymbol{x}+\boldsymbol{y},\boldsymbol{y}+\boldsymbol{z},\boldsymbol{x}+\boldsymbol{z}$

（C）向量组 $\boldsymbol{x}-\boldsymbol{y},\boldsymbol{y}-\boldsymbol{z},\boldsymbol{z}-\boldsymbol{x}$

（D）向量组 $\boldsymbol{x}+\boldsymbol{y},\boldsymbol{y}-\boldsymbol{z},\boldsymbol{x}+2\boldsymbol{z}$

3.（单项选择题） 设函数 $f(x)=\sin(x);g(x)=\cos(x);h(x)=\tan(x)$, 给定 $x\in[-\frac{\pi}{2},\frac{\pi}{2}]$, 则在给出的这三个函数中，下列说法正确的是

（A）$f(x),g(x)$ 线性无关

（B）$f(x),h(x)$ 线性无关

（C）$g(x),h(x)$ 线性无关

(D) $f(x)+g(x),f(x)-g(x)$ 线性无关

4. 给出向量 $\boldsymbol{v_1}=\begin{bmatrix}1\\-1\\0\end{bmatrix},\boldsymbol{v_2}=\begin{bmatrix}1\\0\\-1\end{bmatrix}$, 试写出任意一个向量 $\boldsymbol{v_3}$, 使得 $\boldsymbol{v_1},\boldsymbol{v_2},\boldsymbol{v_3}$ 线性无关。

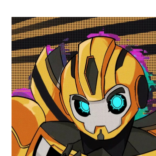
5. 假设向量组 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$ 线性无关，向量 $\boldsymbol{u}\in\textbf{Span}\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$, 且 $\boldsymbol{u}$ 为非零向量。据此判断并讨论向量组 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{n-1},\boldsymbol{u}$ 是否线性无关。

6. 设向量组 $\beta=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$, γ 为 β 的一个子集，证明：若 β 中的向量彼此线性相关，则存在满足一定条件的 γ, 使得 γ 中的向量线性无关。读者可以自行思考：如果 β 线性无关，那么其子集 γ 又满足什么性质？

7. 设 δ 是作用在矩阵 $\mathcal{A}$ 中的列向量上的一种运算，运算 δ 有以下性质：

$$\begin{cases}\delta\left(\begin{bmatrix}\boldsymbol{u} & c\boldsymbol{v}+\boldsymbol{w}\end{bmatrix}\right)=c\delta\left(\begin{bmatrix}\boldsymbol{u} & \boldsymbol{v}\end{bmatrix}\right)+\delta\left(\begin{bmatrix}\boldsymbol{u} & \boldsymbol{w}\end{bmatrix}\right), c\in\mathbb{R}\\ \\ \delta\left(\begin{bmatrix}\boldsymbol{u} & 0\end{bmatrix}\right)=0\ \text{（如果有任意一列为零向量，则整体结果为零）}\\ \\ \delta\left(\begin{bmatrix}\boldsymbol{u} & \boldsymbol{v}\end{bmatrix}\right)=-\delta\left(\begin{bmatrix}\boldsymbol{v} & \boldsymbol{u}\end{bmatrix}\right)\ \text{（交换任意两列，则整体符号改变）}\end{cases}$$

上述性质对于形如 $\mathcal{A}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ 的矩阵同样适用。证明：$\delta(\mathcal{A})\neq 0$ 的充要条件是 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$ 彼此线性无关。（在这道题中 δ 其实就是我们后期要讲的行列式的雏形，该运算基本满足所有行列式的性质。我们还能从中得出得出一个推论：矩阵中列向量（行向量）彼此线性无关的充要条件是矩阵的行列式不为零。对于形如这种涉及 n 个向量的证明，数学归纳法一直都是一个非常有效的证明路径。）



1.3 线性空间和向量、向量组之间的关系

在上一节中，我们重点学习了向量组之间的关系，也就由此引出了这样的一个话题：向量从哪里来？我们知道在平面向量的知识中，向量代表在平面直角坐标系内的一条有向线段，因此平面向量的载体便是平面直角坐标系，我们也把平面直角坐标系所构成的全部区域称为是一个**线性空间（又称作向量空间）（Vector Space)** 我们可以把线性空间理解为所有能够使得向量有意义的区域，因此线性空间也就自然而然包含了非常多的向量，我们便可以把线性空间理解为满足一定条件的全部向量所构成的集合。那么我们会想对于一个线性空间而言，究竟有哪些条件需要满足呢？

我们还以平面直角坐标系 xOy 为例，假设 $\boldsymbol{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$，我们可以利用线性组合的知识将 $\boldsymbol{v}_1$ 分解成如下的形式：

$$\boldsymbol{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

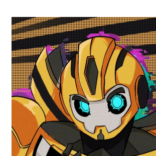
这样一来，我们发现向量 $\boldsymbol{v}_1$ 由两个向量和两个常数共同决定，如果我们设 $\boldsymbol{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$, 我们便发现 $\boldsymbol{v}_2$ 位于空间直角坐标系而非平面直角坐标系中，使得 $\boldsymbol{v}_2$ 不存在于平面直角坐标系内的原因是因为 $\boldsymbol{v}_2$ 的维数不同于平面直角坐标系中向量的维数。我们再设 $\boldsymbol{v}_3 = i\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, 此时我们发现 $\boldsymbol{v}_3$ 是复平面内的一点，意味着 $\boldsymbol{v}_3$ 是含有虚数常数项的线性组合，也不存在于平面直角坐标系 xOy 内。我们把由这些常数构成的集合统称为**数域（Number Field)**, 每一个线性空间都有与之对应的数域，通常用 $\mathbb{F}$ 表示。数域和线性空间二者缺一不可。比如平面直角坐标系中的数域为实数，复平面中的数域为复数；同样每一个线性空间也有与之对应的向量，我们把满足条件的向量组称作是该线性空间的**基底（Basis)**。我会在后面几页的篇幅中重点讨论线性空间的基底。对于数域而言，由于需要涉及高等代数的内容，所以我点到为止，并且在目前阶段，我们将统一把数域特指为全体实数 $\mathbb{R}$。

所以，一个向量是否在一个特定的线性空间内，是由向量本身以及常数项所共同决定的，我们由此提出线性空间（一般用大写字母 U, V, W 等表示）的概念：

定义 1-7
我们称 V 是定义在数域 $\mathbb{F}$ 上的线性空间，如果对于任意的 $\boldsymbol{u}, \boldsymbol{v} \in V; a, b \in \mathbb{F}$, 使得：

$$① \ (ab)\boldsymbol{v} = a(b\boldsymbol{v})$$
$$② \ (a+b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$$
$$③ \ a(\boldsymbol{v}+\boldsymbol{u}) = a\boldsymbol{v} + a\boldsymbol{u}$$

需要说明的是，定义 1-7 并非线性空间的完整定义，在本书中出于各种因素对其进行了简化处理。完整的线性空间定义有 8 条，在此不作详细说明。



我们会在本书中重点研究两到三种线性空间（这些线性空间的数域目前统一确定为 $\mathbb{R}$）：第一种是线性空间 $\mathbb{R}^n$，该线性空间代表了全部 n 维向量所构成的集合，比如 $\begin{bmatrix}1\\0\\3\end{bmatrix} \in \mathbb{R}^3$; $\begin{bmatrix}1\\0\end{bmatrix} \in \mathbb{R}^2$ 等。第二种是线性空间 $\mathbb{P}_n(x)$，即为全部关于 x 的 n 次多项式所构成的集合，比如 $x^2-3x+4 \in \mathbb{P}_2(x)$，$x^5+x^2+2 \in \mathbb{P}_5(x)$ 等。第三种是线性空间 M_{mn}，即为全部 $m \times n$ 矩阵所构成的集合，比如 $\begin{bmatrix}1 & 0 & 0\\0 & 2 & 2\end{bmatrix} \in M_{23}$ 等。读者如果感兴趣可以利用定义 1-7 中的三条性质来自行验证为什么我在上面提到的三种类型满足定义在实数域上的线性空间。

通过上面的例子，我们发现线性空间里面的元素无穷无尽，对于给定的线性空间，我们如何才能更好地研究呢？这里用到的原理和我们第一，二节的方法类似，我们知道线性空间中既然有这么多的向量，那么一定会有一些线性相关的元素存在吧？比如在 $\mathbb{P}_2(x)$ 中，我们知道 x^2+2x+1 与 $2x^2+4x+2$ 彼此线性相关。如同这样的例子不胜枚举，那么我们是否可以进一步地"压缩"线性空间里面的向量个数，使得线性空间的结构可以大幅简化呢？答案是必然的，这也就是为什么我们引入了基底的定义：

定义 1-8
设 V 是定义在数域 $\mathbb{F}$ 上的线性空间，我们称集合 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为线性空间 V 的一组**基底** ***(Basis)***，当且仅当：

① β 中的向量彼此线性无关

② 对任意的 $\boldsymbol{u} \in V$，都有 $\boldsymbol{u} \in \textbf{Span}\left(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\right)$

我们可以简单地从"不多不少"这两个角度来理解线性空间的基底，"不多"对应着定义 1-8 中的第一条，意味着没有"多余"的向量，每一个向量都起到了至关重要的作用；"不少"对应着定义 1-8 中的第二条，意味着线性空间的完整性要得到保障，使得这些向量可以没有遗漏地将每一个线性空间中的向量表示出来。在有了基底的定义之后，我们可以继续定义线性空间的维数：

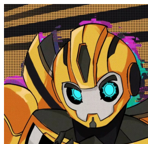
定义 1-9
设 V 是定义在数域 $\mathbb{F}$ 上的线性空间，集合 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为线性空间 V 的一组**基底** ***(Basis)***，则线性空间 V 的**维数** ***(Dimension)*** 为集合 β 中元素的个数，记作 $\dim(V) = |\beta|$。

对线性空间 V 的子集 $\mathcal{P}$ 而言，定义 $\mathcal{P}$ 的维数是其中彼此线性无关的向量的最大数目，定义 $\dim(\{\mathbf{0}\}) = 0$。

定理 1-10

$$\dim(\mathbb{R}^n) = n$$

我们将首次正式地对一个定理进行证明，在以后的章节中有价值，重要的定理我也会进行专门的证明，但读者应当自行尝试去证明每一条定理，通过自行证明定理，可以极大地加深对概念的理解和掌握。



定理 1-10 的证明:

证明. 在 $\mathbb{R}^m$ 中，我们假设向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 为 $\mathbb{R}^m$ 的一组基底。因此，根据定义我们知道向量组 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 线性无关，则由 $\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ 构成的矩阵的列秩为 n。又因为对于任意矩阵而言，其行秩等与列秩，因此该矩阵的行秩为 n。因为我们知道 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 的维数为 m，因此行秩的最大取值为 m，即在 $\mathbb{R}^m$ 中彼此线性无关的向量最大值为 m，所以 $n = m$，故其基底为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m$，因此我们称 $\dim(\mathbb{R}^m) = m$，定理得证。

■

对于定理 1-10 而言，我们还可以再得出一条推论:

定理 1-10 (a)

$$\dim(\mathbb{P}_n(x)) = n + 1$$

我们可以通过一个例子来理解定理 1-10 (a)，比如在 $\mathbb{P}_2(x)$ 中，我们可以选取 $\beta = \{1, x, x^2\}$ 作为一组基，其中由三个线性无关的多项式组成，因此我们称 $\mathbb{P}_2(x)$ 的维数为 3。再由定义 1-9，我们还能得出一条定理。该定理使得我们在判断基底的时候可以减少计算量。

定理 1-11

设 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为线性空间 V 的一个子集，若 β 线性无关，且 $\dim(V) = |\beta|$，则 β 为 V 的一组基。

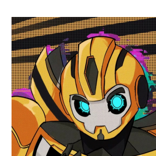
例 1.3.1

在 $\mathbb{R}^3$ 中，判断 $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}; \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}; \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$，判断 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ 是否构成 $\mathbb{R}^3$ 的一组基。

判断是否为基底，我们一定要把是否线性无关放在第一位，那么根据定义，我们只需要看线性齐次方程组 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0$ 的唯一解是否为 $c_1 = c_2 = c_3 = 0$。我们通过研究矩阵 $\left[\begin{array}{ccc|c} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \mathbf{0} \end{array}\right]$，经高斯消元后的到其行最简形矩阵为

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

显然该线性齐次方程组存在唯一平凡解 $c_1 = c_2 = c_3 = 0$，因此 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 线性无关。又因为 $|\beta| = 3 = \dim(\mathbb{R}^3)$，所以根据定理 *1-11*，我们可以直接得出 β 为 $\mathbb{R}^3$ 的一组基。



定理 1-11 的优势在于，对于线性空间 V 一个线性无关的子集 β, 在有 $\dim(V)=|\beta|$ 的情况下，为了确认是否为一组基底，我们可以通过判断线性无关从而直接确定，省去了验证线性组合，从而可以简化我们的运算量。当我们知道 β 为线性空间 V 的一组基底之后，根据性质我们也知道 V 中任意一个向量都是 β 的一个线性组合。我们此时便想知道这个线性组合究竟有着什么样的性质？在此情况下一个向量仅仅有一种线性组合吗？有没有无穷多组满足条件的线性组合？如果我们选择 V 的另外一组基底 γ，这个线性组合又有着什么样的性质？我们通过研究几条定理，来一一回答上述这些问题。

定理 1-12

设 $\beta=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$ 为线性空间 V 的一组基底，那么对于任意的 $\boldsymbol{u}\in V$，有且仅有唯一的一组解 $c_1,c_2,\cdots,c_n$，使得

$$\boldsymbol{u}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_n\boldsymbol{v}_n$$

我们称实数对 $(c_1,c_2,\cdots,c_n)$ 为向量 $\boldsymbol{u}$ 在基底 β 下的**唯一分解坐标 *(Unique Representation)***，为防止混淆和严格区分基底，我们通常写作

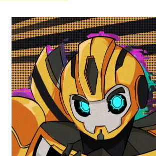
$$[\boldsymbol{u}]_\beta=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}_\beta$$

定理 1-12 的证明：

证明. 给定基底 β, 为了验证 $(c_1,c_2,\cdots,c_n)$ 是否为唯一的分解形式, 我们只需要讨论线性方程组 $\boldsymbol{u}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_n\boldsymbol{v}_n$ 的解的个数。我们不妨设 $\boldsymbol{u}=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}_\beta$，这样一来运用求解线性方程组的知识，写出矩阵形式 $\left[\mathcal{A}\;\middle|\;\boldsymbol{b}\right]$ 我们只需化简形如下式的矩阵：

$$\left[\begin{array}{cccc|c}| & | & & | & u_1\\\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n & \vdots\\| & | & & | & u_n\end{array}\right]$$

我们知道，列向量组 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 线性无关，因此矩阵 $\mathcal{A}$ 满秩，那么根据定义，其行最简型矩阵一定为 $\left[\begin{array}{ccccc|c}1 & 0 & 0 & \cdots & 0 & \lambda_1\\0 & 1 & 0 & \cdots & 0 & \lambda_2\\ & & & \vdots & & \vdots\\0 & 0 & 0 & \cdots & 1 & \lambda_n\end{array}\right]$，所以此时原方程组有且仅有一解。 ■



定理 1-12 说明了两点：第一点即为在给定的基底中，一个向量在该基底里面有且仅有一种分解方式。第二点为向量的坐标和基底的选取密切相关。在不同的基底下，向量的坐标也会发生变化。所以我们从现在开始要格外注意基底的强调。在本书中如果没有特殊说明或者要求的话，所有向量的坐标均默认为在标准基底下的分解, 我们定义 $\mathbb{R}^n$ 的标准基底为：

$$\beta = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}; \cdots; \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

通常我们也使用 $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$ 来代指标准基底，其中 $\boldsymbol{e}_i$ 里面的第 i 个元素为 1，其余元素均为 0。我们之前见到的形如 $\boldsymbol{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ 的向量其实就是在标准基底 $\mathbb{R}^3$ 之下的唯一分解的坐标，写作

$$[\boldsymbol{v}]_E = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}_E = 2\boldsymbol{e}_1 + 3\boldsymbol{e}_2 + 1\boldsymbol{e}_3$$

在标准基底中，向量右下角的角标可以略去，但如果是向量在其他基底下的分解形式，则必须要加上角标。我们接下来研究一下向量在不同的基底之间是如何转化的。

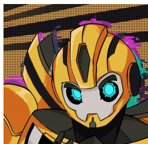
定理 1-13

设 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为线性空间 V 的一组基底, 设向量 $\boldsymbol{u}$ 在基底 β 下的分解坐标为 $[\boldsymbol{v}]_\beta = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_\beta$，那么向量 $\boldsymbol{u}$ 在标准基底 E 下的分解坐标即为 $[c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n]_E$

我们来看另一组 $\mathbb{R}^3$ 的基底 $\gamma = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_E, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_E, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_E \right)$，此时如果我们定义 $[\boldsymbol{v}]_\gamma = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}_\gamma$，那么根据定义 1-12，我们知道向量的坐标其实就是基底向量构成的线性组合。我们于是可以得出：

$$[\boldsymbol{v}]_\gamma = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}_\gamma = 2\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_E + 3\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_E + 1\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_E = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}_E \text{（标准基底）}$$

此时，我们应该格外注意基底与向量坐标之间的顺序关系，在基底中，从左至右第 i 个向量应该对应线性组合中的第 i 项。在本书中，我们默认所有的基底的顺序都是固定的，因此我们便可以用 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 来



表示基底，其中 $\boldsymbol{v}_i$ 即为基底里面从左至右的第 i 个向量，顺序不会发生变化。我们称这样的基底为**秩序基底 (Ordered Basis)**。在本书中，所有的基底均为秩序基底。

例 1.3.2

在 $\mathbb{R}^3$ 中，向量 $\boldsymbol{v}$ 在标准基底 E 下分解的坐标为 $[\boldsymbol{v}]_E = \begin{bmatrix}0\\0\\3\end{bmatrix}_E$。设 $\gamma = \left(\begin{bmatrix}-1\\-1\\1\end{bmatrix}_E, \begin{bmatrix}1\\0\\1\end{bmatrix}_E, \begin{bmatrix}0\\1\\1\end{bmatrix}_E\right)$ 为 $\mathbb{R}^3$ 的另一组基底，据此求 $[\boldsymbol{v}]_\gamma$。

这道题其实就是定理 1-13 中公式的逆运算，我们可以假设 $[\boldsymbol{v}]_\gamma = \begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}_\gamma$ *，那么我们就有这样一个线性方程组：*

$$c_1\begin{bmatrix}-1\\-1\\1\end{bmatrix}_E + c_2\begin{bmatrix}1\\0\\1\end{bmatrix}_E + c_3\begin{bmatrix}0\\1\\1\end{bmatrix}_E = \begin{bmatrix}0\\0\\3\end{bmatrix}_E$$

随后，我们只需要把工作交给我们的老朋友——求解线性方程组和高斯消元了。我们可以写出这样的矩阵：

$$\left[\begin{array}{ccc|c}-1 & 1 & 0 & 0\\-1 & 0 & 1 & 0\\1 & 1 & 1 & 3\end{array}\right] \xrightarrow{\text{经过高斯消元}} \left[\begin{array}{ccc|c}1 & 0 & 0 & 1\\0 & 1 & 0 & 1\\0 & 0 & 1 & 1\end{array}\right]$$

由此我们得出， $c_1 = c_2 = c_3 = 1$*，则根据定义我们知道* $[\boldsymbol{v}]_\gamma = \begin{bmatrix}1\\1\\1\end{bmatrix}_\gamma$。

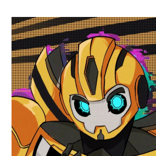
我们再来看一个和多项式有关的例子：

例 1.3.3

已知 $\mathbb{P}_2(x)$ 的标准基底为 $E = \{1, x, x^2\}$, 设线性空间 W 为满足 $f(1) = 0, f'(1) = 0$ 的全体二次多项式构成的集合，求出 W 的一组基。

我们首先设 $f(x) = ax^2+bx+c$, *根据* W *所给的限制条件，我们可以列出这样的一个方程组：* $\begin{cases}a+b+c=0\\2a+b=0\end{cases}$。

求解这个方程组并不困难，我们最终可以解得 $\begin{bmatrix}a\\b\\c\end{bmatrix} = s\begin{bmatrix}1\\-2\\1\end{bmatrix} : s \in \mathbb{R}$。*由此，我们可以取* $(1, -2, 1)_E$ *作为* W *的一组基。即为形如* $f(x) = k(1-2x+x^2) : k \in \mathbb{R}$ *的全体多项式。*



我们现在已经了解向量在标准基底和其它基底之间的转化，那么对于两个非标准基底，它们之间的转化该如何求解呢？

例 1.3.4

设 $\beta = \left(\begin{bmatrix}1\\1\\0\end{bmatrix}_E, \begin{bmatrix}1\\0\\1\end{bmatrix}_E, \begin{bmatrix}0\\1\\1\end{bmatrix}_E\right)$, $\gamma = \left(\begin{bmatrix}1\\2\\0\end{bmatrix}_E, \begin{bmatrix}-1\\0\\1\end{bmatrix}_E, \begin{bmatrix}0\\1\\-2\end{bmatrix}_E\right)$ 为 $\mathbb{R}^3$ 的两组基底，若 $[\boldsymbol{v}]_\beta = \begin{bmatrix}1\\2\\-1\end{bmatrix}$，求 $[\boldsymbol{v}]_\gamma$。

此时，我们需要标准基底 E 作为一个“过渡态”，先求出 $[\boldsymbol{v}]_E$，然后在去求 $[\boldsymbol{v}]_\gamma$。根据定义我们知道

$$[\boldsymbol{v}]_E = 1\begin{bmatrix}1\\1\\0\end{bmatrix}_E + 2\begin{bmatrix}1\\0\\1\end{bmatrix}_E + (-1)\begin{bmatrix}0\\1\\1\end{bmatrix}_E = \begin{bmatrix}3\\0\\1\end{bmatrix}_E$$

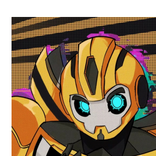
随后我们再设 $[\boldsymbol{v}]_\gamma = \begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}$，因此可以列出如下的线性方程组：

$$\begin{bmatrix}3\\0\\1\end{bmatrix}_E = c_1\begin{bmatrix}1\\2\\0\end{bmatrix}_E + c_2\begin{bmatrix}-1\\0\\1\end{bmatrix}_E + c_3\begin{bmatrix}0\\1\\-2\end{bmatrix}_E$$

不难解得 $\begin{cases} c_1 = \frac{4}{5} \\ c_2 = -\frac{11}{5} \\ c_3 = -\frac{8}{5} \end{cases}$，由此我们得到 $[\boldsymbol{v}]_\gamma = \begin{bmatrix}\frac{4}{5}\\-\frac{11}{5}\\-\frac{8}{5}\end{bmatrix}_\gamma$。

我们不难发现，向量在不同基底下的坐标是不同的，但我们要知道向量本身并没有“坐标”这一属性，我们定义坐标就好比物理学定义参考系一样，通过选取特定的基底，从而使得向量在这个基底下的坐标更加简洁和方便运算。我也说过，在没有特殊强调的时候，所有我们所见的坐标均为向量在标准基底 E 下的分解，我也会用不同的下角标来额外强调。要注意的是在进行不同基底之间的运算时，我们一定要把向量转化到一个共同的基底下进行计算，就好比物理学里面我们要保证一个系统内所有物体的参考系是一致的。

在后续的章节中，我们将把基底与基底之间的相互转化看作线性变换，届时我们将从矩阵以及函数的角度重新审视基底与基底之间的变换，这种变换也被我们称作是矩阵的对角化，就是读者可能已经有所听闻的 $\mathcal{S} = \mathcal{P}^{-1}\mathcal{A}\mathcal{P}$ 形式。在现阶段，读者只需要掌握对于给定的向量，求出其在不同基底下面的坐标即可。基底的概念至关重要，希望读者能够对本章进行反复地阅读，确保能够完全理解本章所提到的概念。



1.3 练习

1. 在本章中我说过向量的坐标会随着其基底的选择而发生变化，但是零向量除外。在 $\mathbb{R}^n$ 中，对于任意基底 β 而言，都有 $[\mathbf{0}]_\beta = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$。为什么？

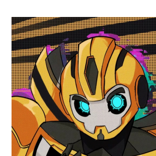
2. 在线性空间 $\mathbb{P}_3(x)$ 中，记集合 $V = \{f(x) \in \mathbb{P}_3(x) : 2f(1) = f'(1), f(1) + f(2) = 0\}$，（其中集合 V 也为一个线性空间，我们称其为 $\mathbb{P}_3(x)$ 的一个子空间。我们会在下一节详细地讨论子空间）试求出 V 的一组基底。

3. 在线性空间 $\mathbb{R}^4$ 中，设在标准基底下分解的向量 $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$。构造出 $\mathbb{R}^4$ 的一组基 β, 使得 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \beta$。

4. 设 $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ 和 $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\}$ 为线性空间 $\mathbb{R}^3$ 的两组基底，那么 $\{\boldsymbol{x}_1 + \boldsymbol{y}_1, \boldsymbol{x}_2 + \boldsymbol{y}_2, \boldsymbol{x}_3 + \boldsymbol{y}_3\}$ 能否也为 $\mathbb{R}^3$ 的一组基底？

5. 在线性空间 $\mathbb{P}_n(x)$ 中，我们设 $\beta = \{x^n; x^n+x^{n-1}; x^n+x^{n-1}+x^{n-2}; \cdots; x^n+x^{n-1}+\cdots+x+1\}$ 为 $\mathbb{P}_n(x)$ 的一组基底（读者可以尝试自行证明 β 为一组基底）。我们设多项式 $f(x) = x^n+2x^{n-1}+3x^{n-2}+\cdots+nx+(n+1)$，据此求出 $[f(x)]_\beta$。（我们设 $\{1, x, x^2, \cdots, x^n\}$ 为 $\mathbb{P}_n(x)$ 的标准基底）

6. 在线性空间 $\mathbb{P}_n(x)$ 中，设 $f(x) \in \mathbb{P}_n(x)$, 设集合 $\beta = \{f^{(0)}(x), xf^{(1)}(x), \cdots, x^n f^{(n)}(x)\}$, 其中 $f^{(k)}(x)$ 代表多项式 $f(x)$ 的 k 阶导数。当 $f(x)$ 满足什么条件时，β 构成 $\mathbb{P}_n(x)$ 的一组基底？



1.4　线性空间之间的关系

在上一节中，我们主要探讨了线性空间与向量的关系。而在本节的内容里面我们将会探讨线性空间之间的关系，我之前说过线性空间就好比满足一定条件的向量所构成的集合，那么线性空间之间的关系和运算也可以类比集合之间的关系和运算，比如在集合中我们有子集这一说，在线性空间里面我们也有子空间的概念。我们知道在集合里面，如果集合 A 的元素个数为 n，那么我们知道 A 的子集个数为 2^n。但是对于线性空间而言，由于其本身就包含了非常多的向量 (绝大多数情况下，包含了无穷多个向量)，因此通过集合的方法去定义线性空间的子空间就显得不太合适，那我们可能想到用线性空间的基底向量去定义子空间。但我们也知道对于一个线性空间而言，基底的选择有无穷多种，并且在不同基底下向量的分解坐标也不相同，所以也不能通过基底来定义子空间。其实，构成子空间的条件要比构成子集的条件要苛刻的多，子空间一定程度上继承了线性空间的一些基本性质，**线性空间的任何子空间自身也是一个线性空间**。我们把这些性质罗列出来，引出本节的第一个重要定义：

定义 1-10
设 V 为定义在数域 $\mathbb{F}$ 上的线性空间，我们定义 W 为线性空间 V 的**子空间 *(Subspace)***，记作 $W \subset V$，如果对于任意的 $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W, \lambda \in \mathbb{F}$，满足：

① $\boldsymbol{w}_1, \boldsymbol{w}_2 \in V$

② $\boldsymbol{w}_1 + \boldsymbol{w}_2 \in W$

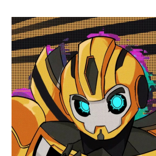
③ $\lambda w_1, \lambda w_2 \in W$

因此我们发现，子空间有着较为严格的限制条件，首先是确保子空间 W 中的向量同样也在 V 中，另外在定义的第二条性质种，我们也称子空间**对于向量加法运算封闭 (Closed Under Vector Addition)**；在定义的第三条性质中，我们也称子空间**对于系数乘法运算封闭 (Closed Under Scalar Multiplication)**。

定理 1-14
设 V 为定义在数域 $\mathbb{F}$ 上的线性空间，W 为线性空间 V 的子空间，则 $\mathbf{0} \in W, \mathbf{0} \in V$。

零向量的标准定义是 $\mathbf{0}\boldsymbol{u} = 0 : \forall \boldsymbol{u} \in V$，并且零向量与基底的选取无关。定理 1-14 的证明并不难，我们可以直接取 $\lambda = 0$，从而根据定义我们有 $\lambda \boldsymbol{v} = \mathbf{0}$。该定理 1-14 同时也是判断是否为线性空间以及子空间的一个最简单直接的方法。

例 1.4.1
在线性空间 $\mathbb{P}_2(x)$ 中，$W = \{f(x) \in \mathbb{P}_2(x) : f(1) = 1\}$ 是否为 $\mathbb{P}_2(x)$ 的一个子空间？
我们先验证零向量，在本题中零向量代表 $f(x) = 0$，即 $[f(x)]_E = (0,0,0)_E$，我们发现 $f(1) = 0 \neq 1$，则 $\mathbf{0} \notin W$，所以 W 不构成子空间。



随后，我们回到矩阵本身，设 $\mathcal{A}$ 为 $m \times n$ 矩阵，我们知道可以用行向量和列向量对矩阵 $\mathcal{A}$ 进行表示，因此我们不妨设由列向量表示的矩阵 $\mathcal{A}$ 为 $\mathcal{A}_c = \begin{bmatrix} | & | & & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \cdots & \boldsymbol{v}_n \\ | & | & & & | \end{bmatrix}$，由行向量表示的矩阵 $\mathcal{A}$ 为 $\mathcal{A}_r = \begin{bmatrix} -- & \boldsymbol{u}_1 & -- \\ -- & \boldsymbol{u}_2 & -- \\ & \vdots & \\ -- & \boldsymbol{u}_m & -- \end{bmatrix}$。我们把 $m \times n$ 矩阵 $\mathcal{A}$ 中由全部列向量构成的线性组合记作 $\textbf{Span}(\mathcal{A}_c)$, 由全部行向量构成的线性组合记作 $\textbf{Span}(\mathcal{A}_r)$, 我们便不难发现其中列向量的维数均为 m，而行向量的维数均为 n, 因此我们知道由列向量所构成的任何线性组合的维数均为 m; 由行向量所构成的任何线性组合的维数均为 n, 所以我们得到 $\textbf{Span}(\mathcal{A}_c) \subset \mathbb{R}^m$, $\textbf{Span}(\mathcal{A}_r) \subset \mathbb{R}^n$。这样一来，我们就建立了子空间和矩阵之间的联系。

定义 1-11

在 $m \times n$ 矩阵 $\mathcal{A}$ 中，我们定义矩阵的**列空间** ***(Column Space)***，记作 $C_{\mathcal{A}}$。同时定义矩阵的**行空间**（***Row Space***），记作 $R_{\mathcal{A}}$。容易得到 $C_{\mathcal{A}} \subset \mathbb{R}^m; R_{\mathcal{A}} \subset \mathbb{R}^n$。我们一般还可以将列空间称作**像空间** ***(Image Space)***，记作 $\textbf{Im}(\mathcal{A})$。

定义 1-12

在 $m \times n$ 矩阵 $\mathcal{A}$ 中，我们定义矩阵 $\mathcal{A}$ 的**核空间** ***(Kernel Space)*** 为满足条件的所有向量 $\boldsymbol{x}$，使得 $\mathcal{A}\boldsymbol{x} = \boldsymbol{0}$。我们将矩阵 $\mathcal{A}$ 的核空间（又称零空间（*Null Space*））记作 $\textbf{Ker}(\mathcal{A})$。

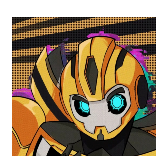
根据矩阵和向量乘法的运算法则，我们知道对于 $m \times n$ 矩阵 $\mathcal{A}$ 而言，如果 $\mathcal{A}\boldsymbol{x}$ 有意义，那么 $\boldsymbol{x}$ 的维数一定为 n, 同时 $\mathcal{A}\boldsymbol{x}$ 的维数为 m。并且，由于我们知道 $\mathcal{A}\boldsymbol{x}$ 实际上就是由矩阵中列向量所构成的线性组合，因此 $\mathcal{A}\boldsymbol{x} \in \textbf{Span}(\mathcal{A}_c) = \textbf{Im}(\mathcal{A})$。我们也知道若 $\mathcal{A}\boldsymbol{x} = 0$, 则 $\boldsymbol{x} \in \mathbb{R}^n; \boldsymbol{x} \in \textbf{Ker}(\mathcal{A})$, 那么此时 $\mathbb{R}^n$ 和 $\textbf{Ker}(\mathcal{A})$ 之间有什么关系呢?

定理 1-15

对任意的 $m \times n$ 实矩阵 $\mathcal{A}$ 而言，$\textbf{Ker}(\mathcal{A}) \subset \mathbb{R}^n$

定理 1-15 的证明

证明. 我们设 $\boldsymbol{x}, \boldsymbol{y} \in \ker(\mathcal{A}), \lambda \in \mathbb{F}$, 根据定义我们知道 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ 且 $\mathcal{A}\boldsymbol{x} = \boldsymbol{0}; \mathcal{A}\boldsymbol{y} = \boldsymbol{0}$, 那么 $\mathcal{A}\boldsymbol{x} + \mathcal{A}\boldsymbol{y} = \mathcal{A}(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{0}$, 则 $\boldsymbol{x} + \boldsymbol{y} \in \textbf{Ker}(\mathcal{A})$。同时，$\mathcal{A}(\lambda\boldsymbol{x}) = \lambda\mathcal{A}\boldsymbol{x} = \boldsymbol{0}$, 则 $\lambda\boldsymbol{x} \in \textbf{Ker}(\mathcal{A})$。因此 $\textbf{Ker}(\mathcal{A}) \subset \mathbb{R}^n$。 ■



例 1.4.2
设矩阵 $\mathcal{A}=\begin{bmatrix}1&0&2&-1\\2&1&0&1\\-1&2&1&0\end{bmatrix}$, 据此分别求出 $\mathbf{Ker}(\mathcal{A}),\mathbf{Im}(\mathcal{A})$ 的一组基底。

我们先求出 $\mathbf{Im}(\mathcal{A})$ 的一组基底:
首先根据定义, 我们知道 $\mathbf{Im}(\mathcal{A})=C_{\mathcal{A}}=\mathbf{Span}\left(\begin{bmatrix}1\\2\\-1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix},\begin{bmatrix}2\\0\\1\end{bmatrix},\begin{bmatrix}-1\\1\\0\end{bmatrix}\right)$, 我们因此得出 $\mathbf{Im}(\mathcal{A})$ 为 4 个向量的线性组合。由于我们知道 $\mathbf{Im}(\mathcal{A})\subset\mathbb{R}^3$, 因此构成 $\mathbf{Im}(\mathcal{A})$ 的基底中向量个数的最大值为 3, 所以 $\mathcal{A}$ 的列向量彼此线性无关。我们因此需要删去 4 个列向量中多余的向量。我们此时可以从第一个向量开始依次选择, 首先 $\begin{bmatrix}1\\2\\-1\end{bmatrix}$ 自身线性无关, 随后我们考虑集合 $\beta=\left(\begin{bmatrix}1\\2\\-1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix}\right)$, 不难验证该集合同样线性无关, 因此我们载选取集合 $\gamma=\left(\begin{bmatrix}1\\2\\-1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix},\begin{bmatrix}2\\0\\1\end{bmatrix}\right)$, 通过求解线性方程组的知识我们也不难求出该集合彼此线性无关。此时线性无关向量的个数已经达到了其基底中向量的最大个数, 所以我们据此构造出的集合 γ 便是 $\mathbf{Im}(\mathcal{A})$ 的一组基底。

我们随后求解 $\mathbf{Ker}(\mathcal{A})$ 的一组基底:
此时, 不难发现我们只需要求解线性方程组 $\mathcal{A}\boldsymbol{x}=\mathbf{0}$, 因此利用我们已经掌握的很熟练的线性方程组的相关知识, 不难写出形如下式的矩阵:

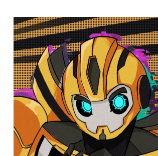
$$\left[\begin{array}{cccc|c}1&0&2&-1&0\\2&1&0&1&0\\-1&2&1&0&0\end{array}\right]$$

对该矩阵进行高斯消元, 不难得到其行最简形式为

$$\left[\begin{array}{cccc|c}1&0&0&\frac{3}{11}&0\\0&1&0&\frac{5}{11}&0\\0&0&1&-\frac{7}{11}&0\end{array}\right]$$

根据求解线性方程组的知识, 我们此时有 3 个前导变量, 1 个滞后变量, 我们选取第四列对应的未知数作为参数 s, 得到其解集为 $s\begin{bmatrix}-\frac{3}{11}\\-\frac{5}{11}\\\frac{7}{11}\\1\end{bmatrix}:s\in\mathbb{R}$, 那么 $\mathbf{Ker}(\mathcal{A})$ 的一组基底即为 $\begin{bmatrix}-\frac{3}{11}\\-\frac{5}{11}\\\frac{7}{11}\\1\end{bmatrix}$。

通过这道题, 我们有什么发现? 我们知道了对于一个 $m\times n$ 的矩阵 $\mathcal{A}$ 而言, $\dim(\mathbf{Im}(\mathcal{A}))$ 即为矩阵中彼此线性无关的列向量的最大数量, 这一就意味着根据定义, $\dim(\mathbf{Im}(\mathcal{A}))=\mathbf{rank}(\mathcal{A})$。在第一节里面, 我定义矩阵的秩等于其行最简形矩阵中"前导变量"的数量, 因此像空间的维数也就等于前导变量的数量。我们都知道,



在行最简形矩阵中，一共只有前导变量和滞后变量两种，其数目之和等于矩阵的列数 n，而从我们对刚才题目的求解看出，矩阵 $\mathcal{A}$ 的零空间又恰好是这些滞后变量的线性组合，这些向量彼此线性无关。也就是说滞后变量的数量也恰好是零空间的维数。至此，我们得出了一个线性代数中重要的定理之一——**秩零定理 (Rank - Nullity Theorem)**。

定理 1-16 （秩零定理）
对任意的 $m \times n$ 矩阵 $\mathcal{A}$ 而言，$\dim(\mathbf{Ker}(\mathcal{A})) + \dim(\mathbf{Im}(\mathcal{A})) = n$

定理 1-16 的证明

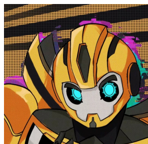
证明. 参见上文。另外我在上文中提到“矩阵 $\mathcal{A}$ 的零空间又恰好是这些滞后变量的线性组合，这些向量彼此线性无关”，这一点我并未详细说明，我们能否用线性组合的性质来印证这一观点？有兴趣的读者不妨尝试。 ■

例 1.4.3
设 $m \times n$ 矩阵 $\mathcal{A}$ 满足 $\mathbf{Ker}(\mathcal{A}) = \mathbf{Im}(\mathcal{A})$, 那么 n 的取值能否为 114514 ？能否为 191981 ？

我们此时注意到 n 的数目很大，并且 m 不确定，因此不太可能去列线性方程组从而求解。因此我们直接使用秩零定理：由题意得 $\dim(\mathbf{Ker}(\mathcal{A})) = \dim(\mathbf{Im}(\mathcal{A}))$, 那么 $\dim(\mathbf{Ker}(\mathcal{A})) + \dim(\mathbf{Im}(\mathcal{A})) = 2\dim(\mathbf{Ker}(\mathcal{A})) = 2\dim(\mathbf{Im}(\mathcal{A})) = n$, 由于线性空间的维数只能为整数，所以 n 必为偶数，因此 n 可以取 114514，不能取 191981。至于怎么去实打实地找到满足 $\mathbf{Ker}(\mathcal{A}) = \mathbf{Im}(\mathcal{A})$ 的矩阵 $\mathcal{A}$, 学有余力的读者可以自行尝试，为了简化运算我们只用考虑 $n = 2$ 的情况。

定义 1-13
对于 $m \times n$ 矩阵 $\mathcal{A}$ 而言，我们定义 $\mathcal{A}$ 的**转置矩阵**（***Transpose Matrix***）为一个 $n \times m$ 矩阵，通常记作 $\mathcal{A}^T$, 使得 $(\mathcal{A})_{ij} = (\mathcal{A}^T)_{ji}$

有了转置矩阵的概念之后，我们就再一次地建立起了列向量和行向量之间的联系。根据矩阵转置的定义，如果我们设 $\mathcal{A} = \begin{bmatrix} | & | & & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \cdots & \boldsymbol{v}_n \\ | & | & & & | \end{bmatrix}$, 那么 $\mathcal{A}^T = \begin{bmatrix} -- & \boldsymbol{v}_1 & -- \\ -- & \boldsymbol{v}_2 & -- \\ & \vdots & \\ -- & \boldsymbol{v}_n & -- \end{bmatrix}$, 根据行秩和列秩相等这一定理，我们也知道矩阵的秩等于其转置矩阵的秩。此处读者可以自行思考，矩阵和该矩阵的转置矩阵的秩零定理中，有哪些相关量?



例 1.4.4

设 $m \times n$ 矩阵 Q 的列向量分别为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$。若 Q 满秩，证明集合 $\beta = \{Q^T\boldsymbol{v}_1, Q^T\boldsymbol{v}_2, \cdots, Q^T\boldsymbol{v}_n\}$ 为 $\mathbb{R}^n$ 的一组基底。

证明. 我们知道向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 的维数为 m，此时 Q^T 为 $n \times m$ 矩阵，因此乘法 $Q^T\boldsymbol{v}_i$ 有意义。通过观察，我们发现 $|\beta| = n = \dim(\mathbb{R}^n)$，由定理 *1-11* 可知，欲证明 β 为 $\mathbb{R}^n$ 的一组基底，我们只需证明 β 中的元素彼此线性无关。我们因此假设存在这样的线性齐次方程：

$$c_1Q^T\boldsymbol{v}_1 + c_2Q^T\boldsymbol{v}_2 + \cdots + c_nQ^T\boldsymbol{v}_n = 0 \ *$$

由于 $c_1, c_2, \cdots, c_n$ 为常数，故我们可以写作

$$Q^T(c_1\boldsymbol{v}_1) + Q^T(c_2\boldsymbol{v}_2) + \cdots + Q^T(c_n\boldsymbol{v}_n) = 0$$

即

$$Q^T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n) = 0$$

由于 Q 满秩，因此根据行秩定理，$\mathbf{rank}(Q) = \dim(\mathbf{Im}(Q)) = n, \dim(\mathbf{Ker}(Q)) = 0$，即 $\mathbf{Ker}(Q) = \{\mathbf{0}\}$，$\dim(\mathbf{Ker}(Q^T)) = m - \mathbf{rank}(Q^T) = m - \mathbf{rank}(Q) = m - n$。又因为 Q 满秩，则 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 彼此线性无关，再由定理 *1-11* 可知，$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为 $\mathbb{R}^n$ 的一组基，则 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 的维数均为 n，则 $m = n$，所以 $\dim(\mathbf{Ker}(Q^T)) = 0$，则 $\mathbf{Ker}(Q^T) = \{\mathbf{0}\}$。因此线性齐次方程 $Q^T\boldsymbol{x} = \mathbf{0}$ 有唯一平凡解 $\boldsymbol{x} = 0$，在本题中即为 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = 0$。又因为这些向量彼此线性无关，则 $c_1 = c_2 = \cdots = c_n = 0$，带入 *(*)* 式中，即 β 线性无关，因此 β 为 $\mathbb{R}^n$ 的一组基底。 ■

在本节的后半部分，我们将会研究不同线性空间之间的关系，准确来说我们会研究在给定的线性空间中，其子空间之间的联系。我们都知道对于集合而言，我们可以定义集合之间的交集与并集，那么对于线性空间而言，我们同样可以定义线性空间的交集与并集。

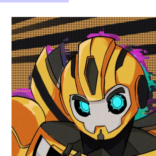
定义 1-14

设 U, V 均为某线性空间 *(*通常为 $\mathbb{R}^n$*)* 的子空间，我们定义线性空间 U, V 的交集为 $U \cap V$; 并集为 $U \cup V$。设 $\boldsymbol{w} \in U \cap V$，则 $\boldsymbol{w} \in U$ 且 $\boldsymbol{w} \in V$；若 $\boldsymbol{w} \in U \cup V$，则 $\boldsymbol{w} \in U$ 或 $\boldsymbol{w} \in V$。

此时，我们自然而然会想 $U \cap V; U \cup V$ 是否也为线性空间呢？在往下阅读我所给出的定理之前，读者不妨自行思考片刻，如果可能，尝试自己写出 $\mathbb{R}^2$ 或 $\mathbb{R}^3$ 的两个子空间，然后通过计算来判断。

定理 1-17

设 U, V 均为某线性空间 *(*通常为 $\mathbb{R}^n$*)* 的子空间，那么对任意的 U, V，可知 $U \cap V$ 为线性空间；$U \cup V$ 为线性空间的充要条件为 $U \subset V$ 或 $V \subset U$。



定理 1-17 的证明

证明. 我们先证对任意的 U, V, 可知 $U \cap V$ 为线性空间。设 $\boldsymbol{x}, \boldsymbol{y} \in U \cap V, c \in \mathbb{F}$, 则 $\boldsymbol{x}+\boldsymbol{y} \in U$ 且 $\boldsymbol{x}+\boldsymbol{y} \in V$, 即 $\boldsymbol{x}+\boldsymbol{y} \in U \cap V$; $c\boldsymbol{v} \in U$ 且 $c\boldsymbol{v} \in V$, 即 $c\boldsymbol{x} \in U \cap V$, 因此 $U \cap V$ 为线性空间。

随后我们再证 $U \cup V$ 为线性空间的充要条件为 $U \subset V$ 或 $V \subset U$, 此时由于涉及到充要条件的证明, 我们需要从前后两个方向进行证明。即:

$(\Longrightarrow)$ 假设 $U \cup V$ 为线性空间, 则需证明 $U \subset V$ 或 $V \subset U$。此时我们可以使用反证法, 即假设 $U \subsetneq V$ 且 $V \subsetneq U$, 此时即存在 $\boldsymbol{u} \in U, \boldsymbol{u} \notin V$ 且 $\boldsymbol{v} \in V, \boldsymbol{v} \notin U$。由于 $U \cup V$ 为线性空间, 且 $\boldsymbol{u} \in U \cup V; \boldsymbol{v} \in U \cup V$, 则 $\boldsymbol{u}+\boldsymbol{v} \in U \cup V$。此时我们考虑 $(\boldsymbol{u}+\boldsymbol{v})-\boldsymbol{v}$, 根据线性空间的性质我们知道 $(\boldsymbol{u}+\boldsymbol{v})-\boldsymbol{v} \in U \cup V$ 即 $\boldsymbol{u} \in U \cup V$, 矛盾。同理考虑 $(\boldsymbol{u}+\boldsymbol{v})-\boldsymbol{u}$, 根据线性空间的性质我们知道 $(\boldsymbol{u}+\boldsymbol{v})-\boldsymbol{u} \in U \cup V$ 即 $\boldsymbol{v} \in U \cup V$, 矛盾。通过反证法, 我们即证明了若 $U \cup V$ 为线性空间, 则 $U \subset V$ 或 $V \subset U$。

$(\Longleftarrow)$ 假设 $U \subset V$ 或 $V \subset U$, 则需证明 $U \cup V$ 为线性空间。由于 U, V 本身即为线性空间, 当 $U \subset V$ 时 $U \cup V = V$; 当 $V \subset U$ 时, $U \subset V = U$, 所以此时 $U \cup V$ 即为线性空间。

因此, $U \cup V$ 为线性空间的充要条件为 $U \subset V$ 或 $V \subset U$ 得证。

■

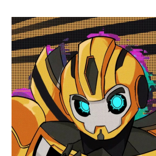
那么对于线性空间而言, 有没有类似于实数之间的加法, 乘法运算呢? 答案是有的, 只不过线性空间之间的加法和乘法运算法则与实数有所不同, 我们先从定义较为简单的乘法运算说起。线性空间 U, V 的乘法运算也被称作是线性空间的**笛卡尔积 (Cartesian Product)**, 通常记作 $U \times V$, 其定义为:

定义 1-15
设 U, V 为线性空间, 我们定义线性空间 U, V 的笛卡尔积 $U \times V$ 为:

$$U \times V = \{(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u} \in U, \boldsymbol{v} \in V\}$$

不难发现, 笛卡尔积的结果其实是一个 "坐标", 其中坐标里的元素分别对应两个线性空间里面的元素, 这也是为什么在平面直角坐标系 xOy 中任意一点 x, y 有横坐标 x 和纵坐标 y。我们称平面直角坐标系为 $\mathbb{R}^2$, 本质上就是笛卡尔积 $\mathbb{R} \times \mathbb{R}$ 的结果。同理我们还知道 $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, 也就是为什么空间直角坐标系中的点分别对应了 x, y, z 三个坐标。那么我们也不难发现, 对于两个线性空间而言, 其笛卡尔积同样为线性空间。

定理 1-18
设 U, V 均为线性空间, 则 $U \times V$ 同样为线性空间。



有兴趣的读者不妨自行证明定理 1-18。我们同时也可以定义线性空间之间的加法运算：

定义 1-16

设 U,V 为线性空间，我们定义线性空间 U,V 的和 $U+V$ 为

$$U+V=\{\boldsymbol{u}+\boldsymbol{v}:\boldsymbol{u}\in U,\boldsymbol{v}\in V\}$$

也就是说，$U+V$ 中的任何一个向量均为 U 中的一个向量 $\boldsymbol{u}$ 与 V 中的一个向量 $\boldsymbol{v}$ 的加和。在平面直角坐标系内，如果我们把 U 当作 x 轴，即 $U=\textbf{Span}\left(\begin{bmatrix}1\\0\end{bmatrix}\right)$，如果我们把 V 当作 y 轴，即 $V=\textbf{Span}\left(\begin{bmatrix}0\\1\end{bmatrix}\right)$，那么 $U+V$ 即为 $\mathbb{R}^2$。或者我们也可以令 $U=\mathbb{R}^2,V=\{\boldsymbol{0}\}$，也能够产生同样的效果。我们由此发现，对于一个向量空间而言，其可以表示为数个子空间的加和。我们其中会重点研究一种特殊的加和，称作是**直和 (Direct Sum)**。我们一起来看直和是如何定义的：

定义 1-17

设 U,V 为线性空间 W 的子空间，我们称 W 为其子空间 U,V 的直和，记作 $W=U\bigoplus V$，如果 U,V 满足以下性质：

$$① \ W=U+V \quad ② \ U\cap V=\{\boldsymbol{0}\}$$

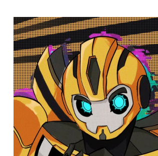
我们不难发现，如果 $W=U\bigoplus V$，那么 U,V 的交集有且只有零向量，这一就意味着这两个线性空间彼此之间没有太多的关联，我们也可以称 U 和 V 线性无关。此时我们不妨在回到基底的概念，此时我们又引出了一条重要的定理：

定理 1-19

设线性空间 U,V,W 满足 $W=U\bigoplus V$，设 β_1 为 U 的一组基底，β_2 为 V 的一组基底，则 $\beta_1\cup\beta_2$ 为 W 的一组基底。

定理 1-19 的证明

证明. 由直和的性质，我们已经知道 $W=\textbf{Span}(\beta_1\cup\beta_2)$，因此我们只需证明 $\beta_1\cup\beta_2$ 线性无关。不妨设 $\beta_1=\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m\}$; $\beta_2=\{v_{m+1},\cdots,\boldsymbol{v}_n\}$，我们考虑线性组合 $(c_1\boldsymbol{v}_1+\cdots+c_m\boldsymbol{v}_m)+(c_{m+1}\boldsymbol{v}_{m+1}+\cdots+c_n\boldsymbol{v}_n)=0$，其中令 $\boldsymbol{x}=(c_1\boldsymbol{v}_1+\cdots+c_m\boldsymbol{v}_m)$，$\boldsymbol{y}=(c_{m+1}\boldsymbol{v}_{m+1}+\cdots+c_n\boldsymbol{v}_n)$，根据直和的性质，我们再一次知道 $\boldsymbol{x}+\boldsymbol{y}=0$ 当且仅当 $\boldsymbol{x}=\boldsymbol{y}=\boldsymbol{0}$，即 $c_1=c_2=\cdots=c_n=0$，因此 $\beta_1\cup\beta_2$ 为 W 的一组基底。 ■



至此，第一章的全部内容到此结束。我们在第一章中主要学习了线性方程组与矩阵和向量之间的关系和运算，那么在下一章——《作用于线性空间之间的变换》中，我们会重点研究线性变换，从宏观的角度重新审视第一章所学的内容。

1.4 练习

---

1. 设 Q 为 4×5 矩阵，且 $\dim(\mathbf{Ker}(Q^T))=2$，那么 Q 中最大线性无关行向量的数目是多少？最大线性无关列向量的数目是多少？

2. 设 $M=\begin{bmatrix} -- & \boldsymbol{v}_1 & -- \\ -- & \boldsymbol{v}_2 & -- \\ & \vdots & \\ -- & \boldsymbol{v}_{2024} & -- \end{bmatrix}$, 若 $\mathbf{Span}\{M^T\boldsymbol{v}_1, M^T\boldsymbol{v}_2,\cdots,M^T\boldsymbol{v}_{2024}\}\subset\mathbb{R}^{2023}$, 则 $\mathbf{rank}(M)\le(\qquad)$; $\dim(\mathbf{Ker}(M^T))\ge(\qquad)$（在括号内填上合适的数字）。

3. 对于矩阵 $P=\begin{bmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \\ 4 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}$, 分别求出 $\mathbf{Ker}(P),\mathbf{Im}(P)$ 的一组基底。

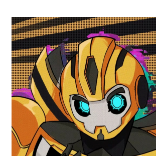
4. 在线性空间 $\mathbb{P}_3(x)$ 中, 已知线性空间 $W=\{f(x)\in\mathbb{P}_3(x): f'(1)+f(1)=0\}$, 分别求出 $\mathbf{Ker}(W),\mathbf{Im}(W)$ 的一组基底。

5. 若线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 无解，证明：$\mathbf{rank}(\begin{bmatrix}\mathcal{A} & \boldsymbol{b}\end{bmatrix})=\mathbf{rank}(\mathcal{A})+1$。

6. 对于任意的线性空间 U,V 而言，存在下列定理：

$$\dim(U+V)=\dim(U)+\dim(V)-\dim(U\cap V)$$

试证明这一定理。（*HINT*：我们不妨设 $\alpha=\{\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n\}$ 为 $U\cap V$ 的一组基底，那么根据定义我们可以对 α 进行扩充，也就是设 $\beta=\alpha\cup\{\boldsymbol{u}_1,\cdots\boldsymbol{u}_n\}$ 为 U 的一组基底；$\gamma=\alpha\cup\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ 为 V 的一组基底。随后我们证明 $\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ 为 $U+V$ 的一组基底，随后便可以得出其维数。）



2 作用于线性空间之间的变换

经过了第一章的学习，我们对于线性代数的一些基本概念已经有所了解。在第一章中，我们主要研究的是对于给定向量和向量组之间的关系。那么在本章节里面我们将会从宏观的角度去进一步研究矩阵和向量之间的关系。我们可以把矩阵和向量的乘法 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 当作是一个函数，其中矩阵便代表了这个函数的法则，$\boldsymbol{x}$ 便代表了函数里的输入值。通过给定的法则我们便可以知道其结果 $\boldsymbol{b}$ 的取值。我们便很想知道这些法则之间究竟有着怎么样的奥秘？这些法则和函数之间的相同点与不同点分别是什么？在线性代数中，我们称这样的运算法则为线性变换，它是线性代数得以发展壮大的媒介，任何矩阵的背后都有其独特的线性变换。通过完成对本章的学习，我们将建立起矩阵和线性变换的关系，届时我们对线性代数的理解便会更上一层楼。

2.1 线性变换和矩阵的关系

对于函数 $y=f(x)$ 而言，我们可以写成映射的形式 $f:X\to Y$, 其中集合 X 代表函数 f 的定义域，集合 Y 代表函数 f 的值域。即 $x\in X$, $y=f(x)\in Y$。那么我们同样可以从函数的角度去考虑我们的老朋友 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$, 我们知道如果 $\mathcal{A}$ 为 $m\times n$ 矩阵，那么 $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{b}\in\mathbb{R}^m$。因此矩阵 $\mathcal{A}$ 同样也可以看作类同于 f 的运算法则，我们一般用 $\mathcal{T}$ 表示这样的运算法则。那么形如 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 的方程即可以表示为 $\mathcal{T}:\mathbb{R}^n\longrightarrow\mathbb{R}^m$。其中 $\mathbb{R}^n$ 即为运算 $\mathcal{T}$ 的定义域，我们一般也用集合 $\mathcal{U}\subset\mathbb{R}^n$ 来表示定义域；用 $\mathcal{V}\subset\mathbb{R}^m$ 来表示值域。我们知道，对于函数 $y=f(x)$ 而言，对给定的 x 而言，有且仅有一个 y 使得 $f(x)=y$，否则便不满足函数的定义。那么对于我们所定义的运算 $\mathcal{T}$ 而言，其当然也有一定的限制条件。我们把这种满足条件的运算 $\mathcal{T}$ 称作是**线性变换 (Linear Transformation)**，其具体定义如下：

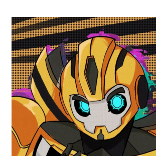
定义 2-1
在数域 $\mathbb{F}$ 上，我们定义 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换，如果对任意的 $\boldsymbol{v},\boldsymbol{u}\in\mathcal{U},\lambda\in\mathbb{F}$, 我们有

$$① \ \mathcal{T}(\boldsymbol{x}+\boldsymbol{y})=\mathcal{T}(\boldsymbol{x})+\mathcal{T}(\boldsymbol{y}) \quad ② \ \lambda\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\lambda\boldsymbol{x})$$

在目前阶段，除非有特殊说明，否则我们一律默认作用于线性变换的数域为全体实数 $\mathbb{R}$。看到这样的定义，读者能否尝试从函数的角度去理解？读者可以尝试构造出一个函数 $g(x)$, 使得为线性变换。

定理 2-1
设 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换，则：

$$① \ \mathcal{T}(\mathbf{0})=\mathbf{0} \quad ② \mathcal{T}(-\boldsymbol{v})=-\mathcal{T}(\boldsymbol{v})$$



定理 2-1 的证明

证明. ① : 根据零向量的定义，$\mathbf{0}=\mathbf{0}+\mathbf{0}$, 则 $\mathcal{T}(\mathbf{0})=\mathcal{T}(\mathbf{0}+\mathbf{0})=\mathcal{T}(\mathbf{0})+\mathcal{T}(\mathbf{0})$, 即 $\mathcal{T}(\mathbf{0})=\mathbf{0}$。
② : 我们取常数 $\lambda=-1$, 则 $\mathcal{T}(-\boldsymbol{v})=\mathcal{T}((-1)\boldsymbol{v})=-\mathcal{T}(\boldsymbol{v})$。 ■

虽然我们发现 $\mathcal{T}(\mathbf{0})=\mathbf{0}$，但我们要注意“此零非彼零”，两个零向量分别指代线性空间 $\mathcal{U}$ 和 $\mathcal{V}$ 的零向量，这一点我们要注意。但是对于任何线性空间而言，零向量的定义是不变的，即 $\mathbf{0}+\boldsymbol{u}=\boldsymbol{u}$。在特定情况下我们会使用下角标来明确零向量所处的线性空间，即定理 2-1 还可以写作 $\mathcal{T}(\mathbf{0}_U)=\mathbf{0}_V$。

例 2.1.1
设作用于 $\mathbb{R}^2$ 上的运算 H 满足 $H(x,y)=(|x|,|y|)$, 那么 H 是否为线性变换?

我们可以先验证定理 *2-1*，即验证 $H(0,0)=(0,0)$，这一点显然满足题意。随后我们设 $\boldsymbol{x}=(x_1,y_1);\boldsymbol{x}_2=(x_2,y_2)$，那么 $H(\boldsymbol{x}_1+\boldsymbol{x}_2)=H((x_1+x_2,y_1+y_2))=(|x_1+x_2|,|y_1+y_2|)$。而 $H(\boldsymbol{x}_1)+H(\boldsymbol{x}_2)=(|x_1|,|y_1|)+(|x_2|,|y_2|)=(|x_1|+|x_2|,|y_1|+|y_2|)$，我们知道 $|x+y|\le|x|+|y|$，因此 $H(\boldsymbol{x}_1+\boldsymbol{x}_2)\ne H(\boldsymbol{x}_1)+H(\boldsymbol{x}_2)$。因此 H 不为线性变换。

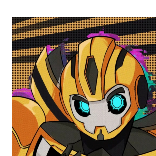
我们不妨再来看一个例子:

例 2.1.2
设运算 $D:\mathbb{P}_n(x)\longrightarrow\mathbb{P}_{n-1}(x)$, 设 $f(x)\in\mathbb{P}_n(x)$, 满足 $D(f(x))=f'(x)$, 那么 D 是否为线性变换?

由于常值函数的导数为零，因此 $f(x)=0$ 时 $f'(x)=0$。根据求导法则，我们很容易知道 $(f(x)+g(x))'=f'(x)+g'(x);(cf(x))'=cf'(x)$，因此 D 为线性变换。

接下来，在我们知道线性变换的抽象定义之后，我们怎么把抽象问题具体化呢？于是我们来一起研究一下线性变换和矩阵之间的关系。我们不妨设 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$ 为一个线性变换，然后我们再设 $\beta=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n\}$ 为线性空间 $\mathcal{U}$ 的一组基底。那么我们知道对任意的 $\boldsymbol{x}\in\mathcal{U}$, 存在线性组合 $\boldsymbol{x}=c_1\boldsymbol{u}_1+c_2\boldsymbol{u}_2+\cdots+c_n\boldsymbol{u}_n$, 并且我们知道向量 $\boldsymbol{x}$ 的坐标为 $[\boldsymbol{x}]_\beta=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}_\beta$。因此，我们还可以根据线性变换的性质得到 $\mathcal{T}(\boldsymbol{x})=\mathcal{T}(c_1\boldsymbol{u}_1+\cdots+c_n\boldsymbol{u}_n)$, 即为

$$\mathcal{T}(\boldsymbol{x})=c_1\mathcal{T}(\boldsymbol{u}_1)+\cdots+c_n\mathcal{T}(\boldsymbol{u}_n)$$



此时，我们设矩阵 $\mathcal{A}$ 的列向量分别为 $\mathcal{T}(\boldsymbol{u}_1),\cdots,\mathcal{T}(\boldsymbol{u}_n)$ 在 $\mathcal{V}$ 的基底 γ 的坐标，那么我们有

$$\mathcal{T}(\boldsymbol{x})=\mathcal{A}\boldsymbol{x}=\begin{bmatrix} | & | & & | \\ [\mathcal{T}(\boldsymbol{u}_1)]_\gamma & [\mathcal{T}(\boldsymbol{u}_2)]_\gamma & \cdots & [\mathcal{T}(\boldsymbol{u}_n)]_\gamma \\ | & | & & | \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}_\beta = c_1[\mathcal{T}(\boldsymbol{u}_1)]_\gamma+\cdots+[\mathcal{T}(\boldsymbol{u}_n)]_\gamma$$

注意到我们最终得到的结果 $\mathcal{T}(\boldsymbol{x})$ 与 $\mathcal{U},\mathcal{V}$ 的基底选取均有关，为了准确，我们会写作 $[\mathcal{T}(\boldsymbol{x})]_\beta^\gamma$。其中下角标表示“定义域的基底”，在本例中即为 $\mathcal{U}$ 的基底 β；上角标表示“值域的基底”，在本题中即为 $\mathcal{V}$ 的基底 γ。值得注意的是，经由上面的推理，我们发现最终结果 $\mathcal{T}(\boldsymbol{x})$ 是在基底 γ 下所得到的结果，因此对于特定的问题而言，我们也要选取特定的基底 γ, 使得运算更为简便。这一点我们会在后面几节当中深入学习。

定理 2-2
设 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换, $\beta=\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n\}$ 为 $\mathcal{U}$ 的一组基底，γ 为 $\mathcal{V}$ 的一组基底，那么对于 $\boldsymbol{x}=c_1\boldsymbol{u}_1+\cdots+c_n\boldsymbol{u}_n\in\mathcal{U}$, 有

$$[\mathcal{T}(\boldsymbol{x})]_\beta^\gamma=c_1[\mathcal{T}(\boldsymbol{u}_1)]_\gamma+\cdots+c_n[\mathcal{T}(\boldsymbol{u}_n)]_\gamma$$

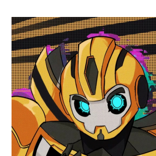
我们这样做的目的还是为了使得计算更加简便，当 $\boldsymbol{x}$ 整体不好计算时，我们可以把 $\boldsymbol{x}$ 写成若干基底向量的线性组合，然后对形式简单的基底向量施加线性变换，最后再进行加和。我们来通过一个具体的例子来看这种简化思想是怎么体现的：

例 2.1.3
在平面直角坐标系 $\mathbb{R}^2$ 中，线性变换 R 满足以下性质：对于任意的 $\boldsymbol{x}\in\mathbb{R}^2$, $R(\boldsymbol{x})$ 的结果为将向量 $\boldsymbol{x}$ 绕坐标原点顺时针旋转角度 $\theta(0\le\theta\le\pi/2)$。那么对于任意的向量 $\boldsymbol{x}=\begin{bmatrix} a \\ b \end{bmatrix}_E$ 而言, 计算其结果 $[R(\boldsymbol{x})]_E$

对于本题中提到的线性变换 $R:\mathbb{R}^2\longrightarrow\mathbb{R}^2$, 此时我们可以统一选取标准基底 $E=\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix},\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$, 然后我们对基底向量施加变换 R：在平面直角坐标系中，通过几何关系便不难发现将 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 顺时针旋转角度 θ 之后得到 $\begin{bmatrix} \cos(\theta) \\ -\sin(\theta) \end{bmatrix}$, 同理不难求出 $R\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)=\begin{bmatrix} \sin(\theta) \\ \cos(\theta) \end{bmatrix}$, 则变换 R 对应的矩阵形式（在标准基底 E 下）即为 $[R]_E=\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$, 将其乘以向量 $\boldsymbol{x}$, 我们得到

$$[R](\boldsymbol{x})=a\begin{bmatrix} \cos(\theta) \\ -\sin(\theta) \end{bmatrix}+b\begin{bmatrix} \sin(\theta) \\ \cos(\theta) \end{bmatrix}=\begin{bmatrix} a\cos(\theta)+b\sin(\theta) \\ -a\sin(\theta)+b\cos(\theta) \end{bmatrix}$$

其中我们所得到的结果为在标准基底 E 下的坐标。



通过例 2.1.3 我们便不难发现，基底的选取要巧妙，我们可以找能够最大程度简化运算量的基底。如果在例 2.1.3 中我们选取 $\beta=\left(\begin{bmatrix}2023\\2024\end{bmatrix},\begin{bmatrix}2024\\-2023\end{bmatrix}\right)$ 的话，不是不能算，但要是像这样去求这些向量绕原点旋转角度 θ 的话，与标准基底相比肯定就会复杂许多。但是，当我们费尽心思去研究如何让选取一组完美的基底的时候，我们同时也在与正确答案"渐行渐远"。为什么我会这么说呢？我们再来看一个 $\mathbb{R}^2$ 的例子：

例 2.1.4

在平面直角坐标系 $\mathbb{R}^2$ 中，线性变换 P 满足以下性质：对于任意的 $\boldsymbol{x}\in\mathbb{R}^2$, $P(\boldsymbol{x})$ 的结果为将向量 $\boldsymbol{x}$ 投影至直线 $y=2x$ 上。那么对于任意的向量 $\boldsymbol{x}=\begin{bmatrix}a\\b\end{bmatrix}$ 而言, 计算其结果 $P(\boldsymbol{x})$

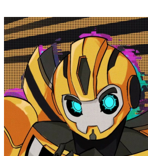
我们是不是可以直接选取这样的一组基底 $\beta=\{\boldsymbol{v}_1,\boldsymbol{v}_2\}$, 使得 $\boldsymbol{v}_1$ 恰在直线上，这样一来 $P(\boldsymbol{v}_1)=\boldsymbol{v}_1=\begin{bmatrix}1\\0\end{bmatrix}_\beta$，投影即为其自身。满足条件的向量很多，我们便可以取 $\boldsymbol{v}_1=\begin{bmatrix}1\\2\end{bmatrix}$；对于 $\boldsymbol{v}_2$ 而言，我们是否也可以找一个与直线垂直的向量？这样 $P(\boldsymbol{v}_2)=0=\begin{bmatrix}0\\0\end{bmatrix}_\beta$，我们可以取 $\boldsymbol{v}_2=\begin{bmatrix}-2\\1\end{bmatrix}$, 那么这样根据定义我们就有该线性变换的矩阵形式 $\begin{bmatrix}P(\boldsymbol{v}_1) & P(\boldsymbol{v}_2)\end{bmatrix}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$

这样一来，我们如果取 xOy 内任意一点 $\begin{bmatrix}a\\b\end{bmatrix}$, 也就是说该向量投影至 $y=2x$ 之后的结果应该是 $a\begin{bmatrix}1\\0\end{bmatrix}+b\begin{bmatrix}0\\0\end{bmatrix}=\begin{bmatrix}a\\0\end{bmatrix}$。这样计算有什么问题?

读者不妨自行思考一下为什么。这个问题尤为重要，它很好地打破了我们的定式思维，当时我在学习这一块的时候也是百思不得其解，所以如果读者实在难以理解，我愿引用我的代数学教授的一句话"*Let the question sink*", 这句话的意思有同于"让子弹飞一会儿"现在不理解太正常不过了，只有当你困惑的时候，才代表你真正学习思考了。我也愿用王国维所提到的三句诗词来概括我们的学习历程：第一阶段：昨夜西风凋敝树，独上高楼，望尽天涯路；第二阶段：衣带渐宽终不悔，为伊消得人憔悴；第三阶段：众里寻他千百度，蓦然回首，那人却在灯火阑珊处。希望读者有所启发。

我们来分析一下问什么我们的结果是错误的。首先是一个定式思维问题：在线性代数中，我们的定式思维是什么呢? 我们的定式思维是向量的坐标默认为标准基底 E, 所以读者会发现很多时候我并没有对向量的基底加以强调，因为我们默认了彼此之间交谈的语言为标准基底，然而随着不同基底的出现，我们有了说不同语言的人，此时我们要见人下菜了。

就像在例 2.1.4 的例子中，我们选取的基底不是我们大家默认的标准基底，这一点并不可怕，可怕的是我们把新的基底和熟知的标准基底混为一谈。我们看到在矩阵 $\begin{bmatrix}P(\boldsymbol{v}_1) & P(\boldsymbol{v}_2)\end{bmatrix}$ 里面, 我们所讨论的是新的基底 β, 然而后面我们做矩阵和向量乘法的时候，$\boldsymbol{x}$ 是什么？它是标准基底里面的向量。它们之间做运算，就如鸡同鸭讲。我在第一章第三节里面就已经强调所有的运算都要最终变成同一基底下的运算。所以这样是行不通的，我们能做的就是把 $\boldsymbol{x}$ 化成为基底 β 下的坐标，然后再计算。但是就算这样，我们最终得到的结果还是在 β 基底下分解的形式，我们还要再进行一次转化，使得它变成我们普遍接受的标准基底。因此在目前阶段，我



们还是老老实实地用标准基底 $E=\{\boldsymbol{e}_1,\boldsymbol{e}_2\}$, 计算 $\boldsymbol{e}_1,\boldsymbol{e}_2$ 在直线 $y=2x$ 上的投影（并不难算）。

既然这样，变换基底岂不是更加复杂了？我会在稍微靠后的章节讲解基变换，到时候我们便会知道，只需要做几个乘法，答案呼之欲出，这个乘法就是大名鼎鼎的 $\mathcal{S}=\mathcal{P}^{-1}\mathcal{A}\mathcal{P}$。对于这些不伦不类的基底，我们到此为止。直到我再次长篇幅提起之前，我们一律使用定式思维，把所有线性空间的基底全部当成标准基底。

定理 2-3

设 $\mathcal{T},\mathcal{S}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换, $\beta=\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n\}$ 为 $\mathcal{U}$ 的一组基底，若对任意的 $\boldsymbol{v}_i\in\beta$，有 $\mathcal{T}(\boldsymbol{v}_i)=\mathcal{S}(\boldsymbol{v}_i)$，则 $\mathcal{T}=\mathcal{S}$。

定理 2-3 的证明

证明. 任取 $\boldsymbol{x}\in\mathcal{U}$, 则有 $\boldsymbol{x}=c_1\boldsymbol{v}_1+\cdots+c_n\boldsymbol{v}_n$, 则 $\mathcal{T}(\boldsymbol{x})=c_1\mathcal{T}(\boldsymbol{v}_1)+\cdots+c_n\mathcal{T}(\boldsymbol{v}_n)=c_1\mathcal{S}(\boldsymbol{v}_1)+\cdots+c_n\mathcal{S}(\boldsymbol{v}_n)=\mathcal{S}(\boldsymbol{x})$。 ∎

例 2.1.5

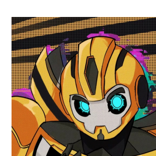
在空间直角坐标系 $\mathbb{R}^3$ 中，线性变换 $P:\mathbb{R}^3\longrightarrow\mathbb{R}^3$ 满足以下性质：$P\left(\begin{bmatrix}1\\-1\\2\end{bmatrix}\right)=\begin{bmatrix}1\\0\\0\end{bmatrix};P\left(\begin{bmatrix}0\\1\\1\end{bmatrix}\right)=\begin{bmatrix}0\\1\\0\end{bmatrix};P\left(\begin{bmatrix}2\\1\\-1\end{bmatrix}\right)=\begin{bmatrix}0\\0\\1\end{bmatrix}$ 据此求 $P\left(\begin{bmatrix}4\\2\\0\end{bmatrix}\right)$

由线性变换的性质，我们可以直接将三个已知向量当成基底向量（这一点需要先验证是否构成基向量）然后我们知道存在这样的线性组合：

$$\begin{bmatrix}4\\2\\0\end{bmatrix}=c_1\begin{bmatrix}1\\-1\\2\end{bmatrix}+c_2\begin{bmatrix}0\\1\\1\end{bmatrix}+c_3\begin{bmatrix}2\\1\\-1\end{bmatrix}$$

不难解出原方程的解为 $c_1=2,c_2=3,c_3=1$ 这样一来

$$\left(\begin{bmatrix}4\\2\\0\end{bmatrix}\right)=2P\left(\begin{bmatrix}1\\-1\\2\end{bmatrix}\right)+3P\left(\begin{bmatrix}0\\1\\1\end{bmatrix}\right)+1P\left(\begin{bmatrix}2\\1\\-1\end{bmatrix}\right)=\begin{bmatrix}2\\3\\1\end{bmatrix}$$

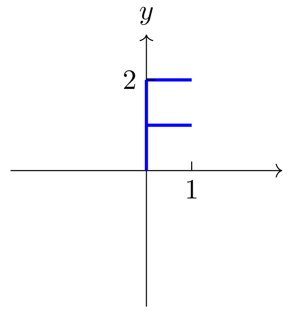


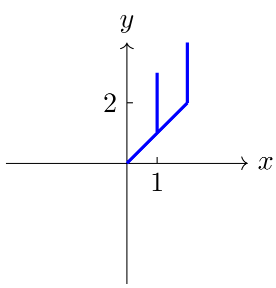
2.1 练习

1. 设 $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4 : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ 为 4 个线性变换，其在标准基底下的变化矩阵分别为

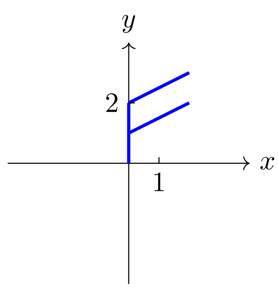
$$\mathcal{T}_1 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}; \mathcal{T}_2 = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}; \mathcal{T}_3 = \begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix}; \mathcal{T}_4 = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix};$$

那么对于如下图所示的图形而言，经过上述的四种变换之后分别对应下列选项中的哪一个图像？

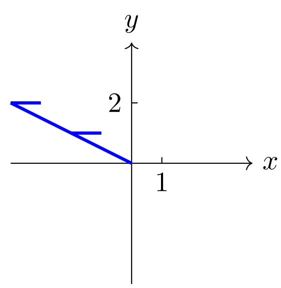




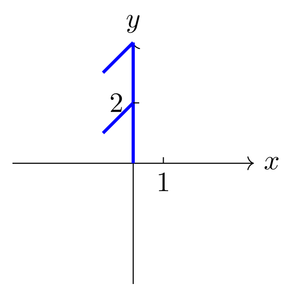
(A)



(B)



(C)



(D)

2. 我们设 $\mathcal{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ 为线性变换，我们任取 $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots \boldsymbol{v}_n \in \mathbb{R}^n$，那么 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 彼此线性无关是 $\mathcal{T}(\boldsymbol{v}_1), \mathcal{T}(\boldsymbol{v}_2), \cdots, \mathcal{T}(\boldsymbol{v}_n)$ 彼此线性无关的 (   )

(A) 充要条件　　(B) 充分但不必要条件　　(C) 必要但不充分条件　　(D) 既不充分也不必要条件

3. 设线性变换 $\mathcal{T} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, 且 $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ 为 $\mathbb{R}^3$ 的一组基底，已知 $\mathcal{T}(\boldsymbol{e}_1 + \boldsymbol{e}_2) = \boldsymbol{e}_3; \mathcal{T}(\boldsymbol{e}_1 + \boldsymbol{e}_3) = \boldsymbol{e}_2; \mathcal{T}(\boldsymbol{e}_2 + \boldsymbol{e}_3) = \boldsymbol{e}_1$, 据此求出 $\mathcal{T}(\boldsymbol{e}_1) + \mathcal{T}(\boldsymbol{e}_2) + \mathcal{T}(\boldsymbol{e}_3)$。

4. 设线性变换 $Q : \mathbb{P}_2(x) \longrightarrow \mathbb{R}$, 已知 $Q(x+2) = 1; Q(1) = 5; Q(x^2 + x) = 0$, 据此求 $Q(2 - x + 3x^2)$。

5. 在平面直角坐标系 xOy 中，我们将抛物线 $y = x^2$ 以 $(0,0)$ 为旋转中心绕原点顺时针旋转角度 $\theta, (0 < \theta < \pi/2)$, 使得旋转过后抛物线与 x 轴相交于 $(5,0)$, 通过建立合适的线性变换模型，求出此时抛物线旋转的角度 θ。

2.2 线性变换的函数性质

在上一章中，我们引出了线性变换的概念以及线性变换的基本性质，那么在本节中我们将通过函数的角度去重新审视线性变换。我说过线性变换就好比是作用于线性空间的函数，那么这个函数是否也具有一些函数特性呢？我们不妨设平面直角坐标系 xOy 内的函数 $y = f(x)$, 我们如果要研究 $f(x)$, 有哪些值得关注的点？我们自然会想到函数的定义域和值域；更进一步会研究函数的零点，单调性和极值等。这些概念我们可以尝试与线性变换 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$ 进行一一对应。

首先我们来看定义域和值域，在函数 $y = f(x)$ 中，定义域和值域均为 $\mathbb{R}$ 的开子集（至于为什么是开子集，这个不是线性代数所关心的问题，感兴趣的读者可以在拓扑学的课程中深入学习欧式空间），我们可以写作 $f : A \subset \mathbb{R} \longrightarrow B \subset \mathbb{R}$，对与线性变换而言，我们知道其本质还是 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$，其中 $\mathcal{A}$ 为变换 $\mathcal{T}$ 的矩阵形式 (再次强调，矩阵 $\mathcal{A}$ 与基底的选取有关，不同的基底对应不同的变化矩阵)，$\boldsymbol{x}\in\mathcal{U}, \boldsymbol{b}\in\mathcal{V}$。我们把所有满足条件的 $\boldsymbol{b}$ 构成的集合称作线性变换的**像 (Image)**, 记作 $\mathbf{Im}(\mathcal{T})$。不难发现这一点和矩阵的像空间定义是相同的，因为我们知道每一个线性变换其实均对应一个矩阵，因此 $\mathbf{Im}(\mathcal{T}) = \mathbf{Span}(C_{\mathcal{A}})$, 即为 $\mathcal{A}$ 中列向量构成的全体线性组合。

定义 2-2
设 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换，我们定义 $\mathcal{T}$ 的像为

$$\mathbf{Im}(\mathcal{T}) := \{\boldsymbol{y} = \mathcal{T}(\boldsymbol{x}) : \boldsymbol{x}\in\mathcal{U}, \boldsymbol{y}\in\mathcal{V}\}$$

因此我们便可以把 $\mathbf{Im}(\mathcal{T})$ 当作是线性变换的"值域"。随后我们联想矩阵里面的另一个定义——核空间，我们知道矩阵 $\mathcal{A}$ 的核空间是所有满足 $\mathcal{A}\boldsymbol{x} = \mathbf{0}$ 的解 $\boldsymbol{x}$，这一点是不是也很像函数的零点？因此我们同样可以定义线性变换的**核 (Kernel)**：

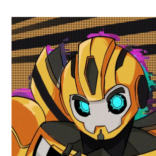
定义 2-3
设 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换，我们定义 $\mathcal{T}$ 的核为

$$\mathbf{Ker}(\mathcal{T}) := \{\mathcal{T}(\boldsymbol{x}) = \mathbf{0} : \boldsymbol{x}\in\mathcal{U}\}$$

我们已经知道 $\mathbf{Im}(\mathcal{T})\subset\mathcal{V}; \mathbf{Ker}(\mathcal{T})\subset\mathcal{U}$，因此我们可以给出定理 1-16 的线性变换形式：

定理 2-4 （秩零定理的线性变换形式）
对任意的线性变换 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$, 有

$$\dim(\mathbf{Im}(\mathcal{T})) + \dim(\mathbf{Ker}(\mathcal{T})) = \dim(\mathcal{U})$$



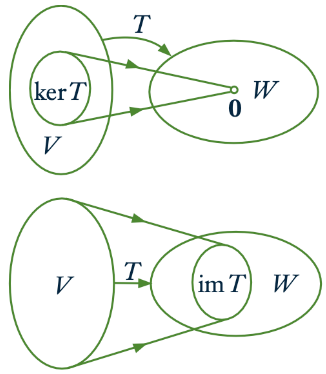


图 1: 核空间与像空间的大致图像关系

图 1 便是一个线性变换中核空间与像空间的大致图像关系。

例 2.2.1

在标准基底 E 下，设线性变换 $\mathbb{T}:\mathbb{R}^3 \longrightarrow \mathbb{R}^3$ 满足

$$\mathcal{T}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y-z \\ x-y+z \\ -x+y+z \end{bmatrix}$$

据此分别求出 $\mathbf{Ker}(\mathcal{T}), \mathbf{Im}(\mathcal{T})$。

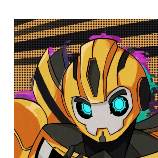
我们根据线性变换的定义，设 E 为标准基底，那么我们有：

$$\mathcal{T}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1+0-0 \\ 1-0+0 \\ -1+0+0 \end{bmatrix}; \mathcal{T}\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0+1-0 \\ 0-1+0 \\ -0+1+0 \end{bmatrix}; \mathcal{T}\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0+0-1 \\ 0-0+1 \\ -0+0+1 \end{bmatrix};$$

因此我们可以得到线性变换 $\mathcal{T}$ 在标准基底 E 下所对应的变化矩阵 $\mathcal{A}$:

$$\mathcal{A} = \begin{bmatrix} | & | & | \\ \mathcal{T}(\boldsymbol{e}_1) & \mathcal{T}(\boldsymbol{e}_2) & \mathcal{T}(\boldsymbol{e}_3) \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

因此，我们只需要研究 $\mathbf{Im}(\mathcal{A}), \mathbf{Ker}(\mathcal{A})$ 即可。我们首先可以研究线性齐次方程组 $\mathcal{A}\boldsymbol{x} = \mathbf{0}$ 从而求解 $\mathbf{Ker}(\mathcal{A})$。不难得出此时 $\mathcal{A}$ 的增广矩阵为 $\left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right]$，通过高斯消元我们最后得到 $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$，因此我们得到 $\mathbf{rank}(\mathcal{A}) = 3$，根据秩零定理，我们可以直接得到 $\mathbf{Im}(\mathcal{A}) = \mathbb{R}^3; \mathbf{Ker}(\mathcal{A}) = \{\mathbf{0}\}$。因此在原线性变换 $\mathcal{T}$ 中，我们得到 $\mathbf{Im}(\mathcal{T}) = \mathbb{R}^3; \mathbf{Ker}(\mathcal{T}) = \{\mathbf{0}\}$。



那么对于任意的线性变换 $\mathcal{T}$ 而言，我们是否必须通过高斯消元的方法，从而才能求出线性变换的核空间与像空间呢？我们先引出几条定义：

定义 2-4
设 $\mathcal{T}:\mathcal{U}\to\mathcal{V}$ 为定义在线性空间 $\mathcal{U},\mathcal{V}$ 之间的线性变换，则：

① 我们称 $\mathcal{T}$ 为**单射** ***(Injective)***，如果对于任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathcal{U}$ 而言，$\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\boldsymbol{y})$ 的充要条件是 $\boldsymbol{x}=\boldsymbol{y}$；

② 我们称 $\mathcal{T}$ 为**满射** ***(Surjective)***，如果对于任意的 $\boldsymbol{y}\in\mathcal{V}$ 而言，都存在 $\boldsymbol{x}\in\mathcal{U}$，使得 $\boldsymbol{y}=\mathcal{T}(\boldsymbol{x})$；

③ 我们称 $\mathcal{T}$ 为**双射** ***(Bijective)***，如果 $\mathcal{T}$ 既为单射，也为满射。

例 2.2.2
在例 2.2.1 中，线性变换 $\mathcal{T}$ 为定义 2-4 中的哪种类型？

我们先判断 $\mathcal{T}$ 是否为单射：根据我们最终的结论，$\textbf{rank}(\mathcal{A})=3$，即为满秩。因此矩阵 $\mathcal{A}$ 中的列向量（为简化符号，我们记作 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$）彼此线性无关，我们此时设 $\boldsymbol{x},\boldsymbol{y}$ 的坐标分别为 $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix};\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$，那么我们不妨设 $\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\boldsymbol{y})$，于是便可以得到这样的线性方程组：

$$x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+x_3\boldsymbol{v}_3=y_1\boldsymbol{v}_1+y_2\boldsymbol{v}_2+y_3\boldsymbol{v}_3$$

整理得

$$(x_1-y_1)\boldsymbol{v}_1+(x_2-y_2)\boldsymbol{v}_2+(x_3-y_3)\boldsymbol{v}_3=0$$

根据线性无关的定义，我们知道 $x_1-y_1=0;x_2-y_2=0;x_3-y_3=0$，即 $x_1=y_1;x_2=y_2;x_3=y_3$，则 $\boldsymbol{x}=\boldsymbol{y}$。同理我们还可以设 $\boldsymbol{x}=\boldsymbol{y}$，然后根据线性变换的定义可以得出 $\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\boldsymbol{y})$，因此 $\mathcal{T}$ 为单射。

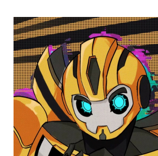
随后我们判断 $\mathcal{T}$ 是否为满射，我们已经知道 $\textbf{Span}\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}=\mathbb{R}^3$，那么对于任意的 $\boldsymbol{y}\in\mathbb{R}^3$，我们均可以找到与之对应的 $\boldsymbol{x}$，使得 $\boldsymbol{y}=\mathcal{A}\boldsymbol{x}$，因此 $\mathcal{T}$ 同样为满射。

由此我们得出 $\mathcal{T}$ 为双射。

那么我们是否有更快速的方法来判断一个线性变换所对应的种类呢？我们来看下面的定理：

定理 2-5
对任意的线性变换 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$，有

① $\mathcal{T}$ 为单射的充要条件是 $\mathcal{T}(\mathbf{0})=\mathbf{0}$；② $\mathcal{T}$ 为满射的充要条件是 $\textbf{rank}(\mathcal{T})=\dim(\mathcal{V})$



定理 2-5 的证明:

证明. 我们首先完成对①的证明:

$(\Longrightarrow)$ 我们假设 $\mathcal{T}$ 为单射, 那么 $\boldsymbol{x}=\boldsymbol{y}$ 可推出 $\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\boldsymbol{y})$, 即 $\mathcal{T}(\boldsymbol{x})-\mathcal{T}(\boldsymbol{y})=\mathbf{0}$, 即 $\mathcal{T}(\boldsymbol{x}-\boldsymbol{y})=\mathcal{T}(\mathbf{0})=\mathbf{0}$。

$(\Longleftarrow)$ 我们假设 $\mathcal{T}(\mathbf{0})=\mathbf{0}$, 假设 $\boldsymbol{x}=\boldsymbol{y}$, 则 $\boldsymbol{x}-\boldsymbol{y}=\mathbf{0}$, 即 $\mathcal{T}(\boldsymbol{x}-\boldsymbol{y})=\mathbf{0}$, 即 $\mathcal{T}(\boldsymbol{x})=\mathcal{T}(\boldsymbol{y})$, 因此 $\mathcal{T}$ 为单射。

然后, 我们再对②进行证明:

$(\Longrightarrow)$ 我们假设 $\mathcal{T}$ 为满射, 根据定义我们有 $\mathbf{Im}(\mathcal{T})=\mathcal{V}$, 即 $\dim(\mathbf{Im}(T))=\mathbf{rank}(\mathcal{T})=\dim(\mathcal{V})$。

$(\Longleftarrow)$ 我们假设 $\mathbf{rank}(\mathcal{T})=\dim(\mathcal{V})$, 则 $\dim(\mathbf{Im}(\mathcal{T}))=\mathbf{rank}(\mathcal{T})=\dim(\mathcal{V})$, 则 $\mathbf{Im}(\mathcal{T})=\mathcal{V}$, 因此 $\mathcal{T}$ 为满射。 ∎

对于定理 2-5, 我们还可以综合以前所学的全部知识, 总结出一条推论:

定理 2-5 (a)
对任意的线性变换 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$, 其中 $\mathcal{A}$ 是变换 $\mathcal{T}$ 的矩阵形式, 则下列说法是等价的:

① $\mathcal{T}$ 为单射; ② $\mathcal{T}(\mathbf{0})=\mathbf{0}$; ③ 矩阵 $\mathcal{A}$ 满秩; ④ 矩阵 $\mathcal{A}$ 的列向量彼此线性无关;

通过观察第 45 页的图 1, 我们能否从集合元素关系的角度去审视线性变换中的单射与满射? 对于一个函数 $f:X\longrightarrow Y$ 而言, 我们这样定义:

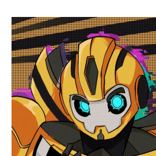
定义 2-5
设 $f:X\longrightarrow Y$ 为定义在 $\mathbb{R}$ 上的函数, $X,Y\subset\mathbb{R}$, 则:

① 我们称 f 为**单射** ***(Injective)***, 如果对于任意的 $x_1,x_2\in X$ 而言, $f(x_1)=f(x_2)$ 的充要条件是 $x_1=x_2$;

② 我们称 f 为**满射** ***(Surjective)***, 如果对于任意的 $y\in Y$ 而言, 都存在 $x\in X$, 使得 $y=f(x)$;

③ 我们称 f 为**双射** ***(Bijective)***, 如果 $\mathcal{T}$ 既为单射, 也为满射。

因此我们不难发现对于函数而言, 单射与满射同样存在。如果定义 $f:\mathbb{R}\longrightarrow\mathbb{R}$, 那么我们熟知的函数 $y=x$, 函数 $y=x^3$; 对数函数 $y=\ln(x)$ 便都是双射函数 (读者可以自行验证)。如果 $f:X\longrightarrow Y$ 为单射函数, 那么我们知道集合 X 与 Y 中的元素应该存在一一对应的关系, 也就是说 Y 中的元素应该足够多, 使得每一个 $x\in X$ 都可以有不同的 $y\in Y$ 与之对应。也就是说最起码要保证 $|X|\le|Y|$, 否则根据抽屉原理我们知道会有最少两个 X 中的不同元素与 Y 中的一个元素对应, 从而不满足单射的性质; 如果 $f:X\longrightarrow Y$ 为满射, 那么 Y 中的所有元素均要有与之相对应的元素。根据函数的定义, 对于给定的 $x\in X$ 而言, 有且仅有一



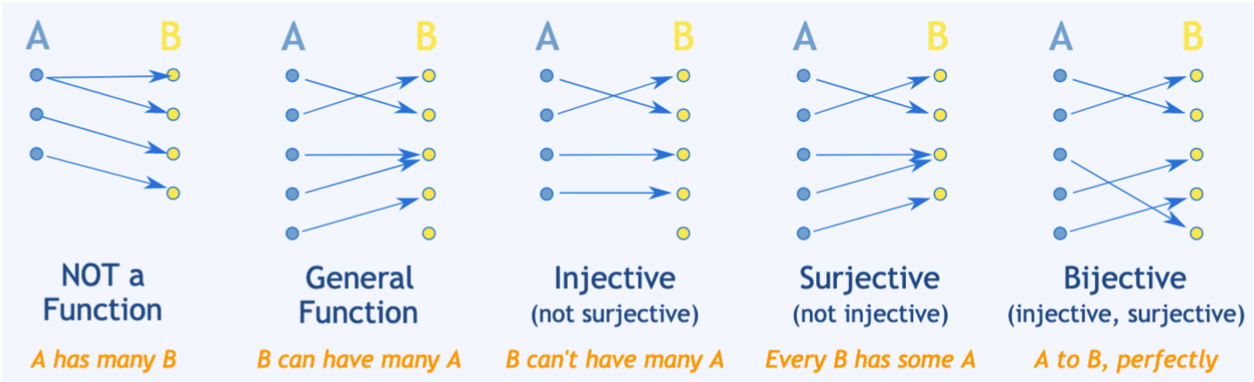


图 2: 本图中从左至右分别对应：不构成函数，一般函数，单射但不为满射，满射但不为单射，双射

个 $y \in Y$ 使得 $f(x) = y$，也就是说 X 中的元素要足够多，使得每一个 $y \in Y$ 都有不同的 $x \in X$ 与之对应，这样就意味着最起码要保证 $|Y| \leq |X|$。

根据上述推理，我们知道了若 $f : X \longrightarrow Y$ 为单射，那么不可能会有 $|X| \geq |Y|$; 若 $f : X \longrightarrow Y$ 为满射，那么不可能会有 $|Y| \geq |X|$。若 $f : X \longrightarrow Y$ 为双射，那么 $|Y| = |X|$。该结论对于函数成立，那么对与线性变换而言，该结论是否成立？答案是肯定的，但是我们要进行一定程度的修改。在线性空间里面，什么性质可以与集合的元素个数类比？答案便是线性空间的维数。维数也可以看作基底向量的个数，所以我们由此便可以得出一条重要定理：

定理 2-6
设 $\mathcal{T} : \mathcal{U} \to \mathcal{V}$ 为定义在线性空间 $\mathcal{U}, \mathcal{V}$ 之间的线性变换，则：

① 若 $\dim(\mathcal{U}) \geq \dim(\mathcal{V})$，则 $\mathcal{T}$ 不可能为单射；

② 若 $\dim(\mathcal{U}) \leq \dim(\mathcal{V})$，则 $\mathcal{T}$ 不可能为满射；

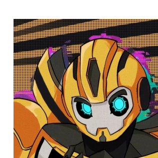
③ 若 $\dim(\mathcal{U}) = \dim(\mathcal{V})$，则 $\mathcal{T}$ 为单射的充要条件是 $\mathcal{T}$ 为满射。

因此，定理 2-6 也是很多时候快速判断某个线性变换种类的一条重要依据，其中第三条尤为重要。当 $\mathcal{U}, \mathcal{V}$ 维数相同时，我们只需要证明出该变换为单射便可以自然推出该变换为满射，反之亦然。所以双射变换往往有着非常好的性质，并且双射变换还能告诉我们两个线性空间 $\mathcal{U}, \mathcal{V}$ 之间的相似性。我们把这种双射变换称作是**同构 (Isomorphism)**，一般读作 $\mathcal{U}$ 同构于 $\mathcal{V}$ ($\mathcal{U}$ *is isomprphic to* $\mathcal{V}$)，写作 $\mathcal{U} \simeq \mathcal{V}$。同构的定义如下：

定义 2-6
设 $\mathcal{U}, \mathcal{V}$ 为线性空间，我们称 $\mathcal{U}$ 同构于 $\mathcal{V}$，记作 $\mathcal{U} \simeq \mathcal{V}$，如果存在线性变换 $\mathcal{T} : \mathcal{U} \longrightarrow \mathcal{V}$，使得 $\mathcal{T}$ 为双射。

同构在满足双射线性变换的同时也是代数结构里面的一个**等价关系 (Equivalence Relation)**，即同构还满足三条性质：

① **自反性 (Reflexive)**：即任何线性空间 $\mathcal{V}$ 同构于自身，有 $\mathcal{V} \simeq \mathcal{V}$。



② **对称性 (Symmetric)**：即 $\mathcal{U} \simeq \mathcal{V}$ 可以推出 $\mathcal{V} \simeq \mathcal{U}$。

③ **传递性 (Transitive)**：即 $\mathcal{U} \simeq \mathcal{V}, \mathcal{V} \simeq \mathcal{W}$ 可以推出 $\mathcal{V} \simeq \mathcal{W}$。

对于第一条自反性而言，我们很好证明，即可以选取最特殊的线性变换 $\mathcal{T}: \mathcal{V} \longrightarrow \mathcal{V}$, 满足 $\mathcal{T}(\boldsymbol{x}) = \boldsymbol{x}$。形如这样的变换我们也称作**单位变换 (Identity Transformation)**, 通常用 I_d 表示，随后我们不难证明 $\mathcal{T}$ 是一个双射线性变换，即 $\mathcal{V} \simeq \mathcal{V}$。对于第二条和第三条而言，我们目前还没有解决的好办法，但完成本章的学习之后我们便可以进行第二，三条的证明。其中第二条便是线性变换的逆变换，与逆矩阵关联密切；第三条便是线性变换的复合，与矩阵的乘法关联密切。

例 2.2.3

试证明 $\mathbb{R}^4 \simeq \mathbb{P}_3(x)$。

证明. 设 E 为 $\mathcal{R}^4$ 的标准基底，我们考虑这样的变换 $\mathcal{T}: \mathbb{R}^4 \longrightarrow \mathbb{P}_3(x)$，使得

$$\mathcal{T}\left(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}\right) = ax^3 + bx^2 + cx + d$$

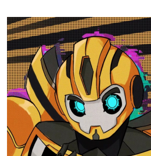
我们先证明 $\mathcal{T}$ 为线性变换：

$$\begin{aligned}
\mathcal{T}\left(\begin{bmatrix} \lambda(a_1+a_2) \\ \lambda(b_1+b_2) \\ \lambda(c_1+c_2) \\ \lambda(d_1+d_2) \end{bmatrix}\right) &= \lambda(a_1+a_2)x^3 + \lambda(b_1+b_2)x^2 + \lambda(c_1+c_2)x + \lambda(d_1+d_2) \\
&= \lambda(a_1x^3 + b_1x^2 + c_1x + d_1) + \lambda(a_2x^3 + b_2x^2 + c_2x + d_2) \\
&= \lambda\mathcal{T}\left(\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{bmatrix}\right) + \lambda\mathcal{T}\left(\begin{bmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix}\right)
\end{aligned}$$

因此，$\mathcal{T}$ 为线性变换。我们发现，$\mathcal{T}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = 0$，因此由定理 *2-5*，我们知道 $\mathcal{T}$ 为单射。又因为其中 $\dim(\mathbb{P}_3(x)) = 4 = \dim(\mathbb{R}^4)$ (定理 *1-10*)，那么根据定理 *2-6* 可知，此时 $\mathcal{T}$ 也为满射，则 $\mathcal{T}$ 为双射。因此我们有 $\mathbb{R}^4 \simeq \mathbb{P}_3(x)$。

■

同构好比几何学里面的相似三角形，我们知道两个三角形如果相似，那么它们的对边成相同比例，对角相



等。对于同构而言，如果 $\mathcal{U} \simeq \mathcal{V}$，那么线性空间 $\mathcal{U}, \mathcal{V}$ 之间也有对应的“相似”关系。简单来说，二者的代数内核是一致的，只不过用了不同的表现方式。我们前面看到的基底之间的变换就是一种同构：它们都是同一个向量在不同基底下的表现形式；我们因此可以将同构理解为同一个“向量”在不同线性空间下的表现形式。线性变换便是在不同的基底，线性空间之间转化的媒介。

在代数结构中，除了同构这一概念，同时也存在**同态 (Homorphism)** 这一概念。同态也可以看做是两个结构之间的相似关系，但是与同构不同，同态所满足的条件不如同构苛刻。我们如果设 f 是定义在两个代数结构之间的同态的话，那么我们就有 $f(x+y) = f(x) + f(y); f(cx) = cf(x)$。所以不难发现，我们所研究的线性变换就是一个天然的同态。我们可以按照映射的方式来考虑同态，设 U, V 为线性空间，我们则可以考虑所有从 V 到 W 上的线性变换 f, 我们把这些线性变换 f 也称作是**线性泛函 (Linear Functional)**。从 U 到 V 满足条件的所有线性泛函所构成的集合我们记作 $\mathbf{Hom}(U, V)$。那么我们自然而然会想，这些变换 f 是否会和向量一样，也存在一组基底？也就是说对于任何一个线性变换 T，其是否也能写成若干个线性变换的线性组合？在现阶段读者可以不用关心这一问题，我会在专门的章节里面引入**对偶空间 (Dual Space)** 的概念，届时我们所研究的不再是由向量构成的线性空间，而是由线性变换所构成的线性空间。对偶空间和线性空间也存在着众多关联，我们会在后续的章节里面重点讨论这一问题。

2.2 练习

---

1. 考虑集合 $A = \{1,2,3,4,5\}; B = \{5,6,7,8\}$, 那么在所有映射 $f : A \longrightarrow B$ 中，有多少个单射？有多少个满射？有多少个双射？

2. 设非零线性变换 $\mathcal{T} : \mathbb{R}^{2024} \longrightarrow \mathbb{R}^{2022}$, 若 $\mathcal{T}$ **不是**满射，则 $\mathbf{rank}(\mathcal{T})$ 的可能取值为（　　　　）

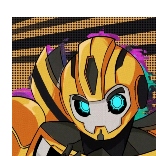
3. 设线性变换 $\mathcal{T} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, 满足 $\mathcal{T}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+kz \\ ky+z \\ x-y \end{bmatrix}, k \in \mathbb{R}$。若 $\mathcal{T}$ 为单射，k 的取值范围是什么？若 $\mathcal{T}$ 为双射，k 的取值范围是什么？若 $\mathcal{T}$ 为满射，k 的取值范围是什么？

4. 设线性空间 V 的维数为 n，证明：$V \simeq \mathbb{R}^n$。

5. 设 $V_1, V_2, \cdots, V_n (n \geq 2)$ 为定义在同一数域 $\mathbb{F}$ 上的线性空间，我们定义线性变换 $T_i : V_i \longrightarrow V_{i+1}, (1 \leq i \leq n-1)$，使得这些线性变换满足以下的性质：

$$\begin{cases} \mathbf{Ker}(T_1) = \{\mathbf{0}\} \\ \mathbf{Ker}(T_{i+1}) = \mathbf{Im}(T_i), 1 \leq i \leq n-2 \\ \mathbf{Im}(T_{n-1}) = V_n \end{cases}$$

证明:$\sum_{i=1}^{n}(-1)^i \dim(V_i) = 0$。



2.3 线性变换的代数性质

我们发现，对于线性空间而言，我们定义了加法和乘法运算，那么有没有除法运算？我们能否对两个线性空间采用“作商”的方式来构造出一个新的线性空间？我们由此引入了**商空间 (Quotient Space)** 的概念，我们首先考虑线性空间 V, W，其中 $W \subset V$。我们给出以下的定义：

定义 2-7
设 V, W 为线性空间且 $W \subset V$。对于给定的 $\boldsymbol{v} \in V$，我们定义 W 的一个**陪集 *(Coset)*** 为：

$$\boldsymbol{v} + W = \{\boldsymbol{v} + \boldsymbol{w} : \boldsymbol{w} \in W\}$$

由此不难发现，对于给定的 $\boldsymbol{v} \in W$，陪集 $\boldsymbol{v} + W$ 即为取遍所有 W 中的向量然后分别与 $\boldsymbol{v}$ 相加所构成的新集合。注意到 $\boldsymbol{v} \in \boldsymbol{v} + W$，因为 $\mathbf{0} \in W, \boldsymbol{v} = \boldsymbol{v} + \mathbf{0}$。那么什么情况下两个不同的陪集表示相同的集合呢？

定理 2-7
设 V, W 为线性空间且 $W \subset V$，设 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$，则 $\boldsymbol{v}_1 + W = \boldsymbol{v}_2 + W$ 的充要条件为 $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in W$

定理 2-7 的证明：

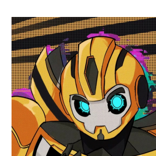
证明. 我们在此将仅证充分性。假设 $\boldsymbol{v}_1 + W = \boldsymbol{v}_2 + W$，那么对 $\boldsymbol{v}_1$ 而言，$\boldsymbol{v}_1 \in \boldsymbol{v}_1 + W$ 且 $\boldsymbol{v}_1 \in \boldsymbol{v}_2 + W$，因此存在 $\boldsymbol{w} \in W$，使得 $\boldsymbol{v}_1 = \boldsymbol{v}_2 + W$，即 $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in W$。我们同理可以得到 $\boldsymbol{v}_2 = \boldsymbol{v}_1 + \boldsymbol{w}'$，即 $\boldsymbol{v}_2 - \boldsymbol{v}_1 \in W$。 ■

那么对于一个给定的 $\boldsymbol{v}$ 而言，我们希望可以通过定理 2-7 来寻找与其性质“相似”的其他向量。我们把这些向量组成的集合 $\mathcal{P}$ 称作是 $\boldsymbol{v}$ 的一个**等价类 (Equivalence Class)**，对于 $\boldsymbol{v}_0 \in \mathcal{P}$，我们有 $\boldsymbol{v} - \boldsymbol{v}_0 \in W$，此时也可以记作 $\boldsymbol{v} \sim \boldsymbol{v}_0$。其同样为一个等价关系，即满足自反性（$\boldsymbol{v} \sim \boldsymbol{v}$），对称性 ($\boldsymbol{v} \sim \boldsymbol{v}_0 \rightarrow \boldsymbol{v}_0 \sim \boldsymbol{v}$)，传递性 ($\boldsymbol{v} \sim \boldsymbol{v}_0; \boldsymbol{v}_0 \sim \boldsymbol{w} \rightarrow \boldsymbol{v} \sim \boldsymbol{w}$)。

随后，我们把 $\boldsymbol{v}$ 的范围扩大，我们取遍所有的 $\boldsymbol{v} \in V$，计算陪集 $\boldsymbol{v} + W$，然后我们取这些陪集的并集，得到**商空间** V/W (读作 V 模 W，同时还可以写作 $V \mod W$)。

定义 2-8
设 V, W 为线性空间且 $W \subset V$。定义商空间 V/W 为

$$V/W = \{\boldsymbol{v} + W : \boldsymbol{v} \in V\}$$



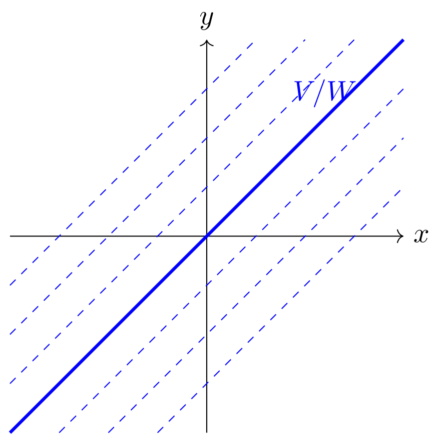
商空间同样为一个线性空间，其满足对加法和乘法运算的封闭，其运算规律满足：

定义 2-9
设 V, W 均为线性空间且 $W \subset V$，设 $\boldsymbol{x}, \boldsymbol{y} \in V, \lambda \in \mathbb{F}$，定义在数域 $\mathbb{F}$ 上的商空间 V/W 满足下列运算：

① $(\boldsymbol{x}+W)+(\boldsymbol{y}+W)=(\boldsymbol{x}+\boldsymbol{y})+W$

② $\lambda(\boldsymbol{x}+W)=\lambda\boldsymbol{x}+W$

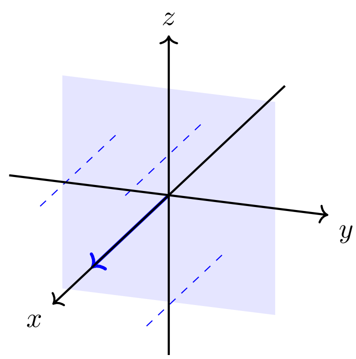
那么，我们该怎么去理解商空间 V/W 呢？设平面直角坐标系 xOy 中，我们有一条过原点的直线 $l: y=x$，我们此时设 $V=\mathbb{R}^2, W=l$ 然后我们构建 V/W。首先根据定义，在所有的陪集 $\boldsymbol{v}+W$ 中，我们取 $\boldsymbol{v}=\mathbf{0}$，那么此时陪集 $\boldsymbol{v}+W$ 所表示的图像即为 W 本身。然后我们考虑形如 $\boldsymbol{v}=\begin{bmatrix} t \\ 0 \end{bmatrix}$ 的所有向量，我们不妨设 $\boldsymbol{w}=\begin{bmatrix} x \\ y \end{bmatrix}$ 那么我们有 $\boldsymbol{v}+\boldsymbol{w}=\begin{bmatrix} x+t \\ y \end{bmatrix}$。对于给定的 t 而言，陪集 $\boldsymbol{v}+W$ 即为将图像 l 往左平移单位 t 之后所得到的新图像。同理，形如 $\boldsymbol{v}=\begin{bmatrix} 0 \\ s \end{bmatrix}$ 所构成的陪集 $\boldsymbol{v}+W$ 即为将图像 l 向上平移单位 s 之后所得到的图像。对于一般形式 $\boldsymbol{v}=\begin{bmatrix} t \\ s \end{bmatrix}$，利用线性组合的知识，我们知道其陪集 $\boldsymbol{v}+W$ 即为将 l 向左平移 t 个单位，再向上平移 s 个单位所得的结果。因此我们将这些情况进行整合，所得到的商空间 V/W 便是所有与 l 平行的直线（见图 (A)）。



(A) 该图展示了 V/W 即代表所有与 l 所平行的直线（包括 l）所构成的集合

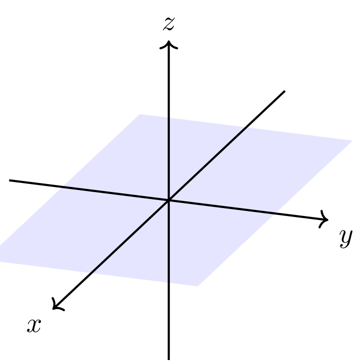
对于 $\mathbb{R}^2$ 而言，其商空间很好理解，我们来看一些更为复杂的例子，考虑 $V=\mathbb{R}^3$，我们先假设 W 代表一条过原点的直线。不失一般性地讲，我们可以把这条直线设为 $W=\textbf{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)$，即为 x 坐标轴。在此情况下我们来考虑 V/W。我们知道，根据等价类的概念，$\boldsymbol{v}_1 \sim \boldsymbol{v}_2$ 如果 $\boldsymbol{v}_1-\boldsymbol{v}_2 \in W$。我们设 $\boldsymbol{v}_1=\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}; \boldsymbol{v}_2=\begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$，

此时若 $\boldsymbol{v}_1 \sim \boldsymbol{v}_2$, 那我们有 $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \begin{bmatrix} a_1 - a_2 \\ b_1 - b_2 \\ c_1 - c_2 \end{bmatrix} \in W$, 根据 W 的定义，此时 $b_1 - b_2 = 0, c_1 - c_2 = 0$，因此在本题中，根据定义我们称 $\boldsymbol{v}_1, \boldsymbol{v}_2$ 彼此是“等价的”如果它们有着相同的 y 坐标和 z 坐标。也就是说对于固定的 y_0 z_0，所有满足形如 $(t, y_0, z_0) : t \in \mathbb{R}$ 的点都相互等价，满足这些条件的点所构成的集合即为经过 $(0, y_0, z_0)$, 并且与 x 轴平行的直线, 这条直线构成了 $y = y_0, z = z_0$ 的一个等价类。因此我们不难发现在本题中，对于一个给定的等价类，x 坐标无关紧要，所以我们完全可以仅仅用坐标 (y, z) 来表示同一个等价类。满足条件的所有等价类即为平面 yOz 上的所有点。因此在本题中 V/W 即表示平面 yOz （如图 (B))。

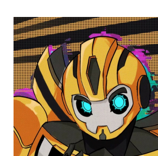


(B) 该图展示了 V/W 即为 yOz 平面 (图中蓝色区域)，其中每一条蓝色虚线即为 V/W 的一个等价类

同样地在 $V = \mathbb{R}^3$ 中, 我们此时假设 W 代表经过原点的平面。不失一般性地讲, 我们设 $W = \textbf{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$, 即为 xOy 平面。然后我们知道 $\boldsymbol{v}_1 \sim \boldsymbol{v}_2$ 如果 $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in W$, 此时若 $\boldsymbol{v}_1 \sim \boldsymbol{v}_2$, 那我们有 $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \begin{bmatrix} a_1 - a_2 \\ b_1 - b_2 \\ c_1 - c_2 \end{bmatrix} \in W$, 根据 W 的定义，此时 $c_1 - c_2 = 0$, 所以此时我们称 $\boldsymbol{v}_1, \boldsymbol{v}_2$ 是等价的，如果其有着相同的 z 坐标。所有形如 $(t, s, z_0) : s, t \in \mathbb{R}$ 的点均等价。我们同样可以仅使用 z 这一个坐标来表示该等价类，该等价类即为到 xOy 的距离为 z 的平面。取遍所有的 z 即构成了商空间 V/W, 在本题中即为 z 坐标轴。



(B) 该图展示了 V/W 即为 z 坐标轴，其中每一个与 xOy 平行的平面即为 V/W 的一个等价类



我们也可以这样来理解商空间，即通过把 V 按照给定的法则"压缩"成维数更小的其他空间。或者我们这样想：在千层饼里面，我们把千层饼看作是三维的，我们把里面近似二维的"薄层"看作是一个陪集，这些陪集叠在一起构成了整张千层饼，而决定陪集位置的便是 z 坐标，即高度。每一个确定位置的高度代表一个陪集，同一个陪集（同一层）上面的元素（芝麻，葱花）等全部等价。在商空间中，其维数由 W, V 的维数所共同决定。存在下列定理：

定理 2-8
设 V, W 为线性空间且 $W \subset V$，则

$$\dim(V/W) = \dim(V) - \dim(W)$$

限于篇幅和读者的理解能力，这个定理我们暂且不证。

在群论，环论，数论等领域均有类似的概念。我们来简单地看一下陪集在数论中的应用：如果我们定义整数集 $\mathbb{Z}$, 我们知道 $\mathbb{Z}$ 即代表全体整数。我们还可以定义 $n\mathbb{Z} := \{nz, z \in \mathbb{Z}, n \in \mathbb{N}\}$, 即为将 $\mathbb{Z}$ 中的元素全部乘以自然数 n 所得到的结果。此时我们考虑集合 $\mathbb{Z}/n\mathbb{Z}$。为简便计算，我们首先考虑 $\mathbb{Z}/2\mathbb{Z}$ 的情况。我们知道 $2\mathbb{Z}$ 代表所有的偶数集，根据定义，我们知道 $a \sim b$ 如果 $a - b \in 2\mathbb{Z}$, 即 $a - b$ 为偶数。在整数中有且仅有两种情况可以满足 $a - b$ 为偶数：a, b 均为奇数；a, b 均为偶数。所以对于 $\mathbb{Z}/2\mathbb{Z}$ 而言，有且仅有奇数集和偶数集这两个等价类，我们一般取两个集合中最小正整数来构建 $\mathbb{Z}/2\mathbb{Z}$，所以我们有

$$\mathbb{Z}/2\mathbb{Z} := \{0, 1\}$$

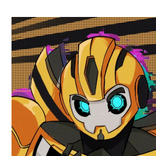
同理得出

$$\mathbb{Z}/n\mathbb{Z} := \{0, 1, \cdots, n-1\}$$

那么在 $\mathbb{Z}/2\mathbb{Z}$ 中，为什么我们说只有两个元素呢？我们来看元素 3, 首先 $1 \in \mathbb{Z}/2\mathbb{Z}$, 并且我们发现 $3 - 1 = 2 \in 2\mathbb{Z}$，因此根据等价类的知识我们说 $3 \sim 1$。同理我们就知道任何奇数均与 1 等价，同时任何偶数均与 0 等价。因此，我们也可以得出在 $\mathbb{Z}/2\mathbb{Z}$ 中的加法表（此时便是我们熟知的二进制）：

$+$	0	1
0	0	1
1	1	0

如果读者对此方面感兴趣，那么不妨尝试阅读有关环论，群论的书籍。形如 $\mathbb{Z}/p\mathbb{Z}$（p 为质数）的结构是整数环的一种，在整数环之间也存在**环同构 (Ring Isomorphism)** 与**环同态 (Ring Homorphism)**。其中环论里面非常著名的定理：**中国剩余定理 (Chinese Reminder Theorem)** 便是环同构的体现：即当 m, n 的最大公约数为 1 时，存在同构 $\mathbb{Z}/mn\mathbb{Z} \simeq \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$。限于篇幅的原因，我将不再过多讲解数论这方面的知识，有兴趣的读者可以自己做额外的阅读。



让我们回到线性变换 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$, 根据同构的性质我们知道要使得 $\mathcal{U}\simeq\mathcal{V}$，我们至少要有 $\dim(\mathcal{U})=\dim(\mathcal{V})$, 那么我们是否可以尝试构造 $\mathcal{U},\mathcal{V}$ 的一些子集，使得我们可以得到一个同构？构造子集的方法便是我们所提到的商空间。由此我们引出线性空间的**第一同构定理 (First Isomorphism Theorem)**：

定理 2-8 (第一同构定理)
设 $\mathcal{T}:\mathcal{U}\longrightarrow\mathcal{V}$ 为一个线性变换，则

$$\mathcal{U}/\,\mathbf{Ker}(\mathcal{T})\simeq\mathbf{Im}(\mathcal{T})$$

定理 2-1 的证明

证明. 我们定义运算 $\mathcal{P}:\mathcal{U}/\mathbf{Ker}(\mathcal{T})\longrightarrow\mathbf{Im}(\mathcal{T});\mathbf{Ker}(\mathcal{T})=K$, 使得

$$\mathcal{P}(\boldsymbol{u}+K)=\mathcal{T}(\boldsymbol{u})$$

跟据商空间的性质，我们知道 $\mathcal{P}(\lambda_1(\boldsymbol{u}_1+K)+\lambda_2(\boldsymbol{u}_2+K))=\mathcal{P}(\lambda_1\boldsymbol{u}_1+\lambda_2\boldsymbol{u}_2+K)=\lambda_1\mathcal{T}(\boldsymbol{u}_1)+\lambda_2\mathcal{T}(\boldsymbol{u}_2)=\lambda_1\mathcal{P}(\boldsymbol{u}_1+K)+\lambda_2\mathcal{P}(\boldsymbol{u}_2+K)$, 因此 $\mathcal{P}$ 为线性变换；我们随后设 $\boldsymbol{u}_1+K=\boldsymbol{u}_2+K$, 则 $\boldsymbol{u}_1-\boldsymbol{u}_2\in K$, 即 $\mathcal{T}(\boldsymbol{u}_1-\boldsymbol{u}_2)=\mathbf{0}$, 所以我们得到 $\boldsymbol{u}_1=\boldsymbol{u}_2$，即 $\mathcal{P}(\boldsymbol{u}_1+K)=\mathcal{P}(\boldsymbol{u}_2+K)$, 于是不难发现 $\mathcal{P}$ 为单射。并且对任意的 $\boldsymbol{v}\in\mathcal{V}$ 而言，我们到可以找到 $\boldsymbol{v}=\mathcal{T}(\boldsymbol{u})=\mathcal{P}(\boldsymbol{u}+K)$, 所以不难发现 $\mathcal{P}$ 也为满射。所以我们称 $\mathcal{P}$ 是一个双射线性变换，则 $\mathcal{U}/\mathbf{Ker}(\mathcal{T})\simeq\mathbf{Im}(\mathcal{T})$。 ■

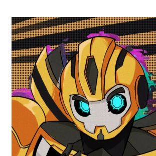
定理 2-8 (a)
设 $\mathcal{A}$ 为 $m\times n$ 矩阵，则

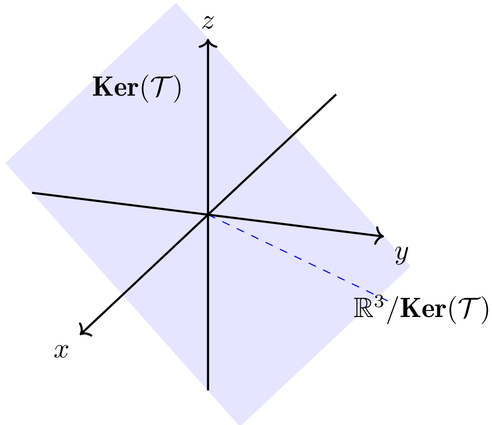
$$\mathbb{R}^n/\,\mathbf{Ker}(\mathcal{A})\simeq\mathbf{Im}(\mathcal{A})$$

我们再一次回到 $\mathbb{R}^3$ 中的一个例子来，我们不妨设线性变换 $\mathcal{T}:\mathbb{R}^3\longrightarrow\mathbb{R}^3$, 然后我们知道线性变换本质就是矩阵与向量的乘法 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$。我们假设 $\mathbf{Ker}(\mathcal{T})$ 是一条经过原点的平面（如图 (C)), 那么我们便可以设 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 为 $\mathbf{Ker}(\mathcal{T})$ 的一组基底，因此我们选定 $\boldsymbol{u}\in\mathbb{R}^3$, 考虑陪集 $\boldsymbol{u}+\mathbf{Ker}(\mathcal{T})$。其中对于任意的 $\boldsymbol{v}\in\mathbf{Ker}(\mathcal{T})$, 我们有线性组合 $\boldsymbol{v}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2$, 即陪集 $\boldsymbol{u}+\mathbf{Ker}(\mathcal{T})=\boldsymbol{u}+c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2;c_1,c_2\in\mathbb{R}$。根据向量的分解定理，我们可以将 $\boldsymbol{u}$ 分解成 $\boldsymbol{u}=\boldsymbol{u}_1+\boldsymbol{u}_2$，使得 $\boldsymbol{u}_1\in\mathbf{Ker}(\mathcal{T})$, 这样一来我们就又有

$$\begin{aligned}\boldsymbol{u}+\mathbf{Ker}(\mathcal{T})&=(\lambda_1\boldsymbol{v}_1+\lambda_2\boldsymbol{v}_2)+\boldsymbol{u}_2+c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2\\&=\boldsymbol{u}_2+[(\lambda_1+c_1)\boldsymbol{v}_1+(\lambda_2+c_2)\boldsymbol{v}_2]\,;\lambda_1,\lambda_2,c_1,c_2\in\mathbb{R}\end{aligned}$$

不难发现 $[(\lambda_1+c_1)\boldsymbol{v}_1+(\lambda_2+c_2)\boldsymbol{v}_2]\in\mathbf{Ker}(\mathcal{T})$, 那么此时我们发现真正决定陪集的只有向量 $\boldsymbol{u}_2$。如果我们考虑 $\boldsymbol{u}=\boldsymbol{u}_1+\boldsymbol{u}_2$ 恰好为一个正交分解，即 $\boldsymbol{u}_2\perp\mathbf{Ker}(\mathcal{T})$, 因为 $\dim(\mathbf{Ker}(\mathcal{T}))=2$, 因此唯一决定 $\boldsymbol{u}_2$ 的有且仅有一个维度的坐标。此时我们发现所有经过原点，且与 $\mathbf{Ker}(\mathcal{T})$ 垂直的向量均构成 $\mathbb{R}^3/\mathbf{Ker}(\mathcal{T})$ 的陪集。因此 $\mathbb{R}^3/\mathbf{Ker}(\mathcal{T})$ 便代表经过原点且与 $\mathbf{Ker}(\mathcal{T})$ 垂直的直线。





(C)

如果此时 $\mathbf{Ker}(\mathcal{T})$ 是一条经过原点的直线，那么此时 $\mathbb{R}^3/\mathbf{Ker}(\mathcal{T})$ 所表示的图象是什么？读者不妨参照上述例子自行推导。

我们知道，在线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 中，设 $\mathcal{A}$ 为 $m\times n$ 矩阵，$\boldsymbol{b}\in\mathbf{Im}(\mathcal{T})$，而 $\mathbf{Im}(\mathcal{T})\simeq\mathbb{R}^n/\mathbf{Ker}(\mathcal{T}))$ 因此我们也可以通过研究 $\mathbb{R}^n/\mathbf{Ker}(\mathcal{T}))$，从而知道其解的一些性质。具体的方法在这里不多赘述，感兴趣的读者可以自行查看相关书籍。除此之外，在商空间里面同样存在**第二同构定理 (Second Isomorphism Theorem)** 与**第三同构定理 (Third Isomorphism Theorem)**，我将不对其进行详细地说明，感兴趣的读者请自行查阅并尝试证明。

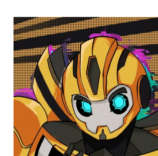
2.3 练习

1. 第一同构定理与秩零定理有什么联系？能否用第一同构定理来证明秩零定理？

2. 假设 $V=\mathbb{R}^3$，当 W 分别等于：① 零空间 $\{\mathbf{0}\}$；② 坐标轴 y 轴；③ 平面 $x-y=0$ 的时候，其商空间 V/W 分别表示什么？其几何含义是什么？

3. 在多项式空间中 (数域为 $\mathbb{R}$)，$\mathbb{P}_2(x)/\mathbb{P}_1(x)$ 表示什么？$\mathbb{P}_3(x)/\mathbb{P}_1(x)$ 表示什么？

4. 假设线性空间 V 的两个子空间为 V_1,V_2，满足 $V=V_1\bigoplus V_2$，通过建立合适的线性变换，证明 $V/V_1\simeq V_2$ 以及 $V/V_2\simeq V_1$。(*HINT：根据直和的定义，对于* $\boldsymbol{v}\in V$，*有* $\boldsymbol{v}=\boldsymbol{v}_1+\boldsymbol{v}_2$，*我们可以设* $T_1:V\longrightarrow V_1$，$T_2:V\longrightarrow V_2$，*使得* $T_1(\boldsymbol{v})=\boldsymbol{v}_1;T_2(\boldsymbol{v})=\boldsymbol{v}_2$。)



2.4 矩阵的乘法与线性变换的复合

我在前面说过，线性变换就好比作用于线性空间上的函数。我们在本章第二节里面也详细学习了线性变换中所体现出的函数的性质。在本节中我们将重点讨论线性变换的复合的本质——矩阵的乘法。在谈论线性变换的复合之前，我们有必要首先引入矩阵的乘法运算，只有这样我们才能更好地学习线性变换的复合。

定义 2-10

设 A 为 $m\times n$ 矩阵，B 为 $k\times r$ 矩阵。当且仅当 $n=k$ 时，矩阵乘法 AB 有意义，并且其结果 $C=AB$ 为 $m\times r$ 矩阵。我们设 A 的行向量分别为 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m$; B 的列向量分别为 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r$。则在 C 中，有

$$C_{ij}=(\boldsymbol{v}_i)^T\cdot(\boldsymbol{u}_j)=(\boldsymbol{v}_i)\cdot(\boldsymbol{u}_j)^T\quad(1\le i\le m;1\le j\le r)$$

根据定义，对于 $C=AB$ 而言，C_{ij} 即为 A 中第 i 行与 B 中第 j 列取“数量积”的结果。我之所以对数量积加上了引号，是因为是因为此时两个向量的维数不同。我们需要先取其中一个向量的转置，然后再求数量积。因此我们注意到，若使得 AB 有意义，我们必须确保矩阵 A 的列数与矩阵 B 的行数相等。我们来通过实际的例子来熟悉矩阵的乘法运算：

例 2.4.1

设 $A=\begin{bmatrix}1&3\\2&-1\\1&1\end{bmatrix},B=\begin{bmatrix}1&0&2\\1&1&-1\end{bmatrix},C=\begin{bmatrix}1&2\\2&-1\end{bmatrix}$。那么在 AB,AC,BC 中，哪些运算有意义？在运算有意义的情况下其结果是多少？

在 AB 中，$A\in M_{32},B\in M_{23}$，因此运算结果有意义且 $AB\in M_{33}$。且

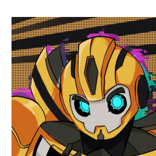
$$AB=\begin{bmatrix}(1,3)\cdot(1,1)&(1,3)\cdot(0,1)&(1,3)\cdot(2,-1)\\(2,-1)\cdot(1,1)&(2,-1)\cdot(0,1)&(2,-1)\cdot(2,-1)\\(1,1)\cdot(1,1)&(1,1)\cdot(0,1)&(1,1)\cdot(2,-1)\end{bmatrix}=\begin{bmatrix}4&3&-1\\1&-1&5\\2&1&1\end{bmatrix}$$

在 AC 中，$A\in M_{32},C\in M_{22}$，因此运算结果有意义且 $AC\in M_{32}$。且

$$AC=\begin{bmatrix}(1,3)\cdot(1,2)&(1,3)\cdot(2,-1)\\(2,-1\ \cdot(1,2)&(2,-1)\cdot(2,-1)\\(1,1)\cdot(1,2)&(1,1)\cdot(2,-1)\end{bmatrix}=\begin{bmatrix}5&-1\\0&5\\3&1\end{bmatrix}$$

在 BC 中，$B\in M_{23},C\in M_{22}$，其中 B 的列数同 C 的行数不相等，因此运算 BC 没有意义。

在定义 2-10 中，我们把向量之间的数量积和矩阵的乘法联系到了一起——这个意义重大，并且在计算机领域有着非常重要的应用。设想在大数据模型中，我们有 n 个 n 维向量，即 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\in\mathbb{R}^n$，然后我们要计算这一组向量中所有的数量积 $\boldsymbol{v}_i\cdot\boldsymbol{v}_j$。在计算机语言中，如果我们只是按照常规方法计算，即一一计算 $\boldsymbol{v}_1\boldsymbol{v}_1,\boldsymbol{v}_1\boldsymbol{v}_2\cdots$，这样一来程序就会变得十分复杂，准确来说其时间复杂度为 $\mathcal{O}(n^3)$（读者没有必要纠结 $\mathcal{O}(n^3)$ 代表什么意思，我们可以简单理解为 $\mathcal{O}(x)$ 中 x 越大，程序越复杂，效率越低。如果读者对这方面感兴趣不



妨阅读有关算法和数据结构的书籍)。于是我们仿照矩阵乘法，把这些向量分别按照行向量和列向量的形式写到矩阵中，然后计算矩阵的乘法。这样做的好处是：算法经过长时间的发展，计算矩阵乘法的算法已经趋于完善。在目前的模型中，矩阵乘法的时间复杂度介于 $\mathcal{O}(n^2)$ 与 $\mathcal{O}(n^3)$ 之间 *(在我编写这本书的时候，人类所能达到的极限值为 $\mathcal{O}(n^2\log n)$)*。当 n 趋于无穷大时，时间复杂度上的微小改变便会对算法整体的效率产生翻天覆地的变化。

矩阵的乘法还可以从矩阵与向量乘法的角度进行审视，下面的定理便是另一种求解矩阵乘法的方式：

定理 2-9

设 A, B 为矩阵，且 AB 有意义。设 $B = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix}$，则

$$AB = \begin{bmatrix} | & | & & | \\ A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_n \\ | & | & & | \end{bmatrix}$$

不难证明该计算法则与定义 2-10 中所提到的法则是等价的。

例 2.4.2

设矩阵 $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix}$，据此找出满足条件的矩阵 B，使得 $AB = C$。

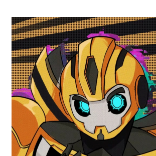
首先根据定义，我们知道 B 是 2×2 矩阵。我们不妨设 $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$，随后根据矩阵乘法的定义，我们有

$$\begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 2a-c & 2b-d \\ a & b \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix}$$

由此我们得出线性方程组 $\begin{cases} 2a-c=-1 \\ 2b-d=3 \\ a=1 \\ b=2 \end{cases}$，不难解得 $\begin{cases} a=1 \\ b=2 \\ c=3 \\ d=1 \end{cases}$。由此我们得到矩阵 $B = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$。

矩阵的加法运算较乘法运算而言更加简单，我们只需要把对应位置的元素相加即可。这也说明当且仅当矩阵 A, B 均为 $m \times n$ 矩阵时，其加法运算 $A + B$ 有意义，记 $C = A + B$，则 $(C)_{ij} = (A)_{ij} + (B)_{ij}$。

根据矩阵乘法运算的法则，如果 A, B 均为 $n \times n$ 矩阵，那么其乘积 AB, BA 也均为 $n \times n$ 矩阵。此时也会有非常多的性质与推论。我们重点研究的对象便是 $n \times n$ 矩阵之间的乘积。在所有的 $n \times n$ 矩阵中，有一种矩阵最为特殊：即由所有 $\mathbb{R}^n$ 的标准基底中的向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 为列向量所构成的矩阵 $I = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix}$。该矩阵我们也称作是**单位矩阵 (Identity Matrix)**，其中 $\boldsymbol{e}_i$ 的模均为 1，且在 $\boldsymbol{e}_i$ 中有且仅有第 i 个位置的



元素为 1，其余元素均为 0。并且 $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \begin{cases} 1(i=j) \\ 0(i \neq j) \end{cases}$。并且我们也知道 I 为 $n \times n$ 矩阵，全体 $n \times n$ 矩阵构成的集合我们写作 M_n，不难发现 M_n 同样构成线性空间。

定理 2-10
设 n 阶矩阵 I 单位矩阵，那么对于任意的 $A \in M_n$, $\boldsymbol{x} \in \mathbb{R}^n$, $AI = IA = A$; $I\boldsymbol{x} = \boldsymbol{x}$

另外矩阵之间还满足一定的运算法则：

定理 2-11
设存在矩阵 A, B, C 使得任意两个矩阵之间的加法，乘法均有意义，设 $\lambda \in \mathbb{R}$，$\boldsymbol{x}$ 为向量，则

① $A(BC) = (AB)C$

② $A(B+C) = AB + AC$

③ $(B+C)A = BA + CA$

④ $\lambda(AB) = (\lambda A)B = A(\lambda B)$

⑤ $A(B\boldsymbol{x}) = (AB)\boldsymbol{x}$ (在运算有意义时)

⑥ $(AB)^T = B^T A^T, (A+B)^T = A^T + B^T$

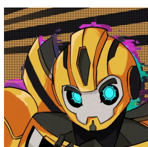
⑦ $(A^T)^T = A$

在定理 2-11 的第一条性质里面，我们因此知道矩阵乘法满足乘法的分配律，因此当涉及多个矩阵的乘法时，我们往往也可以将括号省略不写，比如我们可以直接写作 $ABCD$, 其代表 4 个矩阵的乘积，同时根据乘法的分配律我们有 $ABCD = (AB)(CD) = A(BCD) = (ABC)D = A(BC)D$ 等多种形式。值得注意的是我们只能改变矩阵运算顺序，然而却不能改变这些矩阵所处的位置，因为**根据矩阵乘法的定义，AB 与 BA 并不一定相等**，这一点读者可以自行验证。交换位置之后矩阵的乘法很可能就没有意义，或者结果发生改变。

例 2.4.3
设 A 为 $m \times n$ 矩阵，B 为 $n \times n$ 矩阵，满足 $AB = 0$ 且 A 不为零矩阵。那么是否存在 $n \times n$ 矩阵 C, 使得 $BC = I$?

我们假设这样的矩阵 C 存在。由于 $BC = I$, 根据矩阵乘法的定义，我们可以在等式两边同时乘以矩阵 A, 使得原式仍有意义。即 $A(BC) = A(I)$，根据定理 2-11 中的性质 1 以及定理 2-10，我们有 $(AB)C = A$, 因为我们知道 $AB = 0$，则 $(AB)C = 0$, 即 $A = 0$, 这一点与题设 $A \neq 0$ 矛盾，即不存在这样的矩阵 C, 使得 $BC = I$。

在矩阵乘法中，有一类乘法是我们重点研究的对象：即同为 $n \times n$ 矩阵之间的乘法。这样一来根据定义，



如果 $A, B \in M_n$，那么其乘积 AB, BA 必然存在且有意义（但不一定相等），且乘积所得到的新矩阵也为 $n \times n$ 矩阵，这样一来我们就可以将全体的 $n \times n$ 矩阵当成是线性空间 M_n 中的“向量”来看待，通过研究线性变换来研究矩阵的乘法。我会在本节的后半部分引入线性变换，让我们先把重点还放在研究矩阵的乘法上。我们先引出一条定义：

定义 2-11
我们称矩阵 A 为**对称矩阵** ***(Symmetric Matrix)***，当且仅当 $A = A^T$

我们设 A 为 $m \times n$ 矩阵，那么根据矩阵转置的定义，A^T 为 $n \times m$ 矩阵。若 $A = A^T$，我们则首先要保证二者的行数，列数相同，即 $m = n$。因此只有 $n \times n$ 矩阵存在对称矩阵。由全部 $n \times n$ 对称矩阵构成的集合我们也可以用 $\mathrm{Sym}(M_n)$ 表示。$\mathrm{Sym}(M_n)$ 也是一个线性空间，读者可以自行尝试证明。

例 2.4.4
设 $A, B \in \mathrm{Sym}(M_n)$，证明 $AB \in \mathrm{Sym}(M_n)$ 当且仅当 $AB = BA$。

证明. 我们先证充分性：假设 $AB \in \mathrm{Sym}(M_n)$，那么根据转置矩阵的性质我们有 $AB = (AB)^T$。另外我们也知道 $(AB)^T = B^T A^T$，即 $AB = B^T A^T$。再由 $A, B \in \mathrm{Sym}(M_n)$ 可知 $A = A^T; B = B^T$，因此 $AB = B^T A^T = BA$，充分性得证。

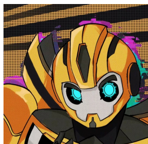
我们随后证必要性：假设 $AB = BA$，那么 $(AB)^T = (BA)^T$，即 $B^T A^T = (BA)^T$。再由 $A, B \in \mathrm{Sym}(M_n)$ 知，$B^T A^T = BA = (BA)^T$，即 $AB = (AB)^T$，故 $AB \in \mathrm{Sym}(M_n)$，必要性得证。 ■

对于 $n \times n$ 矩阵乘法而言，我们引入了**分块矩阵 (Block Matrix)**，即为把矩阵写成是若干个维数更小的 $m \times m$ 矩阵，我们通过研究这些 $m \times m$ 矩阵，也可以得出很多重要的性质和推论。我们假设

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{bmatrix}$$

我们发现，在矩阵 B 中有很多性质比较好的“局部”，比如有一片区域为零，同时左上角区域还存在一个单位矩阵 $I \in M_2$，因此我们可以把这些性质比较好的矩阵分割出来，像这样：

$$A = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{array}\right] = \begin{bmatrix} I_2 & \mathbf{0} \\ P & Q \end{bmatrix}$$



这样一来，我们通过用矩阵的方式来表示 A, 从而最大程度地简化了矩阵 A 的结构。假设此时我们有另外一个矩阵 B, 其中

$$B = \begin{bmatrix} 4 & -2 \\ 5 & 6 \\ 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{bmatrix}$$

那么在计算 AB 时，我们其实就可以把分块矩阵当成一个整体，然后按照常规的矩阵乘法进行计算。比如我们可以将 B 写成这样的分块矩阵：

$$B = \begin{bmatrix} 4 & -2 \\ 5 & 6 \\ -- & -- \\ 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}$$

在进行矩阵的分块时，我们一定要选取合适的位置。我们既想通过分块，使得原矩阵变成数个结构简单的矩阵，同时我们还要考虑每一个分块在与另一个矩阵的对应分块相乘时有意义。按照我们的划分，乘积 AB 即可以写作

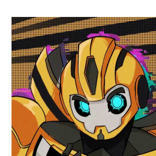
$$AB = \begin{bmatrix} I_2 & \mathbf{0} \\ P & Q \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} I_2X + \mathbf{0}Y \\ PX + QY \end{bmatrix} = \begin{bmatrix} X \\ PX + QY \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 5 & 6 \\ -- & -- \\ 30 & 8 \\ 8 & 27 \end{bmatrix}$$

所以在计算矩阵的乘法时，我们可以将矩阵划分成若干个分区，把每一个分区当成一个整体进行矩阵乘法。对于一种特殊类型的矩阵，存在下列定理：

定理 2-12
假设 $A = \begin{bmatrix} B & X \\ \mathbf{0} & C \end{bmatrix}, A_1 = \begin{bmatrix} B_1 & X_1 \\ \mathbf{0} & C_1 \end{bmatrix}$ 均为分块矩阵，其中 B, B_1 为同等大小的方阵（行数与列数相等）；C, C_1 为同等大小的方阵。那么

$$AA_1 = \begin{bmatrix} B & X \\ \mathbf{0} & C \end{bmatrix} \begin{bmatrix} B_1 & X_1 \\ \mathbf{0} & C_1 \end{bmatrix} = \begin{bmatrix} BB_1 & BX_1 + XC_1 \\ \mathbf{0} & CC_1 \end{bmatrix}$$

该定理的证明便是矩阵乘法定义的直接应用。定理 2-12 还可以引出一条推论：



定理 2-12 (a)

假设 $A=\begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ 为分块矩阵，其中 $I \in M_n$，则对于任意的正整数 k，满足 $A^k = A$。

定理 2-12 (a) 的证明

证明. 首先我们有 $A^2 = \begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$，根据定理 *2-12*，

$$A^2 = \begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} I^2 & IX + X\mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} I & X \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = A$$

因此，不难发现 $A^3 = AA^2 = A$，以此类推。那么我们即有 $A^k = A$。 ■

说了这么多矩阵的乘法，让我们回到线性变换，用线性变换的角度去审视矩阵的乘法。假设我们有线性变换 $T : V \longrightarrow W$, $S : W \longrightarrow U$。我们是否可以构造出一个新的线性变换 $Q : V \longrightarrow U$?

图 3: 线性变换 T, S 的直观图

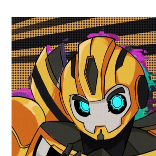
我们假设 $\boldsymbol{v} \in V$，那么通过变换 T，我们知道 $T(\boldsymbol{v}) \in W$，我们不妨设 $T(\boldsymbol{v}) = \boldsymbol{w}$，那么通过变换 S，我们知道 $S(\boldsymbol{w}) \in U$，我们不妨设 $S(\boldsymbol{w}) = \boldsymbol{u}$，然后我们再将 $\boldsymbol{w}$ 进行代换，即 $S(T(\boldsymbol{v})) = \boldsymbol{w}$，我们此时根据定理 2-11 中的推论得到，$S(T(\boldsymbol{v})) = ST(\boldsymbol{v})$，于是我们看到矩阵的乘法实际上也就是两个线性变换的复合。此时 $ST(\boldsymbol{v})$ 也可以记作 $S \circ T$。

定义 2-12

设线性变换 $T : V \longrightarrow W; S : W \longrightarrow U$，我们称线性变换 T 与 S 的**复合**（***Composite***）为 $ST : V \longrightarrow U$，且

$$ST(\boldsymbol{v}) = S(T(\boldsymbol{v}))$$

注意到该法则的顺序可能与我们的正常认知有所不同。在这里 ST 代表 “先 T” 后 “S”，简单来说可以理解为 “先内后外”。就好比复合函数 $f(g(x))$，我们需要先求出内层函数 $g(x)$ 的值，然后在求解外层函数 $f(x)$ 的值。此时 ST 便也相同，因为 ST 其实就是 $S(T(\boldsymbol{x}))$。引入线性变换之后，我们重新回到之前我提出的两个问题：第一，不是所有的矩阵之间都可以做乘法。我们以看图 3 中的变换为例，如果我们定义变换 TS，即



$T(S(\boldsymbol{w})), \boldsymbol{w} \in W$, 那么我们此时发现 $\boldsymbol{u} = S(\boldsymbol{w}) \in U$, 而 $\boldsymbol{u} = S(\boldsymbol{w}) \notin V$。因此这个时候 $T(\boldsymbol{u})$ 即没有意义，因为 $\boldsymbol{u}$ 不在其线性变换所定义的范围之内。在我们给出的使得矩阵 AB 有意义的条件里面，我强调了矩阵 A 的列数等于 B 的行数，现在我们能否用线性变换的知识去解答这一点？读者可以尝试证明；第二，我也说过就算是运算有意义的情况下，AB 也不一定与 BA 相等。为了加深对这一条的理解，我们来一起看下面这个例子：

例 2.4.5

在平面直角坐标系 $\mathbb{R}^2$ 中，选取其基底为标准基底 E。考虑两个线性变换：① $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, 该变换代表将向量 $\boldsymbol{v}$ 以原点为旋转中心顺时针旋转 $\pi/4$；② $S : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, 该变换代表将向量 $\boldsymbol{v}$ 正交投影至 x 坐标轴上。那么分别求出 ST, TS 并判断二者是否相等。

我们首先求出线性变换 T, S 在标准基底下所对应的矩阵。对于 T 而言，我们设其变换矩阵为 A, 有

$$A = \begin{bmatrix} T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) & T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

我们再求线性变换 S 在标准基底下所对应的矩阵。对于 S 而言，我们设其变换矩阵为 B, 有

$$B = \begin{bmatrix} S\left(\begin{bmatrix}1\\0\end{bmatrix}\right) & S\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

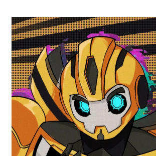
我们先求 TS。根据定义，TS 即 "先 S 后 T，其几何意义即代表先将向量正交投影至 x 轴，然后再绕原点顺时针旋转 $\pi/4$，其计算结果为

$$T(S(\boldsymbol{v})) = (TS)\boldsymbol{v} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix}\right)\boldsymbol{v} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}\boldsymbol{v}$$

我们再求 ST。根据定义，ST 即 "先 T 后 S，其几何意义即代表先将向量绕原点顺时针旋转 $\pi/4$，然后再正交投影至 x 轴，其计算结果为

$$S(T(\boldsymbol{v})) = (ST)\boldsymbol{v} = \left(\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)\boldsymbol{v} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix}\boldsymbol{v}$$

我们发现，$ST \neq TS$



2.4 练习

1. 假设你现在是游戏我的世界中的 Steve, 经过你勇敢地探索，你终于获得了钻石。但是目前的你只有两颗钻石，你在工作台上做出了一个钻石锄（如图 4 所示），随后陷入了漫长的沉思。你突然发现这个工作台同时也是一个 3×3 矩阵 $\mathcal{P}$，钻石和木棍都可以当成是矩阵里面的元素，其余地方（空格）位置的元素我们全部按照 0 来处理。这个时候，你收到了流浪商人送来的礼物，它是一个矩阵：$\mathcal{A}=\begin{bmatrix}1&0&1\\0&1&0\\0&0&0\end{bmatrix}$。你现在可以用你的“工作台矩阵 $\mathcal{P}$”去乘以 $\mathcal{A}$，你便有可能得到一个新的配方 $\mathcal{PA}$!

① 对于流浪商人送给你的“配方 $\mathcal{A}$”而言，这个新的配方对应下面的哪一个选项?

② 假设你有自制配方的能力，你现在可以根据你的喜好去定制属于你的 3×3 矩阵，但要求是你的配方中只能含有数字 0 或 1。那么仅使用图 4 中的钻石锄，以及数个不同的配方，你可以制造出多少种工具？制造出这些工具的配方分别是什么?

图 4: 钻石锄

(a) 钻石镐

(b) 钻石斧

(c) 钻石剑

(d) 钻石铲

2. 下列说法中，哪些是正确的？那些是错误的？如果是错误的，请找出一个反例或者说明原因。

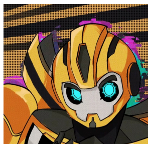
① $A^2=I$ 的充要条件是 $A=I$;

② $AJ=A$ 的充要条件是 $J=I$;

③ 设 A 为 $n\times n$ 矩阵，则 $(A^T)^3=(A^3)^T$

④ A 为对称矩阵的充要条件是 $A+I$ 为对称矩阵

⑤ 若 $A\neq 0$，那么 $A^2\neq 0$



3. 在 AB, BA 有意义的情况下，关于矩阵的乘法，下列哪些说法是正确的?

① 若 A 中有一行元素全部为 0，则 BA 也有一行元素全部为 0

② 若 A 中有一列元素全部为 0，则 BA 也有一列元素全部为 0

③ 若 AB 中有一行元素全部为 0，则 B 也有一行元素全部为 0

④ 若 AB 中有一行元素全部为 0，则 A 也有一行元素全部为 0

4. 在 3×3 矩阵中，找出一个矩阵 $\mathcal{A}$，使得 $\mathcal{A}\neq \mathbf{0}$, $\mathcal{A}^2=\mathbf{0}$。

5. 是否存在满足条件的 2×2 矩阵 A, B, 使得 $AB-BA=I$?

6. 我们称线性变换 $T: V \longrightarrow V$ 为**幂等变换 (Idempotent Transformation)**, 如果其满足 $T^2=T$。此时 T 对应的矩阵也被称作是幂等矩阵

① 证明矩阵 $\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$; $\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$ 为幂等矩阵。

② 在 $\mathbb{R}^3$ 中，T 代表向量到平面 $Ax+By+Cz=0$ 的正交投影 $(A,B,C\neq 0)$。那么对于任意的 A,B,C, 证明 T 为幂等变换。

③ 若 A 为幂等矩阵，证明 A^T 也为幂等矩阵

④ 若 P 为 $n\times n$ 幂等矩阵, 那么对于任意的 $n\times n$ 矩阵 A, $Q=P+AP-PAP$ 同样为幂等矩阵。

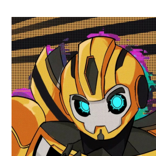
7. 我们称线性变换 $T: V \longrightarrow V$ 为**幂零变换 (Nilpotent Transformation)**, 如果其满足 $T^k=\mathbf{0}$ 其中 $k\leq \dim(V)$。设 $\dim(V)=n$。证明：此时一定有 $T^n=\mathbf{0}$。

8. 设线性变换 $T: V\longrightarrow U$, $S: U\longrightarrow W$。

① 若 ST 为单射变换，证明 T 为单射变换，并且 $\dim(V)\leq \dim(U)$。

② 若 ST 为满射变换，证明 S 为满射变换，并且 $\dim(W)\leq \dim(U)$。

9. 设 A 为 $p\times m$ 矩阵，B 为 $n\times q$ 矩阵，且 $mn=pq$。证明 $M_{mn}\simeq M_{pq}$



2.5 逆矩阵与线性变换的逆变换

让我们回到线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{y}$ 中，我们知道若 $\boldsymbol{x}$ 已知，则可以利用用矩阵与向量的乘法 $\mathcal{A}\boldsymbol{x}$ 来计算结果 $\boldsymbol{y}$。那么我们设想这种情况：若 $\boldsymbol{y}$ 已知，我们该怎么去求解 $\boldsymbol{x}$ 呢？我们可以试着构造出另一个矩阵 $\mathcal{B}$, 使得 $\mathcal{B}\boldsymbol{y}=\boldsymbol{x}$。那么在此情况下矩阵 $\mathcal{A},\mathcal{B}$ 之间满足什么关系？我们可以在等式 $\mathcal{B}\boldsymbol{y}=\boldsymbol{x}$ 两边同时乘以矩阵 $\mathcal{A}$, 得到 $\mathcal{A}(\mathcal{B}\boldsymbol{y})=\mathcal{A}\boldsymbol{x}=\boldsymbol{y}$。那么根据定义，我们有 $\mathcal{A}\mathcal{B}=I$；同理我们也可以在等式 $\mathcal{A}\boldsymbol{x}=\boldsymbol{y}$ 两边同时乘以矩阵 $\mathcal{B}$，那么有 $\mathcal{B}(\mathcal{A}\boldsymbol{x})=\mathcal{B}\boldsymbol{y}=\boldsymbol{x}$。根据定义我们有 $\mathcal{B}\mathcal{A}=I$, 此时我们发现 $\mathcal{A}\mathcal{B},\mathcal{B}\mathcal{A}$ 均有意义，代表 $\mathcal{A},\mathcal{B}\in M_n$，那么此时我们就知道 $\mathcal{A}\mathcal{B}=\mathcal{B}\mathcal{A}=I$, 在此情况下，我们可以把 $\mathcal{B}$ 写作 $\mathcal{A}^{-1}$, 代表矩阵 $\mathcal{A}$ 的**逆矩阵 (Inverse Matrix)**；同理可以把 $\mathcal{A}$ 写作 $\mathcal{B}^{-1}$, 代表矩阵 $\mathcal{B}$ 的**逆矩阵 (Inverse Matrix)**。我们由此引入逆矩阵的概念：

定义 2-13
我们称 $n\times n$ 矩阵 A 可逆，当且仅当存在 $n\times n$ 矩阵 B，使得 $AB=BA=I$。此时我们称 B 为 A 的逆矩阵，通常也可以记作 A^{-1}。

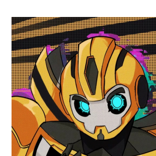
例 2.5.1
设 $B=\begin{bmatrix}-1&1\\1&0\end{bmatrix},A=\begin{bmatrix}0&1\\1&1\end{bmatrix}$, 证明 A 的逆矩阵为 B。

证明. 我们需要验证两项乘积 $AB=BA=I$, 其中 $AB=\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}-1&1\\1&0\end{bmatrix}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$ 且 $BA=\begin{bmatrix}-1&1\\1&0\end{bmatrix}\begin{bmatrix}0&1\\1&1\end{bmatrix}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$。因此根据逆矩阵的定义，我们称 A 的逆矩阵为 B。 ■

我们因此不难发现：若 A 为 B 的逆矩阵，则 B 为 A 的逆矩阵，即逆矩阵满足对称性。有兴趣的读者可以尝试判断逆矩阵是否为一个等价关系，即验证逆矩阵之间是否满足自反性和传递性。

例 2.5.2
证明 $C=\begin{bmatrix}0&0\\1&3\end{bmatrix}$ 不存在逆矩阵。

证明. 我们采用假设法，即假设 C 的逆矩阵为 $B=\begin{bmatrix}a&b\\c&d\end{bmatrix}$, 随后计算 CB 的结果为 $CB=\begin{bmatrix}0&0\\1&3\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}0&0\\a+3c&b+3d\end{bmatrix}$ 不难发现，$(CB)_{11}=0$, 因此无论 a,b,c,d 的取值，CB 不可能为单位矩阵 I, 因此矩阵 C 不存在逆矩阵。 ■



所以我们会想：什么情况下一个矩阵存在逆矩阵呢？首先我们可以确认，**有且仅有 $n \times n$ 矩阵存在逆矩阵**。那么在所有的 $n \times n$ 矩阵中，究竟要满足什么样的性质才能够使得矩阵可逆呢？我们此时不妨从线性变换的角度去研究这个问题。我们知道。每一个矩阵都对应着一个线性变换 $\mathcal{T}$，在选定基底 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 的情况下，该线性变换对应的矩阵即为

$$[\mathcal{T}]_\beta = \begin{bmatrix} | & | & & | \\ [\mathcal{T}(\boldsymbol{v}_1)]_\beta & [\mathcal{T}(\boldsymbol{v}_2)]_\beta & \cdots & [\mathcal{T}(\boldsymbol{v}_n)]_\beta \\ | & | & & | \end{bmatrix}$$

由于我们知道有且仅有 $n \times n$ 矩阵存在逆矩阵，所以线性变换前后所处的线性空间维数不变。为了简化考虑，我们不妨设这样满足条件的线性变换 $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{V}$。随后我们来回顾一下线性变换的函数性质（2.2），我们先看下面几种变换类型：

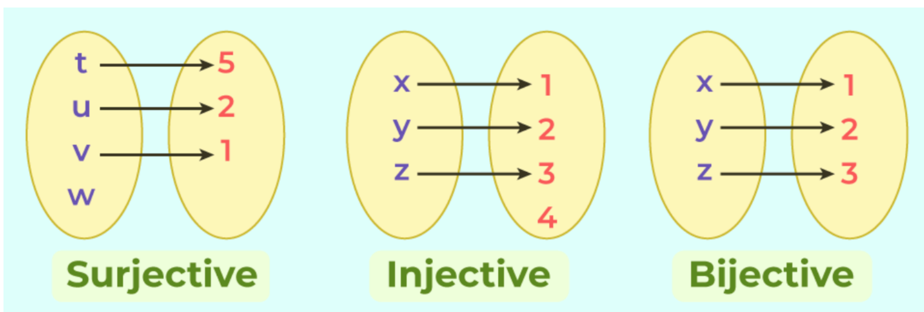


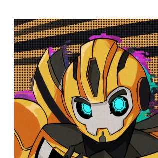
图 6: 本图中从左至右分别代表满射；单射；双射

对于上图中的三种变换，我们来一一研究：先看最左边的图像，很明显该图像对应了一个满射。我们设这个变换为 f，那么我们有 $f(t) = 5; f(u) = 2; f(v) = 1$。但我们发现了一个问题：$f(w)$ 此时没有意义，这样一来这个变换也就自然而言不存在逆变换。因为我们无法找到二者之间“一一对应”的关系（反函数没有意义）。根据函数的定义，对于给定的 x 而言 $f(x)$ 有且仅有一个确定值；对于线性变换也是如此：即对于给定的 $\boldsymbol{x}$ 而言，$\mathcal{T}(\boldsymbol{x})$ 有且仅有一个确定值。对于图 6 中的第二个例子也是同样的问题（函数没有意义），我们无法找到“一一对应”的关系。对于双射的例子而言，由于定义域和值域元素数目相同，因此其函数与反函数都可以找到一一对应的关系。在线性变换中我们所得到的结论相同。

定理 2-13

设 $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{V}$ 为线性变换，则 $\mathcal{T}$ 可逆的充要条件是 $\mathcal{T}$ 为双射。此时我们记 $\mathcal{T}$ 的逆变换为 $\mathcal{T}^{-1}$，满足 $\mathcal{T}\mathcal{T}^{-1} = \mathcal{T}^{-1}\mathcal{T} = I_d$ (单位变换)。设 $\mathcal{T}$ 对应的矩阵形式为 $\mathcal{A}$，那么 $\mathcal{T}^{-1}$ 对应的矩阵形式为 $\mathcal{A}^{-1}$。

我们同时也知道，此时由于线性变换前后所对应的线性空间的维数不发生改变，因此只要得出单射或满射，便可以直接得出该变换为双射。而我们知道，满射即代表矩阵 $[T]_\beta$ 中的列向量可以构成 $\mathcal{T}$ 的一组基底，再结合单射代表列向量彼此线性无关，我们可以再度由行秩定理得到该矩阵的行向量也彼此线性无关，等等。也就是说，有非常多种方法来确定一个线性变换（矩阵）是否可逆，同时矩阵可逆也同时告诉了我们非常多的潜在结论。我们把这些性质总结成为一条定理：



定理 2-14
设 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 为线性变换，记 $[\mathcal{T}]_\beta$ 为该变换的矩阵形式，则下列结论等价（即得出下列结论中的任意一条便可以推出剩余所有结论）：

① $\mathcal{T}$ 为单射

② $\mathcal{T}$ 为满射

③ $\mathcal{T}$ 为双射

④ 线性变换 $\mathcal{T}$ 可逆

⑤ 矩阵 $[\mathcal{T}]_\beta$ 可逆

⑥ $([\mathcal{T}]_\beta)^T$ 可逆

⑦ 矩阵 $[\mathcal{T}]_\beta$ 满秩

⑧ 矩阵 $([\mathcal{T}]_\beta)^T$ 满秩

定理 2-14 其实还可以扩充更多的结论，限于篇幅原因我将省略部分推论。在这里强烈建议读者结合目前所学的知识，尝试去扩充这些结论。比如在线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 中，矩阵可逆是否和解的结构有关系？我们此时在将目光重点放在可逆矩阵上，我们还能引出很多有关可逆矩阵的性质：

定理 2-15
对于 $n\times n$ 可逆矩阵而言，我们有如下的性质：

① 单位矩阵 I 可逆，且 $I^{-1}=I$

② $(A^{-1})^{-1}=A$

③ $(AB)^{-1}=B^{-1}A^{-1}$。一般地，$(A_1A_2\cdots A_k)^{-1}=A_k^{-1}\cdots A_2^{-1}A_1^{-1}$

④ $(A^k)^{-1}=(A^{-1})^k$

⑤ 设 λ 为非零常数，则 $(\lambda A)^{-1}=\dfrac{1}{\lambda}A^{-1}$

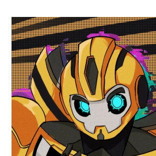
⑥ $(A^T)^{-1}=(A^{-1})^T$

再次强烈建议读者尝试对定理 2-15 中的结论进行证明。此处还是由于篇幅原因，我将不做证明。

例 2.5.3

设 A 为 $n\times n$ 可逆矩阵，若 $A^3-3A+2I=\mathbf{0}$，据此求 A^{-1}。

在这种类型的题里面，我们往往需要运用矩阵的性质进行“配凑”，比如我们在本题中使用 $I=AIA^{-1}$ 这一公式，那么原式即为 $A^3-3A+2AIA^{-1}=\mathbf{0}$，即 $A(A^2-3I+2IA^{-1})=\mathbf{0}$，因为矩阵 A 可逆，所以 $A\neq 0$，则 $A^2-3I+2A^{-1}=\mathbf{0}$，由此我们求得 $A^{-1}=\dfrac{1}{2}(3I-A^2)$。



例 2.5.4

设 A 为 $n\times n$ 矩阵（不一定可逆），若 $A^3=\mathbf{0}$, 证明 $(I-A)^{-1}=I+A+A^2$。

证明. 此时由于矩阵 A 不一定可逆，所以我们不能用含有 A^{-1} 的式子代换，但我们由题意可知，$A^3=\mathbf{0}$, 即 $I-A^3=I$, 即 $I^3-A^3=I$, 我们随后便可以使用立方差公式，得到 $(I-A)(A^2+A+I)=(A^2+A+I)(I-A)=I$, 然后根据逆矩阵的定义，我们就有 $(I-A)^{-1}=A^2+A+I$。

■

定义 2-14
我们称 $n\times n$ 矩阵 A 为**自逆矩阵** ***(Self Inverse Matrix)***，若 $A^2=I$, 即 $A^{-1}=A$。

根据这个定义，我们也不难发现所有的单位矩阵 $I\in M_n$ 均为自逆矩阵。

例 2.5.5

设 A 为 $m\times n$ 矩阵, 且 $X^TX=I_n$。证明 I_m-2XX^T 为自逆矩阵, 同时也为对称矩阵。

证明. 欲证 I_m-2XX^T 自逆，那么根据定义，我们只需证 $(I_m-2XX^T)^2=I_m$, 即

$$I_m^2-2I_m(2XX^T)+(2XX^T)^2=I_m$$

随后根据矩阵乘法的性质，我们得到 $I_m-2(2XX^T)+4(XX^T)(XX^T)=I_m$，即为 $I_m-4(XX^T)+4X\ X^TX\ X^T=I_m$，根据定义，即 $I_m-4XX^T+4XI_nX^T=I_m-4XX^T+4XX^T=I_m$, 那么不难发现，等式左边等于等式右边，即 I_n-2XX^T 为自逆矩阵。由于我们知道对于逆矩阵 $\mathcal{P}$ 而言，$(\mathcal{P}^{-1})^T=(\mathcal{P}^T)^{-1}$, 那么我们有

$$((I_n-2XX^T)^{-1})^T=((I_n-2XX^T)^T)^{-1}$$

根据矩阵的性质我们知道

$$(I_n-2XX^T)^T=(I_n)^T-(2XX^T)^T=I_n-2(X^T)^T(X)^T=I_n-2XX^T$$

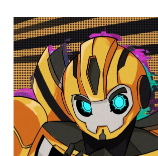
那么根据定义，I_n-2XX^T 也为对称矩阵。。

■

随后，让我们回到线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 中，我们假设矩阵 $\mathcal{A}$ 存在逆矩阵，那么我们便可以在等式的两边同时乘以 $\mathcal{A}^{-1}$, 我们进而得到

$$\boldsymbol{x}=\mathcal{A}^{-1}\boldsymbol{b}$$

其中 $\boldsymbol{x}$ 便是该方程组的唯一解。我们可以仿照 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 的形式，去判断或者求解一个矩阵的逆矩阵。在



线性方程组 $\mathcal{A}\boldsymbol{x}=\boldsymbol{b}$ 中，我们可以把 $\boldsymbol{x}$ 看作是某 $\mathcal{A}$ 的逆矩阵，把 $\boldsymbol{b}$ 看作单位矩阵 I。这样，我们便可以用研究线性方程组的方法来研究逆矩阵了。

定理 2-16

若 A 为 $n\times n$ 可逆矩阵，那么在矩阵 A 中，我们一定可以通过一系列的高斯消元，从而得到单位矩阵 I。从 I 再到 A 的过程也是高斯消元的逆运算。通过对增广矩阵 $\left[A \;\middle|\; I\right]$ 进行高斯消元，我们最终可以得到 $\left[I \;\middle|\; A^{-1}\right]$，其中 A^{-1} 便是矩阵 A 的逆矩阵。

例 2.5.6

求出矩阵 $A=\begin{bmatrix}2&7&1\\1&4&-1\\1&3&0\end{bmatrix}$ 的逆矩阵。

我们只需求解 $\left[A \;\middle|\; I\right]$ 即可，在本题中我们有

$$\left[A \;\middle|\; I\right]=\left[\begin{array}{ccc|ccc}2&7&1&1&0&0\\1&4&-1&0&1&0\\1&3&0&0&0&1\end{array}\right]$$

我们只需对 A 中的 3 列进行高斯消元即可。

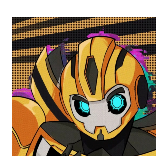
$$\left[\begin{array}{ccc|ccc}2&7&1&1&0&0\\1&4&-1&0&1&0\\1&3&0&0&0&1\end{array}\right]\xrightarrow{2R_3-R_1;2R_2-R_1}\left[\begin{array}{ccc|ccc}2&7&1&1&0&0\\0&1&-3&-1&2&0\\0&-1&-1&-1&0&2\end{array}\right]$$

$$\left[\begin{array}{ccc|ccc}2&7&1&1&0&0\\0&1&-3&-1&2&0\\0&-1&-1&-1&0&2\end{array}\right]\xrightarrow{2R_1-7R_2;R_3+R_2;R1/2;R3/2}\left[\begin{array}{ccc|ccc}1&0&11&4&-7&0\\0&1&-3&-1&2&0\\0&0&-2&-1&1&1\end{array}\right]$$

$$\left[\begin{array}{ccc|ccc}1&0&11&4&-7&0\\0&1&-3&-1&2&0\\0&0&-2&-1&1&1\end{array}\right]\xrightarrow{2R_1+11R_3;2R_2-3R_3;R_1/2}\left[\begin{array}{ccc|ccc}1&0&0&-1.5&-1.5&5.5\\0&1&0&0.5&0.5&-1.5\\0&0&1&-0.5&0.5&-0.5\end{array}\right]$$

因此，从我们最终得出的 $\left[A \;\middle|\; I\right]$ 中，我们得到 $A^{-1}=\begin{bmatrix}-1.5&-1.5&5.5\\0.5&0.5&-1.5\\-0.5&0.5&-0.5\end{bmatrix}$。

需要强调的一点是，当我们使用 $\left[A \;\middle|\; I\right]$ 进行消元时，A 矩阵应可逆，否则我们无法得出形如 $\left[I \;\middle|\; A^{-1}\right]$ 的形式。所以这种方法其实也是判断矩阵 A 是否可逆的方法。这种方法的优点是我们不仅可以判断出矩阵 A 是否可逆，当矩阵可逆时我们还可以直接求出 A 的逆矩阵。当我们完成下一节对行列式的学习之后，我们将



给出逆矩阵的另外一种求法。

我们再通过一道例题来学习分块矩阵中的可逆关系。

例 2.5.7

设 $P=\begin{bmatrix}A & X\\ \mathbf{0} & B\end{bmatrix}$ 和 $Q=\begin{bmatrix}A & \mathbf{0}\\ Y & B\end{bmatrix}$ 为分块矩阵，其中 $A\in M_m$, $B\in M_n$。证明:

① 若 P 为可逆矩阵，则 A,B 均为可逆矩阵，且 $P^{-1}=\begin{bmatrix}A^{-1} & -(A^{-1}XB^{-1})\\ \mathbf{0} & B^{-1}\end{bmatrix}$。

② 若 Q 为可逆矩阵，则 A,B 均为可逆矩阵，且 $Q^{-1}=\begin{bmatrix}A^{-1} & \mathbf{0}\\ -(B^{-1}YA^{-1}) & B^{-1}\end{bmatrix}$。

不难发现，①,②的结构十分相似，我将仅证①，第二问的证明留给读者思考。

证明. 我们先证此时 A,B 均为可逆矩阵：首先我们不难发现 P 为 $(n+m)\times(n+m)$ 矩阵，由于 P 可逆，根据逆矩阵的性质我们知道 P 的行向量彼此线性无关；列向量彼此线性无关。由列向量彼此线性无关我们可知 A 中列向量一定线性无关，又因为 A 为 $m\times m$ 矩阵，则 A 可逆；同理由行向量线性无关可以得出 B 中行向量线性无关，则 B 可逆。

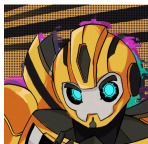
我们可以使用分块矩阵的乘法规则，此时我们只需要验证 $PP^{-1}=P^{-1}P=I_{m+n}$

$$\begin{aligned}PP^{-1}=\begin{bmatrix}A & X\\ \mathbf{0} & B\end{bmatrix}\begin{bmatrix}A^{-1} & -(A^{-1}XB^{-1})\\ \mathbf{0} & B^{-1}\end{bmatrix}&=\begin{bmatrix}AA^{-1}+\mathbf{0} & A(-A^{-1}XB^{-1})+XB^{-1}\\ \mathbf{0}+\mathbf{0} & \mathbf{0}+BB^{-1}\end{bmatrix}\\&=\begin{bmatrix}I_m & -(AA^{-1})(XB^{-1})+XB^{-1}\\ \mathbf{0} & I_n\end{bmatrix}\\&=\begin{bmatrix}I_m & \mathbf{0}\\ \mathbf{0} & I_n\end{bmatrix}\end{aligned}$$

$$\begin{aligned}P^{-1}P=\begin{bmatrix}A^{-1} & -(A^{-1}XB^{-1})\\ \mathbf{0} & B^{-1}\end{bmatrix}\begin{bmatrix}A & X\\ \mathbf{0} & B\end{bmatrix}&=\begin{bmatrix}A^{-1}A+\mathbf{0} & A^{-1}X-(A^{-1}XB^{-1})B\\ \mathbf{0}+\mathbf{0} & \mathbf{0}+B^{-1}B\end{bmatrix}\\&=\begin{bmatrix}I_m & A^{-1}X-(A^{-1}X)(B^{-1}B)\\ \mathbf{0}+\mathbf{0} & I_n\end{bmatrix}\\&=\begin{bmatrix}I_m & \mathbf{0}\\ \mathbf{0} & I_n\end{bmatrix}\end{aligned}$$

因此，$PP^{-1}=P^{-1}P=I_{m+n}$，原式得证。 ■

在一些时候，我们还可以从线性变换的角度入手来判断矩阵是否可逆。比如 2.4 的例 2.4.5，我们用“作图”的方式证明了矩阵乘法中 AB 与 BA 不一定相等。我们同样可以把矩阵当成线性变换，从线性变换的角度有时可以快速判断矩阵是否可逆。



例 2.5.8

判断矩阵 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 是否可逆，如果可逆的话其几何意义是什么?

我们如果考虑线性变换 $\mathcal{Q}: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, 那么在标准基底 E 下，我们设 $[\mathcal{Q}]_E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$。对于任意的 $\boldsymbol{v} = \begin{bmatrix} x \\ y \end{bmatrix}$, 我们有 $\mathcal{Q}\boldsymbol{v} = [\mathcal{Q}]_E\boldsymbol{v} = \begin{bmatrix} y \\ x \end{bmatrix}$, 那么经过观察，$\mathcal{Q}\boldsymbol{v}$ 便是 $\boldsymbol{v}$ 关于 $y = x$ 对称的向量（如图)。

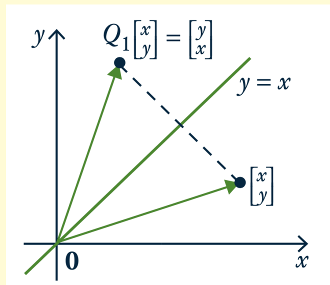


图 7: 线性变换 $\mathcal{Q}$ 的几何意义

所以，这里面便天然存在了“一一对应”的关系，所以我们说线性变换 $\mathcal{Q}$ 可逆，即为矩阵 A 可逆。

这里面也留给读者一个思考：为什么所有的投影变换都不可逆?

2.5 练习

1. 求出下列矩阵的逆矩阵：

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & 5 & -1 \end{bmatrix}; B = \begin{bmatrix} -1 & 4 & 5 & 2 \\ 0 & 0 & 0 & -1 \\ 1 & -2 & -2 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}; C = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

2. 若 $(A^T - 2I_2)^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$, 求 A。

3. 设 A, B 均为 $n \times n$ 矩阵，下列哪些说法是正确的？

① 若 $A \neq \mathbf{0}$, 则 A 可逆；

② 若 A, B 均可逆，则 $A+B$ 可逆；

③ 若 A, B 均可逆，则 $(A^{-1}B)^T$ 可逆；

④ 若 $A^4 = 3I$, 则 A 可逆；

⑤ 若 A^2 可逆, 则 A 可逆；

⑥ 若 A 可逆且**斜对称 (Skew Symmetric)**(即 $A^T = -A$)，则 A^{-1} 可逆且斜对称；

⑦ 若 $AB = B$ 且 $B \neq \mathbf{0}$, 则 A 可逆。

4. 设 $\mathcal{P} = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$, 求 $\mathcal{P}\mathcal{P}^T$ 及 $\mathcal{P}^{-1}$ $(\mathcal{P} \neq \mathbf{0})$。

5. 设 A 为 $n \times n$ 矩阵，若 A 中每一行的全部元素加和均为 0，证明：A 不可逆。

6. 设 P 为 $n \times n$ 矩阵，满足 $P^2 = P$。证明:

① $I - 2P$ 为自逆矩阵；

② $I - aP$ 可逆当且仅当 $a \neq 1$，并验证 $(I - aP)^{-1} = I + \left(\dfrac{a}{1-a}\right)P$。

7. 若矩阵满足 $U^2 = I$，证明：$I + U$ 不可逆，除非 $U = I$。

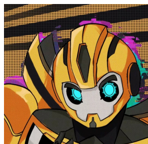
8. 设 A, B 均为 $n \times n$ 矩阵, 证明:

① $A(I + BA) = (I + AB)A$, 且 $(I + BA)B = B(I + AB)$；

② 若 $I + AB$ 可逆，则 $I + BA$ 也可逆，且 $(I + BA)^{-1} = I - B(I + AB)^{-1}A$。

9. 设 A 为 $n \times n$ 矩阵，若 $A^n = \mathbf{0}$, 求 $(I - A)^{-1}$。

10. 设 $n \times n$ 矩阵 J 中的每一个元素均为 1 (即 $(J)_{ij} = 1$ $(1 \leq i, j \leq n)$), 若 $(I - \lambda J)$ 为自逆矩阵和对称矩阵, 求 λ。(用含有 n 的式子表示)



2.6 行列式变换

在本节中，我们将研究一种特殊的线性变换：**行列式 (Determinant)**。简单来说，行列式是一个线性变换 $\mathcal{T}: M_n \longrightarrow \mathbb{R}$, 在这个线性变换（即行列式）中，我们所研究的对象是 $n \times n$ 矩阵，意味着我们将不讨论三行四列矩阵的行列式是什么。同时根据我们所定义的线性变换，行列式的取值（$\mathcal{T}(M)$）为全体实数（在目前阶段我们只研究行列式为实数的情况，行列式为复数的我暂时不做涉及）。实际上行列式还是一个满射变换（但不是单射），这一点在我们了解更多有关行列式的定义之后便能够明白原因。

如果读者尝试过第一章第二节的课后习题（1.2 练习题 7），读者便会明白行列式是一个**多线性映射 (Multi-linear)**, 我们不妨给出多线性映射的定义：

定义 2-14
我们称定义在矩阵 $\mathcal{A} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ 上的变换 $\mathcal{P}$ 为一个多线性映射，如果对于任意的 $\boldsymbol{v}_i$ 以及与 $\boldsymbol{v}_i$ 同维数的向量 $\boldsymbol{u}$ 以及常数 λ, μ，有

$$\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \cdots & \lambda\boldsymbol{v}_i + \mu\boldsymbol{u} & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) = \lambda\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_i & \cdots \boldsymbol{v}_n \end{bmatrix}\right) + \mu\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{u} & \cdots & \boldsymbol{v}_n \end{bmatrix}\right)$$

我们可以简单地理解为矩阵中的每一列都满足线性变换的性质，且不同列之间彼此互不干扰。那么我们便可以得出多线性映射里面的一条定理：

定理 2-17
设 $\mathcal{P}$ 为定义在矩阵 $\mathcal{A}$ 上的多线性映射，若 $\mathcal{A}$ 中存在任意一列元素全部为零，那么 $\mathcal{P}(\mathcal{A}) = 0$。

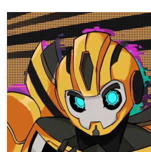
定理 2-17 的证明：

证明. 为了简化运算，且不失一般性，我们不妨设矩阵 $\mathcal{A} = \begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$, 即假设矩阵 $\mathcal{A}$ 的第一列全部为零。那么根据多线性映射还有零向量的性质，我们有

$$\begin{aligned} \mathcal{P}\left(\begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) &= \mathcal{P}\left(\begin{bmatrix} \mathbf{0}+\mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) \\ &= \mathcal{P}\left(\begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) + \mathcal{P}\left(\begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) \end{aligned}$$

那么不难发现，我们有 $\mathcal{P}\left(\begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) = 2\mathcal{P}\left(\begin{bmatrix} \mathbf{0} & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right)$, 因此 $\mathcal{P}(\mathcal{A}) = 0$。 ■

由定义 2-14 还可以引出多线性映射的另外一条性质：



定理 2-18
设 $\mathcal{P}$ 为定义在 $n \times n$ 矩阵 $\mathcal{A}$ 上的多线性映射，则对于任意的非零常数 λ, $\mathcal{P}(\lambda\mathcal{A}) = \lambda^n\mathcal{P}(\mathcal{A})$。

定理 2-17 的证明：

证明. 我们不妨设 $\mathcal{A} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$, 那么根据定义，$\lambda\mathcal{A} = \begin{bmatrix} \lambda\boldsymbol{v}_1 & \lambda\boldsymbol{v}_2 & \cdots & \lambda\boldsymbol{v}_n \end{bmatrix}$, 根据多线性映射的定义，我们有

$$\begin{aligned}\mathcal{P}\left(\begin{bmatrix} \lambda\boldsymbol{v}_1 & \lambda\boldsymbol{v}_2 & \cdots & \lambda\boldsymbol{v}_n \end{bmatrix}\right) &= \lambda\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \lambda\boldsymbol{v}_2 & \cdots & \lambda\boldsymbol{v}_n \end{bmatrix}\right) \\ &= \lambda \cdot \lambda\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \lambda\boldsymbol{v}_n \end{bmatrix}\right) \\ &= \cdots\cdots \\ &= \lambda^n\mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right)\end{aligned}$$

即 $\mathcal{P}(\lambda\mathcal{A}) = \lambda^n\mathcal{P}(\mathcal{A})$。 ■

在所有的多线性映射中，我们再定义一种特殊的多线性映射，我们称作为**交替多线性映射 (Alternating Multi-linear)**。其定义如下：

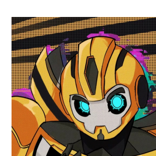
定义 2-14
我们称多线性映射 $\mathcal{P}$ 为交替多线性映射，如果当 $n \times n$ 矩阵 $\mathcal{A}$ 中含有任意两列完全相同的元素时，$\mathcal{P}(\mathcal{A}) = 0$。

根据这个定义，我们还能得出交替多线性映射的一条性质：

定理 2-18
设 $\mathcal{P}$ 为交替多线性映射，对于任意的 $n \times n$ 矩阵 $\mathcal{A}$, 我们交换矩阵 $\mathcal{A}$ 中的任意两列，记交换之后的矩阵为 $\mathcal{B}$, 则 $\mathcal{P}(\mathcal{A}) = -\mathcal{P}(\mathcal{B})$。

定理 2-18 的证明：

证明. 同样地，为了简化运算以及不失一般性，我们假设交换 $\mathcal{A}$ 的第一，二列。即 $\mathcal{A} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$；$\mathcal{B} = \begin{bmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$。此时我们考虑 $\delta = \mathcal{P}\left(\begin{bmatrix} \boldsymbol{v}_1 + \boldsymbol{v}_2 & \boldsymbol{v}_1 + \boldsymbol{v}_2 & \boldsymbol{v}_3 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right)$。根据定义 *2-14*，我们知道此时 $\delta = 0$, 再结合多线性映射的性质，我们有 $\delta = \mathcal{P}[\boldsymbol{v}_1, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n] + \mathcal{P}[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n] + \mathcal{P}[\boldsymbol{v}_2, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n] + \mathcal{P}[\boldsymbol{v}_2, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n] = 0$, 因此不难发现 $\mathcal{P}(\mathcal{A}) = -\mathcal{P}(\mathcal{B})$。 ■



在所有的交替多线性映射中，我们将会重点研究其中一种——**行列式 (Determinant)**，我们先一起看一下行列式的定义：

定义 2-15
假设我们存在一种交替多线性映射 $\det: M_n \longrightarrow \mathbb{R}$，使得 $\det(I_n) = 1$，我们称该多线性多线性映射 det 为行列式。通常用 $\det(A)$ 来表示矩阵 A 的行列式。

我们现在已经知道了行列式的定义，那么我们该如何去求解一个具体的矩阵的行列式呢？我们在此之前先引入一些有关排列组合的知识：假设我们现在有 n 个元素：$1, 2, 3, \cdots, n$ 构成的集合 T_n，我们定义一种作用于这 n 个元素上面的双射 π，使得对任意的 $m \in T_n, \pi(m) \in T_n$。由于 π 为双射，所以我们知道当 $m \neq n$ 时，$\pi(m) \neq \pi(n)$, 且对于任意的 $r \in T_n$，都存在相应的 $t \in T_n$, 使得 $\pi(t) = r$。满足这种条件的映射 π 被我们称作是关于 1 ——n 的一个**排列（Permutation)**。我们很多时候也用一个 $2 \times n$ 矩阵 π 来表示一个排列，在矩阵的第 i 列里面，$\pi_{1i} = i; \pi_{2i} = \pi(i)$。下面便是一个当 $n = 5$ 时的例子：

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \pi(1) & \pi(2) & \pi(3) & \pi(4) & \pi(5) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix}$$

那么对于给定的 n 而言，我们把所有满足条件的不同排列 π 所构成的集合称作是关于 n 的一个**排列群 (Permutation Group)**，通常用 S_n 表示。

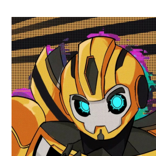
定理 2-19
对于排列群 S_n 而言，$|S_n| = n!$。

该定理的证明显而易见，再次不多赘述。我们此时需要引入一个至关重要的概念，即排列的**符号 (Sign)**:

定义 2-16
设排列 $\pi \in S_n$, 我们定义排列 π 的符号为 $\mathrm{sgn}(\pi) = (-1)^k$, 其中 k 代表在排列 π 中, 所有满足 $\begin{cases} i < j \\ \pi(i) > \pi(j) \end{cases}$ 的有序数对 (i, j) 的数量。

我们通过定义 2-16，一起来看一下上文提到的排列 $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix}$ 的符号。首先我们知道 $\pi(1) = 4; \pi(2) = 2; \pi(3) = 3; \pi(4) = 5; \pi(5) = 1$, 那么我们先固定 $i = 1$，我们发现:$1 < 2, \pi(1) > \pi(2); 1 < 3, \pi(1) > \pi(3); 1 < 5, \pi(1) > \pi(5)$。对于元素 1 而言有三组满足条件的数对；然后再看元素 2，我们发现：$2 < 5; \pi(2) > \pi(5)$，只有一组满足条件的数对；对于元素 3: $3 < 5; \pi(3) > \pi(5)$，有一对满足条件的数对；对于元素 4: $4 < 5, \pi(4) > \pi(5)$，有一对满足条件的数对。对于元素 5 而言，由于 5 是最大的元素，因此不存在 j 使得 $5 < j$，所以有 0 对满足条件的数对。那么综合起来一共有 6 组满足条件的数对，此时排列 π 的符号即为 $\mathrm{sgn}(\pi) = (-1)^6 = 1$。

通过排列符号的定义，我们不难发现排列的符号 $\mathrm{sgn}(\pi)$ 的取值为 1 或 -1。在 $n \times n$ 矩阵中，我们便可



以把行列式看作是一种矩阵元素之间的排列。我们下面给出计算行列式的其中一个公式：

定理 2-20

设 A 为 $n\times n$ 矩阵，其中 (a_{ij}) 代表矩阵 A 中第 i 行，第 j 列位置的元素。那么矩阵 A 的行列式为：

$$\det(A)=\sum_{\pi\in S_n}\left(\operatorname{sgn}(\pi)\prod_{i=1}^{n}(a_{\pi(i)\,i})\right)$$

例 2.6.1：

计算二阶矩阵 $\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}$ 的行列式。

首先根据定义，我们要找出排列群 S_2 里面的全部元素。此时有且仅有两种排列，即

$$\pi_1=\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix};\pi_2=\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$$

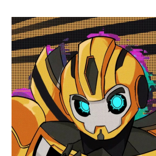
那么，在排列 π_1 中，我们发现满足 $\begin{cases}i<j\\ \pi_1(i)>\pi_i(j)\end{cases}$ 的有序数对个数为 0，因此 $\operatorname{sgn}(\pi_1)=(-1)^0=1$；同理在 π_1 中我们有且仅有一对满足条件的排列 $(1<2;\pi_2(1)>\pi_2(2))$，则 $\operatorname{sgn}(\pi_2)=(-1)^1=-1$。随后，根据定理 2-20，我们有

$$\begin{aligned}\det(A)&=\sum_{\pi\in S_2}\left(\operatorname{sgn}(\pi)\prod_{i=1}^{2}(a_{\pi(i)\,i})\right)\\ &=\operatorname{sgn}(\pi_1)(a_{\pi_1(1)1})(a_{\pi_1(2)2})+\operatorname{sgn}(\pi_2)(a_{\pi_2(1)1})(a_{\pi_2(2)2})\end{aligned}$$

随后，根据排列的性质，我们知道 $\pi_1(1)=1;\pi_1(2)=2;\pi_2(1)=2;\pi_2(2)=1$，我们将这些数值代入，

$$\begin{aligned}&=(+1)a_{11}a_{22}+(-1)a_{21}a_{12}\\ &=a_{11}a_{22}-a_{21}a_{12}\end{aligned}$$

因此对于二阶矩阵的行列式而言，我们可以简记为“左上乘右下，减右上乘左下”。那么二阶矩阵的行列式有没有什么几何意义呢？我们设想在平面直角坐标系 xOy 内有向量 $\boldsymbol{v}_1=\begin{bmatrix}a\\ b\end{bmatrix}$ 与 $\boldsymbol{v}_2=\begin{bmatrix}c\\ d\end{bmatrix}$，我们假设 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 不重合，然后我们考虑以 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 为两条邻边所构成的平行四边形。经过割补法等运算，不难求出这个平行四边形的面积为 $|ad-bc|$，也正是 $|\det(\boldsymbol{v}_1,\boldsymbol{v}_2)|$。因此二阶行列式的绝对值的几何意义也是在平面直角坐标系内由 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 为邻边所围成的平行四边形的面积。



例 2.6.2：

计算三阶矩阵 $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ 的行列式。

我们首先找出 S_3 中的全部元素，它们分别是：

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}; \pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}; \pi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

$$\pi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}; \pi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}; \pi_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

此时读者可以尝试自行计算各个排列的符号，参照 *2.6.1* 中的方法，我们最终得到

$$\det(A) = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}$$

同样地，三阶行列式的绝对值的几何意义是空间直角坐标系中由三组列向量所围成的平行六面体的体积。

当 n 逐渐增大时，我们所分析的排列数目也会增多，计算也会变得愈发复杂，那么有没有一种可以相对快速地计算行列式的方法呢？当然有，并且答案就在定理 2-20 中。我们只需要对定理 2-20 中的公式做一些变形转化，便可以得到一个相对简便的计算公式。我们下面将进行详细地推理：

对于 $n \times n$ 矩阵 A 而言，首先回顾我们所给出的定义：

$$\det(A) = \sum_{\pi \in S_n} \left(\operatorname{sgn}(\pi) \prod_{i=1}^{n} (a_{\pi(i)\,i}) \right)$$

此时，我们选定元素 j，考虑满足条件的下面数组排列：

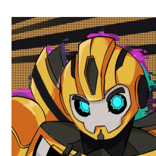
$$\pi_1(j) = 1, \pi_2(j) = 2, \cdots, \pi_k(j) = k$$

那么对于我们选定的 j 而言，考虑满足条件的所有排列 $\pi_i(j) = i$, 原式中 $\left(\operatorname{sgn}(\pi) \prod_{i=1}^{n} (a_{\pi(i)\,i})\right)$ 即可以写作 $a_{ij} \cdot \left(\sum\limits_{\pi \in S_n, \pi_i(j)=i} \left(\operatorname{sgn}(\pi) \prod\limits_{k \neq j} a_{\pi(k)k} \right) \right)$，所以当我们取遍所有的 $j \in \{1, 2, \cdots, n\}$ 时，我们有

$$\det(A) = \sum_{i=1}^{n} a_{ij} \left(\sum_{\pi \in S_n, \pi(j)=i} \operatorname{sgn}(\pi) \prod_{k \neq j} a_{\pi(k)k} \right)$$

此时，我们将重点研究 $\left(\sum_{\pi \in S_n, \pi(j)=i} \operatorname{sgn}(\pi) \prod_{k \neq j} a_{\pi(k)k}\right)$ 一项。为了方便研究，我们定义两个特殊的排列 f, g，使得

$$\begin{pmatrix} (1 & 2 & \cdots & j-1) & (j & j+1 & \cdots & n-1 & n) \\ & & & & \downarrow f & & & & \\ (1 & 2 & \cdots & j-1) & (j+1 & j+2 & \cdots & n & j) \end{pmatrix}$$



$$\begin{pmatrix} (1 & 2 & \cdots & i-1) & (i & i+1 & \cdots & n-1 & n) \\ & & & & \downarrow g & & & & \\ (1 & 2 & \cdots & i-1) & (i+1 & i+2 & \cdots & n & i) \end{pmatrix}$$

我们此时, 在矩阵 $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$ 中，我们考虑一个新的 $(n-1)\times(n-1)$ 矩阵：

$$\mathcal{Q} = \begin{bmatrix} a_{g(1)f(1)} & a_{g(1)f(2)} & \cdots & a_{g(1)f(n-1)} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{g(n-1)f(1)} & a_{g(n-1)f(2)} & \cdots & a_{g(n-1)f(n-1)} \end{bmatrix}$$

此时根据我们对 f,g 的定义可知，矩阵 $\mathcal{Q}$ 与 i,j 的取值有关，我们不妨来通过设出确定的 i,j，来考虑矩阵 $\mathcal{Q}$ 的性质。

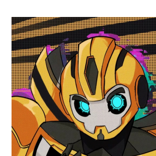
我们首先设 $i=j=1$, 那么根据 f,g 的定义，我们有$\mathcal{Q} = \begin{bmatrix} a_{g(1)f(1)} & a_{g(1)f(2)} & \cdots & a_{g(1)f(n-1)} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{g(n-1)f(1)} & a_{g(n-1)f(2)} & \cdots & a_{g(n-1)f(n-1)} \end{bmatrix}$

$$= \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

我们再看一组取值：设 $i=n, j=1$, 那么同样地，我们有$\mathcal{Q} = \begin{bmatrix} a_{g(1)f(1)} & a_{g(1)f(2)} & \cdots & a_{g(1)f(n-1)} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{g(n-1)f(1)} & a_{g(n-1)f(2)} & \cdots & a_{g(n-1)f(n-1)} \end{bmatrix}$

$$= \begin{bmatrix} a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{(n-1)2} & a_{(n-1)3} & \cdots & a_{(n-1)n} \end{bmatrix}$$

读者此时不妨多取几组 i,j 的值来寻找规律。事实上，当我们取特定的 $1 \leq i,j \leq n$ 时，我们会发现所得到的 $(n-1)\times(n-1)$ 矩阵 $\mathcal{Q}$ 即为原矩阵中除去第 i 行与第 j 列之后的矩阵。



定义 2-17
在 $n\times n$ 矩阵 A 中，我们定义 A_{ij} 为矩阵 A 去除第 i 行与第 j 列之后所得到的 $(n-1)\times(n-1)$ 矩阵。我们称 A_{ij} 为 A 在第 i 行，第 j 列的**辅助因子矩阵 *(Cofactor Matrix)***，我们同时定义 A 在第 i 行，第 j 列位置的**辅助因子 *(Cofactor)***(也称“代数余子式”) 为 $A^{ij}=(-1)^{(i+j)}\det(A_{ij})$。

让我们继续上文的推理：我们此时如果设 $1\le\alpha,\beta\le n-1$，设矩阵 $A=(a_{ij})$，那么我们不难证明 $A_{ij}=\left(a_{g(\alpha)f(\beta)}\right)$。那么，结合我们在之前所得到的 $\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{k\ne j}a_{\pi(k)k}\right)$ 式子，我们将其用含有 f,g 的式子表示，同时定义 $b_{\alpha\beta}=a_{g(\alpha)f(\beta)}$，则我们有

$$\begin{aligned}\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{k\ne j}a_{\pi(k)k}\right)&=\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{k\ne j}b_{[g^{-1}(\pi(k))f^{-1}(k)]}\right)\\&=\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{t=1}^{n-1}b_{g^{-1}[\pi(f(t))]t}\right)\end{aligned}$$

我们此时定义一个新的排列 $\pi'=g^{-1}\pi f$，其中 $\pi(j)=i$。那么根据排列的知识，我们有 $\operatorname{sgn}(\pi)=\operatorname{sgn}(g)\operatorname{sgn}(\pi')\operatorname{sgn}(f^{-1})=(-1)^{i+j}\operatorname{sgn}(g^{-1}\pi f)$，因此综合我们之前所得，我们有

$$\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{k\ne j}a_{\pi(k)k}\right)=(-1)^{i+j}\sum_{\pi'\in S_{n-1}}\operatorname{sgn}(\pi')\prod_{t=1}^{n-1}b_{\pi'(t)t}=A^{ij}$$

于是，我们再结合之前的式子，把最开始的 $\sum a_{ij}$ 添加到其中来，我们得到：

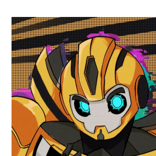
$$\det(A)=\sum_{i=1}^{n}a_{ij}\left(\sum_{\pi\in S_n,\pi(j)=i}\operatorname{sgn}(\pi)\prod_{k\ne j}a_{\pi(k)k}\right)=\sum_{i=1}^{n}a_{ij}A^{ij}$$

我们作为总结，给出下面的定理：

定理 2-21 (LaPlace 定理)
设 A 为 $n\times n$ 矩阵，对任意的 $1\le i,j\le n$，有

$$\begin{cases}\det(A)=a_{i1}A^{i1}+a_{i2}A^{i2}+\cdots+a_{in}A^{in}=\displaystyle\sum_{j=1}^{n}(-1)^{(i+j)}a_{ij}\det(A_{ij})\ \text{（行列式的按行展开法则）}\\ \det(A)=a_{1j}A^{1j}+a_{2j}A^{2j}+\cdots+a_{nj}A^{nj}=\displaystyle\sum_{i=1}^{n}(-1)^{(i+j)}a_{ij}\det(A_{ij})\ \text{（行列式的按列展开法则）}\end{cases}$$

定理 2-21 是本章中最重要的定理之一，它告诉我们计算矩阵的行列式的时候，我们可以利用“降维打击”的方法，把一个高阶行列式化简成低阶行列式，这种化简方法便是定理 2-21。我们来看几个具体的例子：



例 2.6.3：

计算 $\begin{bmatrix} 1 & 3 & 2 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{bmatrix}$ 的行列式。

根据定理 *2-21*，我们不妨将该矩阵按第一行展开，那么我们有

$$\begin{aligned} \det(A) &= 1\det(A_{11}) - 3\det(A_{12}) + 2\det(A_{13}) \\ &= \det\begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix} - 3\det\begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} + 2\det\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \end{aligned}$$

我们知道，对于二阶行列式而言，其结果便是"左上乘右下减去右上乘左下"，所以我们很快便能得到

$$\det(A) = 1 \times (-3) - 3 \times (-3) + 2 \times (3) = 12$$

定理 2-22

在一个 $n \times n$ 矩阵中，若 $(-1)^{i+j} = 1$，我们记该位置的符号为 $+$，若 $(-1)^{i+j} = -1$，我们记该位置的符号为 $-$，那么该矩阵对应的符号矩阵为

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & & & & \end{bmatrix}$$

我们可以简记为：位于 $i = j = 1$ 位置的符号永远为正，然后任何相邻位置的符号不同。

随后我们会讨论一些求解矩阵行列式的技巧与推论。首先根据定理 2-21 我们知道，不管我们按照哪一行或者哪一列来展开矩阵，所得的行列式的结果不变，因此在展开矩阵的时候我们要挑选尽量形式简单的行或列，比如有元素 0 的行或列，这样一来我们就可以减少计算量。同时要知道行列式也是一个交替多线性映射，那么矩阵的行列式也就满足所有交替多线性映射的性质。

定理 2-23

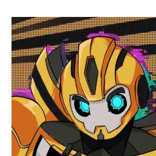
在一个 $n \times n$ 矩阵 A 中，记 A 的行列式为 $\det(A)$，则：

① 若 A 中含有任意一行或一列的元素全部为零，或者 A 中含有完全相同的两行或两列，则 $\det(A) = 0$；

② 交换 A 中的任意两行或者两列，行列式的符号改变；

③ 对于任意的非零常数 λ，满足 $\det(\lambda A) = \lambda^n \det(A)$；

④ 将一行或一列中的元素乘以同一非零常数，然后加到另一行或另一列的对应元素上，$\det(A)$ 不变



我们发现，定理 2-23 中的第 1，2，3 条我们均以在交替多线性映射中证明，我们在此仅证明第四条结论：

定理 2-23-4 的证明：

证明. 设 $A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ 为了简化运算同时不失一般性，我们便假设将第二列的元素乘以非零常数 λ，然后加到第一列上，即 $A' = \begin{bmatrix} \boldsymbol{v}_1 + \lambda\boldsymbol{v}_2 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$，根据交替多线性映射的性质我们有

$$\det(A') = \det(\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}) + \lambda\det(\begin{bmatrix} \boldsymbol{v}_2 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix})$$

即

$$\det(A') = \det(\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}) + 0 = \det(A)$$

因此行列式的值不变。 ■

定理 2-24
在一个 $n \times n$ 矩阵 A, B 中，记 A, B 的行列式分别为 $\det(A), \det(B)$，则：

① $\det(AB) = \det(A)\det(B)$;

② $\det(A) \neq 0$ 的充要条件是矩阵 A 中行向量，列向量彼此线性无关；

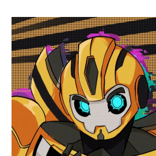
③ 当 $\det(A) \neq 0$ 时，$\det(A^{-1}) = \dfrac{1}{\det(A)}$

④ $\det(A) \neq 0$ 的充要条件是矩阵 A 可逆；

同样地，定理 2-24 还可以进行扩充，这里由于篇幅原因不再展开，读者不妨停下来仔细思考一下，尝试把我们目前所学的概念联系到一起，这样是一种非常有效的复习方法。

定义 2-18
在 $n \times n$ 矩阵 A 中，我们定义一个新的矩阵 A'，其中 $(A')_{ij} = A^{ij}$，即新矩阵中位置 (i, j) 的元素为 A 在位置 (i, j) 的辅助因子，则矩阵 $(A')^T$ 被我们称作是 A 的**伴随矩阵 *(Adjugate Matrix)***，通常用 $\operatorname{adj}(A)$ 表示。

对于伴随矩阵而言，我们再引出一条非常重要的定理：



定理 2-25

对于 $n \times n$ 矩阵 A 而言，记 $\mathrm{adj}(A)$ 为 A 的伴随矩阵，则

$$A(\mathrm{adj}(A)) = (\mathrm{adj}(A))A = (\det(A))I$$

特别地，如果 A 可逆，我们有：

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

该定理的证明将留作练习供读者思考。其中所采用的证明方法参见定理 2-21 与定义 2-17。我们来看一下该定理在求解线性方程组中的应用：假设我们需要求解线性方程组 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$, 我们设

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

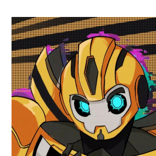
我们此时规定 $\mathcal{A}$ 可逆，则 $\boldsymbol{x} = \mathcal{A}^{-1}\boldsymbol{b}$, 再由定理 2-25，我们得到

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} &= \frac{1}{\det(\mathcal{A})}(\mathrm{adj}(\mathcal{A}))\boldsymbol{b} \\ &= \frac{1}{\det(\mathcal{A})} \begin{bmatrix} \mathcal{A}^{11} & \mathcal{A}^{21} & \cdots & \mathcal{A}^{n1} \\ \mathcal{A}^{12} & \mathcal{A}^{22} & \cdots & \mathcal{A}^{n2} \\ \vdots & \vdots & & \vdots \\ \mathcal{A}^{1n} & \mathcal{A}^{2n} & \cdots & \mathcal{A}^{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{aligned}$$

由此我们不难得到：

$$\begin{aligned} x_1 &= \frac{1}{\det(\mathcal{A})}\left[b_1\mathcal{A}^{11} + b_2\mathcal{A}^{21} + \cdots + b_n\mathcal{A}^{n1}\right] \\ x_2 &= \frac{1}{\det(\mathcal{A})}\left[b_1\mathcal{A}^{12} + b_2\mathcal{A}^{22} + \cdots + b_n\mathcal{A}^{n2}\right] \\ &\vdots \\ x_n &= \frac{1}{\det(\mathcal{A})}\left[b_1\mathcal{A}^{1n} + b_2\mathcal{A}^{2n} + \cdots + b_n\mathcal{A}^{nn}\right] \end{aligned}$$

我们不难发现，根据之前所给出的定义，形如 $b_1\mathcal{A}^{1i} + b_2\mathcal{A}^{2i} + \cdots + b_n\mathcal{A}^{ni}$ 便可以是看作把矩阵 $\mathcal{A}$ 中的第 i 列替换成 $\boldsymbol{b}$ 之后所得到的新矩阵的行列式。这样一来我们便发现了一种全新的求解线性方程组的方法，这便



是**克拉默法则 (Cramer's Rule)**。

定理 2-26 (克拉默法则)

设 $n \times n$ 矩阵 $\mathcal{A}$ 可逆，在线性方程组 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 中，其解 $x_1, x_2, \cdots, x_n$ 分别为

$$x_1 = \frac{\det(\mathcal{A}_1)}{\det(\mathcal{A})}; x_2 = \frac{\det(\mathcal{A}_2)}{\det(\mathcal{A})}; \cdots ; x_n = \frac{\det(\mathcal{A}_n)}{\det(\mathcal{A})}$$

其中 $\mathcal{A}_k$ 代表将矩阵 $\mathcal{A}$ 的第 k 列替换成 $\boldsymbol{b}$ 所得出的新矩阵。

至此，第二章的内容到此结束。我知道线性变换所涵盖的东西实在太多，但无奈篇幅有限，只好忍痛割爱。在第三章里面，我将从各个角度讲解线性变换的应用。第二章的内容便是第三章的垫脚石，希望读者认真反复地研读该章节，为自己的线性代数水平打下扎实的基础。

2.6 练习

---

1. 计算下列矩阵的行列式。

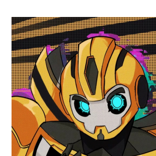
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & a & 0 \\ b & c & d \\ 0 & e & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & b & c \\ b & c & 1 \\ c & 1 & b \end{bmatrix}$$

2. 我们在 $n \times n$ 矩阵中，将对角线: $(A)_{11} - (A)_{22} - \cdots - (A)_{nn}$ 称作是矩阵 A 的主对角线，我们定义矩阵 A 为**上三角矩阵 (Upper Triangular Matrix)**，如果 A 中所有位于主对角线以下的元素均为 0；我们定义矩阵 A 为**下三角矩阵 (Upper Triangular Matrix)**，如果 A 中所有位于主对角线以上的元素均为 0。

① 利用排列的知识，证明：当 A 为上三角矩阵或下三角矩阵时，其行列式等于主对角线上的元素的乘积。

② 设 M 可以写成分块矩阵 $M = \begin{bmatrix} X & A \\ \mathbf{0} & Y \end{bmatrix}$, 其中 X, Y 均为方阵。证明：$\det(A) = \det(X)\det(Y)$, 同样的结论对与形如 $\begin{bmatrix} X & \mathbf{0} \\ B & Y \end{bmatrix}$ 的矩阵同样适用。

③ 求 $\mathcal{P} = \begin{bmatrix} -1 & -1 & 1 & 0 \\ 2 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & 3 & -1 & 2 \end{bmatrix}$ 的行列式 $\det(\mathcal{P})$。



3. 下列说法中哪些是正确的？哪些是错误的？如果错误，试着举出一个反例或说明原因。

① $\det(A+B)=\det(A)+\det(B)$；

② $\det(A)=\det(\mathbf{RREF}(A))$;

③ $\det(A)=\det(A^T)$;

④ 若 $A^T=-A$, 则 $\det(A)=-1$;

⑤ $\det(A^TA)>0$;

⑥ $\det(I+A)=1+\det(A)$;

⑦ 若 A 中有一行元素全部为零，则 $\mathrm{adj}(A)$ 中也有一行元素全部为零;

⑧ 若 A 可逆，则 $\mathrm{adj}(A)$ 可逆;

4. 证明：对任意的 3×3 矩阵 A 而言，$A^2+I\neq\mathbf{0}$。

5. 证明：对任意的 $n\times n$ 矩阵 A,B 而言，$\det(A+B^T)=\det(A^T+B)$。

6. 若 $\det\begin{bmatrix} a & b \\ c & d\end{bmatrix}=-2$, 求 $\det\begin{bmatrix} 2 & -2 & 0 \\ c+1 & -1 & 2a \\ d-2 & 2 & 2b\end{bmatrix}$。

7. 设 $Q=\begin{bmatrix} 1+x & 2 & 3 & 4 \\ 1 & 2+x & 3 & 4 \\ 1 & 2 & 3+x & 4 \\ 1 & 2 & 3 & 4+x\end{bmatrix}$, 据此求 $\det(Q)$。

8. 若 $\det\begin{bmatrix} x & y & 0 & 0 \\ 0 & x & y & 0 \\ 0 & 0 & x & y \\ y & 0 & 0 & x\end{bmatrix}=\det\begin{bmatrix} 1 & a & a^2 & a^3 \\ a & a^2 & a^3 & 1 \\ a^2 & a^3 & 1 & a \\ a^3 & 1 & a & a^2\end{bmatrix}=0$, 据此求 x,y,a。

9. 若 $A^{-1}=\begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 3 \\ 3 & 1 & -1\end{bmatrix}$, 求 $\mathrm{adj}(A)$。

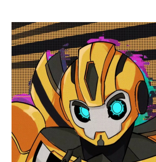
10. 设 A 为 $n\times n$ 矩阵 $(n\geq 2)$, 证明:

① 对于任意的非零常数 λ, 有 $\mathrm{adj}(\lambda A)=\lambda^{n-1}\,\mathrm{adj}(A)$；

② $\mathrm{adj}(A^{-1})=(\mathrm{adj}(A))^{-1}$；

① $\mathrm{adj}(A^T)=(\mathrm{adj}(A))^T$；

④ $\mathrm{adj}(\mathrm{adj}(A))=(\det(A))^{n-2}A$



3 线性变换之间的转化

在上一章中，我们学习了线性变换的基本性质，我们知道当线性变换选定基底之后便有了矩阵。因此对于同一个线性变换而言，如果选定的基底不同，所表达出的矩阵也就不同。这些矩阵不同基底下面的矩阵之间有着怎样的联系？完成本章的学习，读者便会对这一概念有着更加深刻的理解。

3.1 不同基底之间的转换

如果读者此时还有印象，想必会记得在第二章第一节里面我举了一个例子：求出把向量投影到直线 $y = 2x$ 上的变换矩阵。那么理想情况下便是取两个分别与直线平行，垂直的向量作为基底，此时的线性变换矩阵便会十分简单。但是我们最后却混淆了不同的基底，导致计算结果出错。我们现在来重新审视一下这个问题：

首先我们考虑线性变换 $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{V}$ 以及线性空间 $\mathcal{V}$ 的两组基底

$$\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}; \gamma = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n\}$$

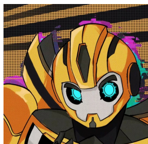
我们假设选取 β 基底作为我们的“参考系”，那么对于任意的一个向量 $\boldsymbol{v} \in \mathcal{V}$, 我们记该向量在基底 β 下的坐标为 $[\boldsymbol{v}]_\beta$。此时我们设 $[\mathcal{T}]_\beta$ 为线性变换在基底 β 下的表达形式，那么只有这样，我们知道 $[\mathcal{T}]_\beta[\boldsymbol{v}]_\beta$ 才有意义。如果此时我们假设 $[\mathcal{T}]_\gamma$ 为线性变换在基底 γ 下的表达形式，此时我们知道乘法 $[\mathcal{T}]_\gamma[\boldsymbol{v}]_\beta$ 没有意义，我们需要将二者转化到同一个基底里面。我们不妨假设 $[\mathcal{T}]_\gamma$ 的形式非常简便，所以我们会考虑把 $[\boldsymbol{v}]_\beta$ 转化为 $[\boldsymbol{v}]_\gamma$。我们假设线性变换 $\mathcal{T}_\beta^\gamma : [\boldsymbol{v}]_\beta \longrightarrow [\boldsymbol{v}]_\gamma$ 便是实现这一转化的媒介。设该变换的矩阵形式为 $\mathcal{P}$, 那么根据定义，我们有

$$\mathcal{P} = \begin{bmatrix} | & | & & | \\ [(\boldsymbol{v}_1)]_\gamma & [(\boldsymbol{v}_2)]_\gamma & \cdots & [(\boldsymbol{v}_n)]_\gamma \\ | & | & & | \end{bmatrix}$$

此时，我们便可以使用 $[\mathcal{T}]_\gamma[\boldsymbol{v}]_\gamma$ 进行计算, 记 $[\mathcal{T}]_\gamma[\boldsymbol{v}]_\gamma = [\boldsymbol{w}]_\gamma$。我们最后需要做的即为将该向量重新变化至我们所选定的基底 β 中。为此我们便可以使用 $\mathcal{T}_\beta^\gamma$ 的逆变换。下面的图便是一个非常直观的例子：

$$\begin{tikzcd}
[\boldsymbol{v}]_\beta \arrow[r, "{[\mathcal{T}]_\beta}"] \arrow[d, "\mathcal{P}"'] & {[\mathcal{T}\boldsymbol{v}]_\beta} \\
[\boldsymbol{v}]_\gamma \arrow[r, "{[\mathcal{T}]_\gamma}"'] & {[\mathcal{T}\boldsymbol{v}]_\gamma} \arrow[u, "\mathcal{P}^{-1}"]
\end{tikzcd}$$

由此不难看出，对于任意的 $[\boldsymbol{v}]_\beta$, 有 $[\mathcal{T}]_\beta[\boldsymbol{v}]_\beta = (\mathcal{P}^{-1}[\mathcal{T}]_\gamma\mathcal{P})\,[\boldsymbol{v}]_\beta$。这样一来，我们就完成了不同基底之间的转换。那么现在我们就可以回到之前所提到的例子了：



例 3.1.1

在平面直角坐标系 $\mathbb{R}^2$ 中，线性变换 $\mathcal{P}$ 满足以下性质：对于任意的 $\boldsymbol{x} \in \mathbb{R}^2$, $\mathcal{P}(\boldsymbol{x})$ 的结果为将向量 $\boldsymbol{x}$ 投影至直线 $y = 2x$ 上。那么对于任意的向量 $\boldsymbol{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ 而言，计算其结果 $\mathcal{P}(\boldsymbol{x})$

我们选取两组基底：$\beta = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}; \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$，不难发现 β 即为标准基底。随后我们选取两个特殊的向量：与直线 $y = 2x$ 重合的向量以及与直线 $y = 2x$ 垂直的向量。那么我们便可以选取一组新的基底：$\gamma = \left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)$，此时我们知道线性变换在基底 γ 下的矩阵形式便很容易求得：

$$[\mathcal{P}]_\gamma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

此时，如果要求出 $[\mathcal{P}]_\beta$，我们便可以使用形如上一页图中的转换。我们设线性变换 $\mathcal{Q} : [\boldsymbol{v}]_\beta \longrightarrow [\boldsymbol{v}]_\gamma$，那么我们便可以求出矩阵 $\mathcal{Q}$，矩阵 $\mathcal{Q}$ 即为

$$\mathcal{Q} = \begin{bmatrix} | & | \\ \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)_\gamma & \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)_\gamma \\ | & | \end{bmatrix}$$

如果要求 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}_\gamma$，则我们需要求满足条件的线性组合：

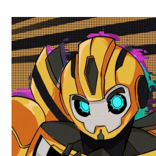
$$\lambda_1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

此时我们有 $\lambda_1 = -\frac{2}{5}; \lambda_2 = \frac{1}{5}$，同样我们可以求 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}_\gamma$，因此我们便可以求得 $\mathcal{Q}$ 为

$$\mathcal{Q} = \begin{bmatrix} -\frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

那么此时，参考我们上一页给出的图表，我们有

$$\mathcal{P}(\boldsymbol{x}) = \left(\begin{bmatrix} -\frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} \end{bmatrix}\right)^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} \end{bmatrix} (\boldsymbol{x})$$



那么对于这种形如 $\mathcal{A}=\mathcal{Q}^{-1}\mathcal{P}\mathcal{Q}$ 的形式，我们用矩阵的相似来定义该关系。

定义 3-1
我们称 $n\times n$ 矩阵 $\mathcal{A},\mathcal{Q}$ 为**相似矩阵（*Similar Matrix*）**，一般记作 $\mathcal{A}\simeq\mathcal{Q}$，如果存在 $n\times n$ 可逆矩阵 $\mathcal{Q}$，使得 $\mathcal{A}=\mathcal{Q}^{-1}\mathcal{P}\mathcal{Q}$。

不难看出，相似矩阵之间还是一个等价关系，即相似矩阵满足自反性，对称性和传递性。有兴趣的读者可以自行证明。

定理 3-1
设 $A,B,C\in M_n$，则

① $A\simeq A$

② $A\simeq B\Longrightarrow B\simeq A$

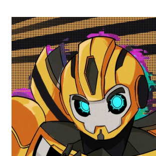
③ $A\simeq B; B\simeq C\Longrightarrow A\simeq C$

同时，相似矩阵之间还满足一条重要的定理：

定理 3-2
设矩阵 $A\simeq B$，则 $\det(A)=\det(B)$

定理 3-2 的证明：

证明. 根据定义，若 $A\simeq B$，那么存在可逆矩阵 Q，使得 $A=Q^{-1}BQ$，那么 $\det(A)=\det(Q^{-1}BQ)=\det(Q^{-1})\det(B)\det(Q)=\dfrac{1}{\det(Q)}\det(B)\det(Q)=\det(B)$，因此定理 3-2 得证。 ■



3.2 线性变换（矩阵）的特征向量与特征值

在函数中，我们接触到过“不动点”的概念，比如像 $f(f(x)) = C$（C 为常数）这样的例子。当我们知道某个函数具有类似的性质时，在研究函数的复合时便会容易很多。那么对于线性变换而言，我们能够找到一个线性变换里面的“不变量”呢？即类似 $T(T(\boldsymbol{x})) = \boldsymbol{y}$ 这样的形式。我们用**特征向量（Eigenvectors）**和**特征值（Eigenvalues）**来定义线性变换里面的这一“不变”的关系。

定义 3-2
在线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 中，如果存在常数 $\lambda\in\mathbb{F}$ 以及**非零**向量 $\boldsymbol{v}\in\mathcal{V}$，使得

$$\mathcal{T}(\boldsymbol{x}) = \lambda\boldsymbol{x}$$

成立，我们则称 λ 为线性变换 $\mathcal{T}$ 的一个特征值，$\boldsymbol{x}$ 为与特征值 λ 所对应的特征向量。

定理 3-3
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 的一个特征值为 λ，与该特征值所对应的特征向量为 $\boldsymbol{v}$。那么对于任意的正整数 n，有

$$\mathcal{T}^n(\boldsymbol{v}) = \lambda^n\boldsymbol{v}$$

因此，我们得到了同函数里面的“不动点”所类似的推论。在此情况下当我们计算线性变换自身的复合时，便可以利用特征向量和特征值的关系来进行简便运算。这一点在后面章节的矩阵对角化里面会重点提及。我们现在需要知道的是：并不是所有的线性变换都会有特征向量和特征值；同理可知一个线性变换也可能会有多个特征向量和特征值；一个特征向量仅能和与之对应的特征值组合，否则就不具备定义 3-2 中的性质。

那么我们便想知道对于给定的线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$，我们该如何计算该线性变换的特征向量和特征值呢？我们一起来看下面的推理过程：

我们假设在线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 中，λ 为一个特征值，$\boldsymbol{v}$ 为该特征值所对应的特征向量。那么根据定义，我们有

$$\mathcal{T}(\boldsymbol{v}) = \lambda\boldsymbol{v}$$

我们将原等式进行移项，得到

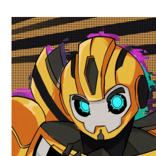
$$\mathcal{T}(\boldsymbol{v}) - \lambda(\boldsymbol{v}) = 0$$

根据单位变换的性质，设 I 为单位矩阵，$\mathcal{T}$ 的矩阵形式为 $\mathcal{A}$，那么有 $I\boldsymbol{v} = \boldsymbol{v}$，因此

$$\mathcal{A}(\boldsymbol{v}) - \lambda(I\boldsymbol{v}) = 0$$

即

$$\boldsymbol{v}(\mathcal{A} - \lambda I) = 0$$



根据定义，我们强调 $\boldsymbol{v} \neq \mathbf{0}$, 那么就有

$$\mathcal{A} - \lambda I = \mathbf{0}$$

对于零矩阵而言，我们可以使用行列式的性质，即

$$\det(\mathcal{A} - \lambda I) = 0$$

定义 3-3

在线性变换 $\mathcal{T} : \mathcal{V} \longrightarrow \mathcal{V}$ 中，设 $\mathcal{T}$ 在某一选定基底下的矩阵形式为 A, 那么我们定义 $\mathcal{T}$ 在该基底下的 **特征多项式 *(Characteristic Polynomial)*** $C_{\mathcal{T}}(x)$ 为

$$C_{\mathcal{T}}(x) = \det(A - \lambda x)$$

其中，$\mathcal{T}$ 的特征值即为 $C_{\mathcal{T}}(x)$ 的零点。

由于我们知道, 当 λ 为特征值时，与其对应的特征向量 $\boldsymbol{v}$ 满足 $\boldsymbol{v}(\mathcal{T} - \lambda I) = 0$, 也就是说

$$\boldsymbol{v} \in \mathbf{Ker}(\mathcal{T} - \lambda I)$$

在这里，当特征值 λ 给定时，我们便也得知了全体的关于 λ 的特征向量即可以构成 $\mathcal{T} - \lambda I$ 的核空间。我们也可以将其称为是关于特征向量 λ 的特征空间，计作 $Eig_{\mathcal{T}}(\lambda)$。任何 $\boldsymbol{u} \in Eig_{\mathcal{T}}(\lambda)$ 均为与特征值 λ 对应的特征向量。特征空间也是 $\mathcal{V}$ 的一个字空间。有兴趣的读者可以自行证明。

例 3.2.1

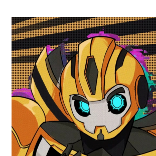
设线性变换 $\mathcal{P} : \mathcal{V} \longrightarrow \mathcal{V}$ 满足 $\mathcal{P}^2 = \mathcal{P}$ ，证明：$\mathcal{P}$ 的特征值为 0 或 1。

证明. 设 λ 为线性变换 $\mathcal{P}$ 的一个特征值；$\boldsymbol{v}$ 是与之对应的一个特征向量。由特征值的定义可知，$\mathcal{P}\boldsymbol{v} = \lambda\boldsymbol{v}$ ，再由题意可知，$\mathcal{P}^2\boldsymbol{v} = \lambda^2\boldsymbol{v} = \lambda\boldsymbol{v} = \mathcal{P}\boldsymbol{v}$, 因此有 $\lambda^2 = \lambda$, 即解得 $\lambda = 0; 1$。 ■

此时我们想研究一下相似矩阵，我们之所以称作相似矩阵，是因为二者之间存在着相似的结构。我们目前仅仅知道相似矩阵的行列式相同，但实际上相似矩阵也有着相同的特征向量。这一点我们同样会在后面几节矩阵的对角化里面重点提及。

定理 3-4

若矩阵 $\mathcal{A} \simeq \mathcal{B}$, 则 $\mathcal{A}, \mathcal{B}$ 具有相同的特征值。



定理 3-4 的证明:

证明. 根据相似矩阵的定义，我们知道存在可逆矩阵 $\mathcal{Q}$, 使得 $\mathcal{A}=\mathcal{Q}^{-1}\mathcal{B}\mathcal{Q}$ 我们设 λ 为一个特征值，那么我们有 $\det(\mathcal{A}-\lambda I)=\det(\mathcal{Q}^{-1}\mathcal{B}\mathcal{Q}-\lambda I)$，我们对此式进行等价变换，有

$$\det(\mathcal{A}-\lambda I)=\det(\mathcal{Q}^{-1}\mathcal{B}\mathcal{Q}-\lambda\mathcal{Q}^{-1}I\mathcal{Q})$$

即为

$$\det(\mathcal{A}-\lambda I)=\det(\mathcal{Q}^{-1}\mathcal{Q}(\mathcal{B}-\lambda I))$$

再由行列式的性质，我们知道

$$\begin{aligned}\det(\mathcal{Q}^{-1}\mathcal{Q}(\mathcal{B}-\lambda I))&=\det(\mathcal{Q}^{-1}\mathcal{Q})\det(\mathcal{B}-\lambda I)\\&=\frac{1}{\det(\mathcal{Q})}\det(\mathcal{Q})\det(\mathcal{B}-\lambda I)\\&=\det(\mathcal{B}-\lambda I)\end{aligned}$$

即我们得到 $\det(\mathcal{A}-\lambda I)=\det(\mathcal{B}-\lambda I)$, 二者的特征多项式相同，因此也具有相同的特征值。 ■

值得注意的是，虽然相似矩阵的特征值相同，但是由于矩阵结构毕竟不同，所以相似矩阵的特征向量（特征向量构成的空间）不一定相等。

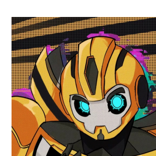
因此，我们现在知道如何计算给定矩阵（线性变换）的特征向量和特征值。欲求特征值，我们只需要求解方程 $\det(A-xI)=0$ 的解。回顾行列式的知识，我们想一下什么情况下矩阵的行列式能够很容易计算出来？我们会想到上三角矩阵和下三角矩阵（若读者不熟悉此概念请查阅 2.6 练习 2），此时矩阵的行列式即为主对角线上元素的乘积。我们此时不妨在特殊一点：考虑一个主对角线以外的位置全部为零的矩阵，即 $(A)_{ij}=0$ 若 $i\neq j$。形如这样的矩阵我们称作是**对角矩阵 (Diagonal Matrix)**，我们设对角矩阵 A 主对角线上的元素从左上方到右下方依次为 $\lambda_1,\lambda_2,\cdots,\lambda_n$。由于此时矩阵的结构简单，为了简化处理我们也可以将该对角矩阵 A 计作 $\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$。同上（下）三角矩阵类似，对角矩阵的行列式即为对角线上全体元素的乘积, 此时我们也不难发现，对角矩阵的特征多项式 $\det(A-xI)$ 中，有

$$(A-xI)_{ij}=\begin{cases}\lambda_i-x & (i=j)\\ 0 & (i\neq j)\end{cases}$$

此时，我们便有

$$\det(A-xI)=\prod_{i=1}^{n}(\lambda_i-x)$$

定理 3-5
对角矩阵，上（下）三角矩阵的特征值即为主对角线上的全部元素。



为了更好地判断一个线性变换（矩阵）的特征值，我们通常会运用因式分解，将矩阵的特征多项式写成形如 $p(x)=(\lambda_1-x)^{k_1}(\lambda_2-x)^{k_2}\cdots(\lambda_j-x)^{k_j}$ 的形式。此时，我们要考虑复数根的情况了，但是我会在题目中强调数域，没有特殊说明的时候我们还是仅考虑特征向量为实数的情况。

随后，我们在给出两个至关重要的定义：特征值的**代数重数** (**Algebraic Multiplicity**) 以及**几何重数** (**Geometric Multiplicity**), 这两个定义在后面的章节里面对于对角化的判断不可或缺。

定义 3-4

在线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 中，设 λ 为线性变换的一个特征值，则

① λ 的**代数重数** ***(Algebraic Multiplicity)*** 为特征多项式中 $(\lambda-x)$ 一项的次数；

② λ 的**几何重数** ***(Geometric Multiplicity)*** 为特征空间 $Eig_{\mathcal{T}}(\lambda)$ 的维数

例 3.2.2

设矩阵 $A=\begin{bmatrix}1&0&0\\1&2&0\\1&0&1\end{bmatrix}$, 求出 A 的特征值，以及这些特征值所对应的几何重数与代数重数。

A 的特征多项式为 $\det\left(\begin{bmatrix}1&0&0\\1&2&0\\1&0&1\end{bmatrix}-\begin{bmatrix}x&0&0\\0&x&0\\0&0&x\end{bmatrix}\right)$, 即 $C_A(x)=\det\left(\begin{bmatrix}1-x&0&0\\1&2-x&0\\1&0&1-x\end{bmatrix}\right)$, 根据行列式的展开法则，我们按第一行展开，即有

$$\begin{aligned}C_A(x)&=(1-x)[(2-x)(1-x)]\\&=(1-x)^2(2-x)\end{aligned}$$

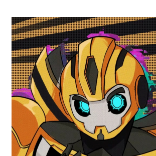
因此，我们发现矩阵 A 的特征值为 $\lambda_1=1;\lambda_2=2$, 其中根据定义，λ_1 的代数重数为 2；λ_2 的代数重数为 1。

对于 λ_1 而言，根据定义有 $Eig_A(\lambda_1)=\mathbf{Ker}(A-I)$, 即

$$Eig_A(\lambda_1)=\mathbf{Ker}\left(\begin{bmatrix}0&0&0\\1&1&0\\1&0&0\end{bmatrix}\right)$$

通过对矩阵 $A-I$ 进行高斯消元，我们得到

$$\mathbf{RREF}(A-I)=\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}$$



因此，我们知道 $\mathbf{Ker}(A-I)=\mathbf{Span}\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)$，不难发现其维数为 1。即特征值 $\lambda_1=1$ 的几何重数为 1。

在看特征值 λ_2，根据定义，我们得到

$$Eig_A(\lambda_2)=\mathbf{Ker}(A-2I)=\mathbf{Ker}\left(\begin{bmatrix}-1&0&0\\1&0&0\\1&0&-1\end{bmatrix}\right)$$

对 $A-2I$ 进行高斯消元，得到

$$\mathbf{RREF}(A-2I)=\begin{bmatrix}1&0&0\\0&0&0\\0&0&-1\end{bmatrix}$$

因此，$\mathbf{Ker}(A-2I)=\mathbf{Span}\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right)$，其维数为 2，因此特征值 λ_2 对应的几何重数为 2。

在以后的计算中，为方便文字说明，对于一个特征值 λ 而言，我们将使用 $m_\lambda(a)$ 代表其代数重数，使用 $m_\lambda(g)$ 代表其几何重数。

例 3.2.3

证明 $n\times n$ 矩阵 A 可逆 (存在逆矩阵) 的充要条件是 A 不存在为 0 的特征值。

证明. 我们可以证明其逆反命题：即 A 不可逆的充要条件是存在为 0 的特征值。

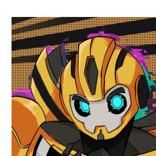
$(\Longrightarrow)$
假设 A 不可逆，根据定义我们有 $\det(A)=0$，在特征多项式 $\det(A-xI)$ 中正好对应 $x=0$ 的解，因此 A 的一个特征值为 0。

$(\Longleftarrow)$
假设 A 存在为 0 的特征值，则其特征多项式为 $C_A(x)=\det(A-xI)$，将 $x=0$ 代入特征多项式，得 $\det(A-0I)=\det(A)=0$，根据行列式的定义，$\det(A)=0$ 说明 A 不可逆，因此得证。

■

此时，我们想研究一种特殊的情况：我们设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$，其中 $\boldsymbol{v}\in\mathcal{V}$，我们假设 $\dim(\mathcal{V})=n$，那么我们考虑由以下向量为线性组合所构成的字空间：

$$\beta=\{\boldsymbol{v},\mathcal{T}(\boldsymbol{v}),\mathcal{T}^2(\boldsymbol{v}),\cdots,\mathcal{T}^{n-1}(\boldsymbol{v})\}$$



如果我们假设 β 彼此线性无关，那么根据定义，其即为 $\mathcal{V}$ 的一组基底，根据定义我们便知道 $\mathcal{T}^n(\boldsymbol{v}) \in \mathcal{V}$，并且存在常数 $c_0, c_1, \cdots, c_{n-1}$，使得

$$\mathcal{T}^n(\boldsymbol{v}) + c_0\boldsymbol{v} + c_1\mathcal{T}(\boldsymbol{v}) + \cdots + \mathcal{T}^{n-1}(\boldsymbol{v}) = 0$$

此时，我们设矩阵 $\mathcal{B}$ 为线性变换 $\mathcal{T}$ 在基底 β 下的矩阵表达形式，那么我们有

$$\mathcal{B} = \begin{bmatrix} | & | & & | \\ [\mathcal{T}(\boldsymbol{v})]_\beta & [\mathcal{T}(\mathcal{T}(\boldsymbol{v}))]_\beta & \cdots & [\mathcal{T}(\mathcal{T}^{n-1}(\boldsymbol{v}))]_\beta \\ | & | & & | \end{bmatrix}$$

根据定义，我们便有

$$\mathcal{B} = \begin{bmatrix} 0 & 0 & & \cdots & -c_0 \\ 1 & 0 & & \cdots & -c_1 \\ 0 & 1 & & \cdots & -c_2 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{bmatrix}$$

因此其特征多项式为

$$\begin{aligned} C_{\mathcal{B}}(x) &= \det(\mathcal{B} - xI) \\ &= \det \begin{bmatrix} -x & 0 & & \cdots & -c_0 \\ 1 & -x & & \cdots & -c_1 \\ 0 & 1 & & \cdots & -c_2 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -x - c_{n-1} \end{bmatrix} \end{aligned}$$

其中，我们将该行列式从第 n 列展开，那么有

$$\det(\mathcal{B} - xI) = (-1)^{2n}(-x - c_{n-1})\det(\mathcal{B}_{nn}) + \sum_{j=1}^{n-1}(-1)^{n+j}(-c_{j-1})\det(\mathcal{B}_{jn})$$

其中，我们观察到 $\mathcal{B}_{nn}$ 为下三角矩阵，因此 $\det(\mathcal{B}_{nn}) = (-x)^{n-1}$，同时 $\mathcal{B}_{jn}$ 为形如 $\begin{bmatrix} M_j & \mathbf{0} \\ \mathbf{0} & N_j \end{bmatrix}$ 的分块矩阵，且

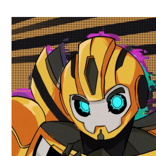
$$\det(\mathcal{B}_{jn}) = \det(M_j)\det(N_j) = (-x)^{j-1}(1)^{n-j}$$

由此可知

$$\det(\mathcal{B} - xI) = (-1)^{2n}(-x - c_{n-1})x^{n-1} + \sum_{j=1}^{n-1}(-1)^{2n}(-c_{j-1})(-x)^{j-1}$$

即

$$\det(\mathcal{B} - xI) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$$



此时，令 $x=\mathcal{T}$, 我们便有

$$\det(\mathcal{B}-\mathcal{T}I)=\mathcal{T}^n+c_{n-1}\mathcal{T}^{n-1}+\cdots+c_1\mathcal{T}+c_0I=0$$

不失一般性地讲，我们有 $C_{\mathcal{B}}(\mathcal{T})=0$, 由于我们事先已经知道，同一线性变换特征多项式与基底的选取无关，因此我们便得出了著名的 **Cayley-Hamilton** 定理：

定理 3-6 （Cayley-Hamilton）定理

设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, 若 $\mathcal{T}$ 的特征多项式为 $C_{\mathcal{T}}(x)$, 那么 $C_{\mathcal{T}}(\mathcal{T})=0$。

需要注意的是，在我们之前的推导中，我们默认了 $\{\boldsymbol{v},\mathcal{T}(\boldsymbol{v}),\mathcal{T}^2(\boldsymbol{v}),\cdots,\mathcal{T}^{n-1}(\boldsymbol{v})\}$ 彼此线性无关，但是在实际情况下往往这些向量线性相关。我们需要明白的是 $Cayley-Hamilton$ 定理对任何前后维数不变的线性变换以及 $n\times n$ 矩阵均成立。我们上面的推理仅仅是一种特殊情况下的证明。完整的证明需要更多的引理，在后续的章节我们会重新证明。因此在此我们不多做叙述。

3.2 练习

---

1. 求出下列矩阵的特征向量与特征值：

$$A=\begin{bmatrix}2&1&1\\0&1&0\\1&-1&2\end{bmatrix}\quad B=\begin{bmatrix}0&1&0\\3&0&1\\0&0&0\end{bmatrix}\quad C=\begin{bmatrix}1&1&-3\\2&0&6\\1&-1&5\end{bmatrix}$$

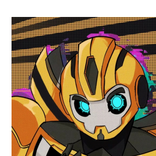
2. 我们称线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 在选定基底下对应的矩阵形式 A 为**幂零矩阵 (Nilpotent Matrix)**, 如果其满足 $A^k=\mathbf{0}$ 其中 $k\geq 1$。

① 证明主对角线全部元素为零的上（下）三角矩阵为幂零矩阵。

② 证明幂零矩阵有且仅有 0 这一个特征值。(无论数域为实数或复数该结论均成立)

③ 证明 $n\times n$ 幂零矩阵 A 的特征多项式为 $C_A(x)=x^n$。

3. 设 A 为 $n\times n$ 可逆矩阵，且 λ 为一个特征值（在例题中我们已经证明此时 $\lambda\neq 0$），证明 $\dfrac{1}{\lambda}$ 为 A^{-1}



的一个特征值。此时设 $C_A(x), C_{A^{-1}}(x)$ 分别为 A, A^{-1} 的特征多项式，试证明：

$$C_{A^{-1}}(x) = \frac{(-x)^n}{\det(A)} C_A(\frac{1}{x})$$

4.* 在复数域中，证明矩阵 $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ 的特征值为 $e^{i\theta}$ 和 $e^{-i\theta}$。

5. 设 A 为 $n \times n$ 矩阵，若 A 中的每一列（或每一行）元素之和均为同一常数 λ, 证明：λ 即为 A 的一个特征值。

6. 设分块矩阵 $A = \begin{bmatrix} B & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix}$, 其中 A, B, C 均为方阵。设 A, B, C 的特征多项式分别为 $C_A(x), C_B(x), C_C(x)$。

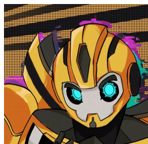
① 证明：$C_A(x) = C_B(x)C_C(x)$；

② 若 $\boldsymbol{x}$ 为 B 的一个特征向量；$\boldsymbol{y}$ 为 C 的一个特征向量，证明：$\begin{bmatrix} \boldsymbol{x} \\ \mathbf{0} \end{bmatrix}$; $\begin{bmatrix} \mathbf{0} \\ \boldsymbol{y} \end{bmatrix}$ 均为 A 的特征向量。

7. 在 $n \times n$ 矩阵 A 中，设 λ 为 A 的一个特征值，证明：

① $k\lambda$ 为矩阵 kA 的一个特征值；

② $\lambda^3 + 2\lambda + 3$ 为矩阵 $A^3 - 2A + 3I$ 的一个特征值。



3.3 线性变换（矩阵）的对角化

在特征向量与特征值一节中，我们知道形如 $\mathcal{T}(\boldsymbol{x}) = \lambda\boldsymbol{x}$ 的形式，其中 λ 为线性变换的特征值，$\boldsymbol{x}$ 为与其对应的特征向量。我们在本节中将尝试扩大 $\boldsymbol{x}$ 的可能取值范围——把范围扩大到线性变换所作用的整个线性空间 $\mathcal{V}$。此时，我们能否尝试利用向量分解的形式，对任意的 $\boldsymbol{v} \in \mathcal{V}$, 有 $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_n$, 使得每一个 $\boldsymbol{v}_i$ 都是和特征值 λ_i 有关的向量？这样一来我们便有

$$\mathcal{T}(\boldsymbol{v}) = \lambda_1\boldsymbol{v}_1 + \cdots + \lambda_n\boldsymbol{v}_n$$

此时我们对其中的每一项再进行研究。

定理 3-7
设 λ_1, λ_2 为线性变换 $\mathcal{T}$ 的两个不同的特征值，则 $Eig_{\mathcal{T}}(\lambda_1) \cap Eig_{\mathcal{T}}(\lambda_2) = \mathbf{0}$

该定理告诉了我们不同的特征空间之间没有交集（虽然其交集存在零向量，但是根据我们的定义，零向量不在我们所考虑的特征向量中）。这也就意味着特征空间彼此线性无关。那么这些线性无关的特征空间能否全部包含线性空间 $\mathcal{V}$ 呢？我们由此引出矩阵对角化的概念：

定义 3-5
在线性变换 $\mathcal{T}: \mathcal{V} \longrightarrow \mathcal{V}$ 中，我们称 $\mathcal{T}$ **可对角化** ***(Diagonalizable)***，如果存在互不相同的字空间 $\mathcal{W}_1, \mathcal{W}_2, \cdots, \mathcal{W}_n$ 以及常数 $\lambda_1, \lambda_2, \cdots, \lambda_n \in \mathbb{F}$, 使得：

① $\mathcal{V} = \bigoplus_{i=1}^{n} \mathcal{W}_i$

② 对于任意的 $\boldsymbol{w} \in \mathcal{W}_i$, 有 $\mathcal{T}(\boldsymbol{w}) = \lambda_i\boldsymbol{w}$

该定理告诉我们：线性变换（矩阵）可对角化需要满足两个条件：第一，其作用的线性空间 $\mathcal{V}$ 是若干个字空间的**直和 (Direct Sum)**(不清楚此概念的读者请阅读第一章第四节的内容)，这些字空间其实就是 $\mathcal{T}$ 的特征空间；第二：对于每一个特征空间 $\mathcal{W}_i$ 而言，对于 $\boldsymbol{w} \in \mathcal{W}_i$, 满足 $\mathcal{T}(\boldsymbol{w}) = \lambda_i\boldsymbol{w}$, 这里面 λ_i 即为和特征空间 $\mathcal{W}_i$ 相对应的特征值。

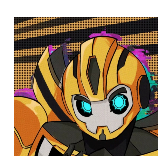
此时我们便知道，根据直和的定义，对于任意的 $\boldsymbol{v} \in \mathcal{V}$, 都存在 $\boldsymbol{w}_i \in \mathcal{W}_i$, 使得

$$\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2 + \cdots + \boldsymbol{w}_n$$

因此，再根据矩阵对角化的概念，我们有

$$\mathcal{T}(\boldsymbol{v}) = \lambda_1\boldsymbol{w}_1 + \lambda_2\boldsymbol{w}_2 + \cdots + \lambda_n\boldsymbol{w}_n$$

其中，λ_i 即为与特征空间 $\mathcal{W}_i$ 对应的特征值。



此时，由于特征空间彼此线性无关，因此 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n$ 也彼此线性无关。根据直和的定义，此时 $\gamma = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n\}$ 也构成 $\mathcal{V}$ 的一组基底, 如果我们考虑线性变换 $\mathcal{T}$ 在基底 γ 下面的矩阵形式，不难得到

$$[\mathcal{T}]_\gamma = \begin{bmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

此时，我们也联想到了基底之间的变换，如果设 $[\mathcal{T}]_E$ 为该线性变换在标准基底下的表达形式的话，那么我们知道，存在可逆矩阵 $\mathcal{Q}$，使得

$$[\mathcal{T}]_\gamma = \mathcal{Q}^{-1}[\mathcal{T}]_E\mathcal{Q}$$

我们再一次回顾形如下图的转化关系：

$$\begin{tikzcd}
{[\boldsymbol{v}]_\gamma} \arrow[r, "{[\mathcal{T}]_\gamma}"] \arrow[d, "\mathcal{Q}"'] & {[\mathcal{T}\boldsymbol{v}]_\gamma} \\
{[\boldsymbol{v}]_E} \arrow[r, "{[\mathcal{T}]_E}"'] & {[\mathcal{T}\boldsymbol{v}]_E} \arrow[u, "\mathcal{Q}^{-1}"']
\end{tikzcd}$$

其中可逆矩阵 $\mathcal{Q}$ 即为从标准基底 γ 变化到基底 E 的变化矩阵。此时不妨设在标准基底 $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$ 中，有

$$\boldsymbol{v} = c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + \cdots + c_n\boldsymbol{e}_n$$

同时，$\boldsymbol{v}$ 还满足

$$\boldsymbol{v} = \lambda_1\boldsymbol{w}_1 + \lambda_2\boldsymbol{w}_2 + \cdots + \lambda_n\boldsymbol{w}_n$$

根据定义，我们有：

$$\begin{bmatrix} | & & | \\ [\mathcal{T}(\boldsymbol{e}_1)]_\gamma & \cdots & [\mathcal{T}(\boldsymbol{e}_n)]_\gamma \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}_\gamma = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}_E$$

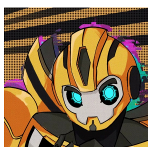
形如 $\begin{bmatrix} | & & | \\ [\mathcal{T}(\boldsymbol{e}_1)]_\gamma & \cdots & [\mathcal{T}(\boldsymbol{e}_n)]_\gamma \\ | & & | \end{bmatrix}$ 的矩阵即为可逆矩阵 $\mathcal{Q}$。

不难发现，我们得到

$$\lambda_1[\mathcal{T}(\boldsymbol{e}_1)]_\gamma + \cdots + \lambda_n[\mathcal{T}(\boldsymbol{e}_n)]_\gamma = \boldsymbol{v}$$

根据定义，我们便有 $[\mathcal{T}(\boldsymbol{e}_i)]_\gamma = \boldsymbol{w}_i$, 其中 $\boldsymbol{w}_i$ 为特征向量。

由此我们得出，可逆矩阵 $\mathcal{Q}$ 的第 i 列即为线性变换中与 λ_i 所对应的特征向量。在此时如果我们发现某一个特征空间的维数为 k ($k > 1$) 时，我们需要将该特征空间的特征值在对角线上重复 k 次，然后选取构成其基底的 k 个向量作为不同的特征向量依次写在相应的列里面。此时，我们便得出了判断线性变换（矩阵）是否可对角化的第一个定理：



定理 3-8
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, 那么 $\mathcal{T}$ 可对角化的条件是存在可逆矩阵 $\mathcal{Q}$ 与对角矩阵 $\mathcal{P}$, 使得

$$\mathcal{Q}^{-1}\mathcal{T}\mathcal{Q}=\mathcal{P}$$

其中 $\mathcal{Q}$ 的列向量即为 $\mathcal{T}$ 的特征向量；$\mathcal{P}$ 主对角线上的元素即为相对应的特征值。

定理 3-7 也可以叙述为：设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, 那么 $\mathcal{T}$ 可对角化的条件是存在可逆矩阵 $\mathcal{Q}$ 与对角矩阵 $\mathcal{P}$, 使得

$$\mathcal{Q}\mathcal{P}\mathcal{Q}^{-1}=\mathcal{T}$$

其中 $\mathcal{Q}$ 的列向量即为 $\mathcal{T}$ 的特征向量；$\mathcal{P}$ 主对角线上的元素即为相对应的特征向量。

随后，我们进一步来探究线性变换（矩阵）可对角化的条件。设线性空间 $\mathcal{V}$ 的维数为 n。假设此时线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 有 n 个不同的特征值，那么我们便知道此时就有 n 个彼此线性无关的特征向量，由这些特征向量作为列向量构成的矩阵可逆，也就满足了矩阵对角化的条件。由此我们可以再引出一条定理：

定理 3-9
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, $\dim(\mathcal{V})=n$。若 $\mathcal{T}$ 存在 n 个互不相同的特征值，则 $\mathcal{T}$ 可对角化。

如果 $\mathcal{T}$ 的特征值个数小于 n，此时并不意味着 $\mathcal{T}$ 不能对角化，我们只需要找出 n 个彼此线性无关的特征向量即可。此时我们注意到由于特征空间的个数小于 n, 因此一定存在维数大于 1 的特征空间，自然在该空间里面也可以找到多个彼此线性无关的特征向量。在这种情况下该特征空间对应的特征值在对角矩阵中会重复出现，出现次数与其维数相同。

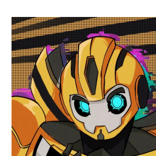
定理 3-9 (a)
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, $\dim(\mathcal{V})=n$。若 $\mathcal{T}$ 存在 n 个彼此线性无关的特征向量，则 $\mathcal{T}$ 可对角化。

此时，我们再从几何重数与代数重数的角度去理解这个问题。我们知道，特征值的几何重数代表其特征空间的维数，也代表着该特征空间里面彼此线性无关的特征向量的个数。因此我们还可以引出一条推论：

定理 3-9 (b)
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, $\dim(\mathcal{V})=n$。若 $\mathcal{T}$ 的所有特征值的几何重数之和为 n，则 $\mathcal{T}$ 可对角化。

再进一步的深入讨论之前，我们引出一条定理，来说明几何重数与代数重数之间的关系。

定理 3-10
设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$, 则对于 $\mathcal{T}$ 的任意一个特征值 λ 而言，其代数重数大于等于其几何重数。



定理 3-9 的证明我们在此省略，读者不必知道其证明方法。值得注意的是，任何一个特征值的代数重数均大于等于 1, 并且根据定义，所有特征值的代数重数之和为 n（线性空间的维数）。

例 3.3.1

设线性变换 $\mathcal{T}: \mathcal{V} \longrightarrow \mathcal{V}$, 已知 $\dim(\mathcal{V}) = 5$, 且 $\mathcal{T}$ 的特征值为 $0, 1, 2$。若 $\mathbf{rank}(\mathcal{T}) = 2$, 证明：$\mathcal{T}$ 可对角化。

证明. 由秩零定理可知，$\dim(\mathbf{Ker}(\mathcal{T})) = 3$, 即 $\dim(\mathbf{Ker}(\mathcal{T} - 0I)) = 3$, 则特征值 0 对应的特征空间的维数为 3，特征值 $\lambda = 0$ 的几何重数即为 3。也就是说 $\lambda = 0$ 的代数重数大于等于 3, 由于全体特征值的代数重数之和为 5, 且 $1, 2$ 这两个特征值的代数重数不为零，则 $\lambda = 0$ 的代数重数为 3, 特征值 $1, 2$ 的代数重数均为 1，且已知任何特征值的代数重数均大于等于其几何重数，所以特征值 $1, 2$ 的几何重数为 1 或 0。若特征值的代数重数为 0, 意味着 $\mathbf{Ker}(\mathcal{T} - \lambda I) = \{\mathbf{0}\}$, 也就是说 $\mathcal{T} - \lambda I$ 的列向量彼此线性无关。由于 $\mathbf{rank}(\mathcal{T}) \neq 5$, 所以 $\mathcal{T}$ 的列向量彼此线性相关，根据线性组合的知识也就得出 $\mathcal{T} - \lambda I$ 的列向量彼此线性相关。所以特征值 $1, 2$ 的几何重数均为 1。此时根据定理 *3-8* （*b*), 所有特征值的几何重数之和为 5 ($\dim(\mathcal{V})$, 因此 $\mathcal{T}$ 可对角化。 ■

那么我们现在已经万事俱备，只欠东风了：我们一起引出最重要的对角化定理：

定理 3-11
设线性变换 $\mathcal{T}: \mathcal{V} \longrightarrow \mathcal{V}$, $\dim(\mathcal{V}) = n$。则 $\mathcal{T}$ 可对角化的充要条件是对于任何一个 $\mathcal{T}$ 的特征值 λ, 其几何重数等于其代数重数。

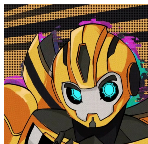
定理 3-12
任何对称矩阵 (即 $A = A^T$) 均可对角化。

这一定理的证明我们暂时忽略，在后续章节的合适位置我们会专门对这一定理进行证明，现阶段该定理可以作为一个有用的结论使用。

例 3.3.2

设 a 为任意实数，在下列矩阵中，哪一个矩阵**不可**对角化？

(A) $\begin{bmatrix} 1 & a & -1 \\ 0 & 2 & a \\ 0 & 0 & 3 \end{bmatrix}$ (B) $\begin{bmatrix} 1 & a & 0 \\ a & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix}$ (C) $\begin{bmatrix} 1 & 1 & a \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ (D) $\begin{bmatrix} 1 & 1 & a \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix}$



(A) 选项：根据上三角矩阵的性质，A 的特征值为 $1,2,3$，因此可知 A 有 3 个互不相同的特征值，根据定理 *3-8*，A 可对角化。

(B) 选项：不难发现 B 为对称矩阵，因此根据定理 *3-11*，B 可对角化。

(C) 选项：根据上三角矩阵的性质，C 的特征值为 $1,2$，其中 $\lambda=1$ 的代数重数为 1，经计算其几何重数为 1；$\lambda=2$ 的代数重数为 2，经计算其几何重数为 2。由定理 *3-10* 可知，每一个特征值的代数重数均等于几何重数，则 C 可对角化。

(D) 选项：根据上三角矩阵的性质，C 的特征值为 $1,2$，其中 $\lambda=1$ 的代数重数为 1，经计算其几何重数为 1；$\lambda=2$ 的代数重数为 2，经计算其几何重数为 1。由定理 *3-10* 可知，对于 $\lambda=2$ 而言其几何重数不等于代数重数，因此 D 不可对角化。

3.3 练习

---

1. 判断下列矩阵能否对角化：

$$A=\begin{bmatrix}1&2\\3&2\end{bmatrix}\quad B=\begin{bmatrix}2&-4\\-1&-1\end{bmatrix}\quad C=\begin{bmatrix}2&1\\4&-1\end{bmatrix}$$

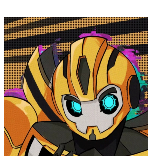
2. 特征多项式相同的 $n\times n$ 矩阵一定都可以（或都不可以）对角化吗？不妨使用 $A=\begin{bmatrix}2&3&-3\\1&0&-1\\1&1&-2\end{bmatrix}, B=\begin{bmatrix}0&1&0\\3&0&1\\2&0&0\end{bmatrix}$ 来验证这一猜想。

3. 设线性变换 $\mathcal{T}:\mathcal{V}\longrightarrow\mathcal{V}$ 满足 $\mathcal{T}^2(\boldsymbol{v})=\boldsymbol{v}$, $\forall\boldsymbol{v}\in\mathcal{V}$。证明：$\mathcal{T}$ 可对角化。

4. 设 $n\geq 1$，定义线性变换 $\mathcal{T}:\mathbb{P}_n(x)\longrightarrow\mathbb{P}_n(x)$，满足：

$$\mathcal{T}(f(x))=f(x)+f^{(1)}(x)+f^{(2)}(x)+\cdots+f^{(n)}(x)$$

其中 $f^{(k)}(x)$ 代表多项式 $f(x)$ 的 k 阶导数。判断 $\mathcal{T}$ 能否对角化。



3.4 线性递归系统

在自然界中有非常多的随时间而变化的变量。例如某物种种群数量随时间而变化；中矿石放射性元素随含量时间增长而衰减等。我们可以尝试用线性变换的知识去理解这些问题：我们可以将所要研究的变量当做一个向量 $\boldsymbol{v}_0$, 然后通过已知条件我们也可以求出一个变换矩阵 A, 这样一来 $\boldsymbol{v}_1 = A\boldsymbol{v}_0$ 就可以表示为该变量经过时间等因素的影响，在下一状态时的数量。我们也可以利用该模型来预测其未来的变化趋势。我们一起来看一个例子：

假设存在某鸟类种群，在该鸟类种群中我们将其分为幼年鸟和成年鸟两种。并且经过一段时间的调查我们得出以下关系：

① 每一年新孵化的幼年鸟的数量是前一年存活下来的成年鸟的数量的两倍；

② 每一年中约有 $\dfrac{1}{2}$ 的成年鸟能够存活下来；

③ 每一年中约有 $\dfrac{1}{4}$ 的幼年鸟能够存活下来并且成长为成年鸟

现在经调查得知该种群有 100 只成年鸟和 40 只幼年鸟，那么试问在许多年后该鸟类种群数目会发生何等变化？

我们可以设幼年鸟在第 k 年的数量为 j_k, 设成年鸟在第 k 年的数量为 a_k, 那么我们可以根据上述关系，得到：

$$\begin{cases} a_{k+1} = \dfrac{1}{2}a_k + \dfrac{1}{4}j_k \\ \\ j_{k+1} = 2a_k \end{cases}$$

此时我们设 $\boldsymbol{v}_k = \begin{bmatrix} a_k \\ j_k \end{bmatrix}$, 设 $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}$, 那么根据题设关系，我们就有

$$\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k$$

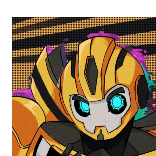
因此，结合初始条件 $\boldsymbol{v}_0 = \begin{bmatrix} 100 \\ 40 \end{bmatrix}$, 我们还可以得到

$$\boldsymbol{v}_k = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}^k \begin{bmatrix} 100 \\ 40 \end{bmatrix}$$

接下来，我们便一起研究一下矩阵 A。其中经过计算我们得出 A 的特征多项式为

$$C_A(x) = (x-1)(x+\frac{1}{2})$$

因此 A 有两个不同的特征值 $\lambda_1 = 1, \lambda_2 = -\frac{1}{2}$。那么根据定理 3-9 可知，$A$ 可对角化。且我们可以计算出 $\lambda_1 = 1$ 对应的特征向量 $\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$；$\lambda_2 = -\frac{1}{2}$ 对应的特征向量 $\boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$。



也就是说，存在可逆矩阵 $\mathcal{P}=\begin{bmatrix}1 & -1\\ 2 & 4\end{bmatrix}$ 与对角矩阵 $\mathcal{D}=\begin{bmatrix}1 & 0\\ 0 & -\frac{1}{2}\end{bmatrix}$, 使得

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

即

$$A^k=\mathcal{P}\mathcal{D}^k\mathcal{P}^{-1}=\begin{bmatrix}1 & -1\\ 2 & 4\end{bmatrix}\begin{bmatrix}1^k & 0\\ 0 & (\frac{1}{2})^k\end{bmatrix}\begin{bmatrix}\frac{2}{3} & \frac{1}{6}\\ -\frac{1}{3} & \frac{2}{3}\end{bmatrix}$$

$$=\frac{1}{6}\begin{bmatrix}4+2(-\frac{1}{2})^k & 1-(-\frac{1}{2})^k\\ 8-8(-\frac{1}{2})^k & 2+4(-\frac{1}{2})^k\end{bmatrix}$$

即

$$\boldsymbol{v}_k=\begin{bmatrix}a_k\\ j_k\end{bmatrix}=A^k\boldsymbol{v}_0=\begin{bmatrix}4+2(-\frac{1}{2})^k & 1-(-\frac{1}{2})^k\\ 8-8(-\frac{1}{2})^k & 2+4(-\frac{1}{2})^k\end{bmatrix}\begin{bmatrix}100\\ 40\end{bmatrix}$$

$$=\frac{1}{6}\begin{bmatrix}440+160(-\frac{1}{2})^k\\ 880-640(-\frac{1}{2})^k\end{bmatrix}$$

所以我们得到

$$a_k=\frac{220}{3}+\frac{80}{3}(-\frac{1}{2})^k \qquad j_k=\frac{440}{3}-\frac{320}{3}(-\frac{1}{2})^k$$

此时，当 k 很大时，我们知道 $(-\dfrac{1}{2})^k\to 0$, 因此在很多年之后，种群的数量会趋于稳定，即

$$\begin{cases}a_k\approx \frac{220}{3}\\ j_k\approx \frac{440}{3}\end{cases}$$

定义 3-6

在 $n\times n$ 矩阵 A 中，如果存在满足条件的向量和非负整数 k, 使得

$$\boldsymbol{v}_{k+1}=A\boldsymbol{v}_k$$

我们称该系统为一个**线性递归系统** ***(Linear Dynamical System)***。

我们设在线性递归系统 $\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k$ 中，矩阵 A 可对角化，设 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$；与之对应的特征向量为 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$。设矩阵 $\mathcal{P} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$, $\mathcal{D} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, 那么

$$A^k \boldsymbol{v}_0 = (\mathcal{P}\mathcal{D}^k\mathcal{P}^{-1})\boldsymbol{v}_0 = \mathcal{P}\mathcal{D}^k(\mathcal{P}^{-1}\boldsymbol{v}_0)$$

我们令 $\boldsymbol{b} = \mathcal{P}^{-1}\boldsymbol{v}_0 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$，则

$$\begin{aligned} \boldsymbol{v}_k &= \mathcal{P}\mathcal{D}^k(\mathcal{P}^{-1}\boldsymbol{v}_0) \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} \mathrm{diag}(\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k) \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} b_1\lambda_1^k \\ b_2\lambda_2^k \\ \vdots \\ b_n\lambda_n^k \end{bmatrix} \end{aligned}$$

因此，我们也得出了一个公式：

$$\boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{u}_1 + b_2\lambda_2^k\boldsymbol{u}_2 + \cdots + b_n\lambda_n^k\boldsymbol{u}_n$$

其中

$$\boldsymbol{v}_0 = b_1\boldsymbol{u}_1 + b_2\boldsymbol{u}_2 + \cdots + b_n\boldsymbol{u}_n$$

此时，不失一般性地讲，我们不妨假设 λ_1 是 A 中的一个**主导特征值 (Dominant Eigenvalue)**，我们给出以下定义：

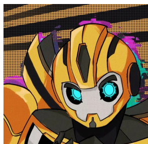
定义 3-7
在 $n \times n$ 矩阵 A 中，我们称 λ 为矩阵 A 的主导特征值，如果其代数重数为 1，且 λ 是所有特征值里的最大值。

那么此时我们便有

$$\boldsymbol{v}_k = \lambda_1^k \left[b_1\boldsymbol{u}_1 + b_2\left(\frac{\lambda_2}{\lambda_1}\right)^k \boldsymbol{u}_2 + \cdots + b_n\left(\frac{\lambda_n}{\lambda_1}\right)^k \boldsymbol{u}_n \right]$$

当 k 很大时，我们便会发现 $\left(\dfrac{\lambda_j}{\lambda_1}\right)^k \to 0 \ \ (j > 1)$, 所以决定 $\boldsymbol{v}_k$ 的其实只与主导特征值有关，即

$$\boldsymbol{v}_k \approx \lambda_1^k b_1 \boldsymbol{u}_1$$



定理 3-13
在线性递归系统 $\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k\ (k > 0)$ 中，假设 $\boldsymbol{v}_0, A$ 均给定，且 A 可对角化。设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 A 的特征值；$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ 为相对应的特征向量，令 $\mathcal{P} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$，那么我们有

$$\boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{u}_1 + b_2\lambda_2^k\boldsymbol{u}_2 + \cdots + b_n\lambda_n^k\boldsymbol{u}_n$$

其中

$$\boldsymbol{b} = \mathcal{P}^{-1}\boldsymbol{v}_0 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

如果 A 还存在主导特征值 λ_i，那么

$$\boldsymbol{v}_k = \lambda_i^k b_i \boldsymbol{u}_i \quad \text{（对很大的 } k \text{ 而言）}$$

前面几页的推理便是对该定理的证明。

例 3.4.1

在某线性递归系统中，给定 $x_0 = 1, x_1 = -1$，且

$$x_{k+2} = -x_{k+1} + 2x_k \quad (k > 0)$$

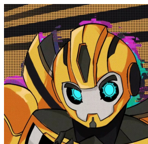
据此求 x_k 的通项。

我们由题意可设 $\boldsymbol{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix}$，由 $\boldsymbol{v}_{k+1} = \begin{bmatrix} x_{k+1} \\ x_{k+2} \end{bmatrix} = \begin{bmatrix} x_{k+1} \\ 2x_k - x_{k+1} \end{bmatrix} = A\boldsymbol{v}_k$ 可以求出其变换矩阵为

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

经计算，A 的特征值为 $\lambda_1 = -2; \lambda_2 = 1$，其对应的特征向量为 $\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}; \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$，则 $\mathcal{P} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}, \boldsymbol{b} = \mathcal{P}^{-1}\boldsymbol{v}_0 = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$，由此可得

$$\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = b_1\lambda_1^k\boldsymbol{u}_1 + b_2\lambda_2^k\boldsymbol{u}_2 = (-2)^k \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} + 1^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{即} x_k = \frac{1}{3}\left[2(-2)^k + 1\right]$$



形如 $\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k$ 的线性递归系统还有一种特殊情况：如果矩阵 $A\boldsymbol{v}_k$ 代表的是下一阶段 ($\boldsymbol{v}_{k+1}$) 所发生的事件的概率，我们把这种结构也可以称作**马尔可夫链（Markov Chain)**。

定义 3-8
在一个变化的系统中，如果该变化的发展仅仅取决于该系统当前的状态，而不取决于该系统之前的状态，我们就将该系统称作为一个马尔可夫链。

在一个马尔可夫链中，我们可以假设这个系统存在 n 中不同的状态，并且该系统当前处于状态 j，那么我们也会得到在该状态下，该系统下一状态转化为状态 k 的概率 p_{kj} (如下图) ，这样一来所得到的 $n \times n$ 矩阵 $P = (p_{ji})$ 便被称作是**过渡矩阵 (Transition Matrix)**。

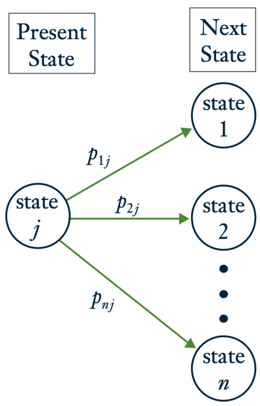


图 8

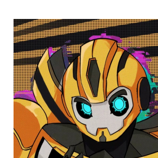
根据定义，矩阵 P 的第 j 列为 $\begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}$, 不难发现

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1$$

其中 p_{kj} 即表示在当前状态为 j 时，经过一次变化到状态 k 的概率。其中 $0 < p_{kj} < 1$, 故 P 也被称作**随机矩阵 (Stochastic Matrix)**。我们设 $s_i^{(m)}$ 代表系统经过 m 次变化之后 (或者说在第 m 阶段) 最终处于状态 i 的概率，那么此时定义

$$\boldsymbol{s}_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \\ \vdots \\ s_n^{(m)} \end{bmatrix} \quad (m = 0, 1, 2 \cdots)$$

我们将 $\boldsymbol{s}_m$ 称作**状态向量 (State Vector)**，其中 $\boldsymbol{s}_0$ 称作**初始状态 (Initial State)**, 在一个马尔可夫链中，$P, \boldsymbol{s}_0$ 均为已知量。



定理 3-14
设 P 为一个有 n 个状态的马尔可夫链的随机矩阵（过渡矩阵），若 $\boldsymbol{s}_m$ 代表该马尔可夫链在第 m 阶段时的状态向量，则

$$\boldsymbol{s}_{m+1} = P\boldsymbol{s}_m = P^{m+1}\boldsymbol{s}_0$$

该定理的证明我们暂时忽略。

例 3.4.2

在一处草原中，某狼群会对 R_1, R_2, R_3 三个不同位置的羊群进行攻击，已知该狼群的攻击有以下规律：

① 若该狼群在某一天攻击了一处羊群，那么下一天该狼群仍攻击该羊群的概率和不攻击该羊群的概率相等;

② 若该狼群在某一天攻击了 R_1 处的羊群，那么下一天该狼群不会攻击 R_2 处的羊群;

③ 若该狼群在某一天攻击了 R_2 或 R_3 处的羊群，那么下一天该狼群攻击 R_2 和 R_3 处的羊群的概率相等。

现在已知该狼群在周一攻击了 R_1 处的羊群，那么在周四该狼群仍攻击 R_1 处的羊群的概率是多少?

根据题意，我们可以得出以下的关系表 (该表格即为过渡矩阵)：

	R_1	R_2	R_3
R_1	0.5	0.25	0.25
R_2	0	0.5	0.25
R_3	0.5	0.25	0.5

$$P = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0 & 0.5 & 0.25 \\ 0.5 & 0.25 & 0.5 \end{bmatrix}$$

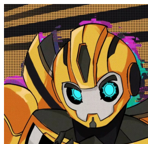
我们根据第一条规律便可以得知，$p_{11} = p_{22} = p_{33} = 0.5$。$P$ 的第一列表示当前状态为 R_1 时，下一状态的概率，因此由第二条规律可知 $p_{21} = 0$，再由其列中元素之和为 1 可知 $p_{31} = 0.5$。随后再利用规律 3，便可以求出 P 中的全部元素。我们设周一时对应的状态为 $\boldsymbol{s}_0$，根据定义，$\boldsymbol{s}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$，那么我们就可以知道 $\boldsymbol{s}_3$ 即可表示周四时的状态向量。我们有

$$\boldsymbol{s}_3 = P^3(\boldsymbol{s}_0) = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0 & 0.5 & 0.25 \\ 0.5 & 0.25 & 0.5 \end{bmatrix}^3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{11}{32} \\ \frac{6}{32} \\ \frac{15}{32} \end{bmatrix}$$

由此得知在周四该狼群仍然攻击 R_1 处的羊群的概率为 $\dfrac{11}{32}$。

在例 3.4.1 中，我们发现当经过很长一段时间之后，鸟类种群中成年鸟和幼年鸟的数目趋于稳定。对于一些马尔可夫链而言，我们也存在一个使得状态向量 $\boldsymbol{s}$ 收敛于某一位置的向量 $\boldsymbol{x}$。此时马尔可夫链趋于稳定，我们有

$$\boldsymbol{x} = P\boldsymbol{x}$$



我们称满足条件的 $\boldsymbol{x}$ 为**稳定态向量 (Steady-State Vector)** 我们对该等式进行移项，得到

$$(I-P)\boldsymbol{s}=\boldsymbol{0}$$

需要注意的是并不是所有的马尔可夫链都具备这个性质，只有**正则矩阵 (Regular Matrix)**）对于该结论才成立。正则矩阵 P 指的是存在正整数 m，使得 P^m 的每一个元素均大于零的矩阵。

定理 3-15

设 P 为一个有 n 个状态的马尔可夫链的随机矩阵（过渡矩阵），设 P 同时为正则矩阵，那么存在唯一的向量 $\boldsymbol{s}$，使得

$$\boldsymbol{s}=P\boldsymbol{s}$$

其中，$\boldsymbol{s}$ 中的元素均为正数且元素之和为 1。同时由状态向量构成的数列 $\boldsymbol{s}_0,\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots$ 收敛于 $\boldsymbol{s}$。

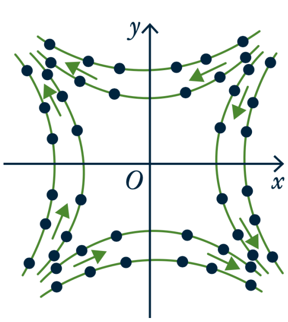
3.4 练习

---

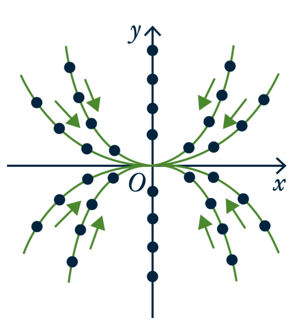
1. 在线性递归系统 $\boldsymbol{v}_{k+1}=A\boldsymbol{v}_k$ 中，我们给出

$$A_1=\begin{bmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{3}\end{bmatrix}\quad A_2=\begin{bmatrix}\frac{3}{2} & 0\\ 0 & \frac{4}{3}\end{bmatrix}\quad A_3=\begin{bmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{bmatrix}$$

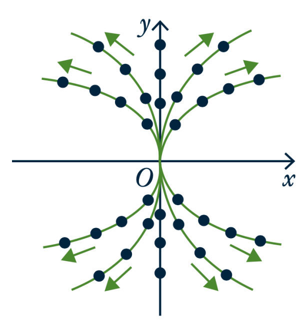
如果我们将 $\boldsymbol{v}_0,\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots$ 看作数列，依次标在坐标系中，用箭头来表示其变化趋势。那么下面的三个选项中哪一幅图可以准确地对应上面的矩阵？



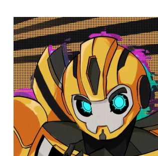
(a)



(b)



(c)



2. 甲要爬一段有 k 级台阶的楼梯，已知甲每一步可以爬一级或两级台阶，要求最后甲可以刚好达到最顶端。记 s_k 为甲刚好爬上台阶所用的不同种类数，当 $k=1$ 时，我们有 $s_1=1$，$k=2$ 时有 $s_2=2$。据此求出 s_k 的通项公式。(*HINT*：我们有 $s_{k+2}=s_{k+1}+s_k$，据此建立线性递归模型进行求解。)

3. 假设有足够多的红色，蓝色，白色木板，将这些木板从下往上堆成一叠。设 x_k 表示当木板为 k 个时，所有的使白色木板彼此不相邻的堆叠方式。据此求出 x_k 的通项公式。(*HINT*：我们有 $x_{k+2}=2x_{k+1}+2x_k$，据此建立线性递归模型进行求解。并且可以参考第二题！)

4. 在一个核反应堆中含有 α 粒子和 β 粒子。每经过一秒就会有一个 α 粒子分解为三个 β 粒子；同时也会有一个 β 粒子分解为一个 α 粒子和两个 β 粒子。在某一时刻该封闭系统内有一个 α 粒子，那么在经过 $t=20$ 秒之后该系统里面有多少个 α 粒子和 β 粒子?

5. 在某机器学习模型中，我们给出如下图所示的网格迷宫。某 *AI* 需要从一个位置出发，然后通过图中标出的路径进行节点之间的移动从而到达下一相邻位置。已知 *AI* 在每一处特定节点位置时所选择的路径概率相等。

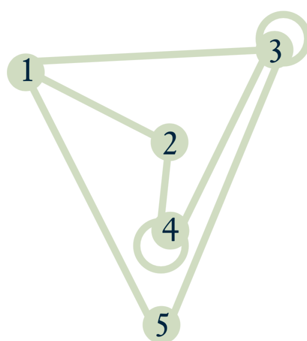


图 10: 某机器学习网格迷宫

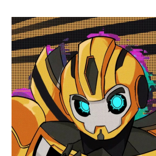
① 假设在某一时刻 *AI* 位于节点 1，据此求出经过三次移动之后，该 *AI* 又回到节点 1 处的概率；

② 假设在某一时刻 *AI* 位于节点 2，那么在经过次数足够多的移动之后该 *AI* 最有可能处于哪一处节点？

6. 设 $0<p<1; 0<q<1$ 及随机矩阵 $P=\begin{bmatrix}1-p & q\\ p & 1-q\end{bmatrix}$。

① 证明：$\dfrac{1}{p+q}\begin{bmatrix}q\\ p\end{bmatrix}$ 为 P 的稳定态向量；

② 证明：P^m 收敛于 $\dfrac{1}{p+q}\begin{bmatrix}q & q\\ p & p\end{bmatrix}$。(*HINT*：先尝试用数学归纳法证明 $P^m=\dfrac{1}{p+q}\begin{bmatrix}q & q\\ p & p\end{bmatrix}+\dfrac{(1-p-q)^m}{p+q}\begin{bmatrix}p & -q\\ -p & q\end{bmatrix}$)



4 作用于内积空间的变换

在前面三章的学习中，我们主要学习了作用于线性空间的线性变换。在本章中我们将引入一种全新的空间：**内积空间** ***(Inner Product Space)***。内积空间与向量空间存在着诸多的相同之处和不同之处，我们也可以理解为内积空间是一定程度上对向量空间进行扩充而得到的。在内积空间中，我们所谈论的数域要进行扩充：引入复数域 $\mathbb{C}$, 在此基础上我们会重新定义内积空间，以及内积空间的性质，最后我们也会从线性变换的角度去审视内积空间。

4.1 内积空间的基本定义与性质

正如本章序言所讲，在本章中我们所考虑的数域 $\mathbb{F}$ 为复数域 $\mathbb{C}$ 或实数域 $\mathbb{R}$。

内积空间我们可以理解为是在线性空间的基础上增加了增加了一种二元运算 $f:(\boldsymbol{v}_1,\boldsymbol{v}_2)\mapsto\mathbb{F}$:

$$f:\mathcal{V}\times\mathcal{V}\longrightarrow\mathbb{F}$$

如果设 $\mathcal{V}$ 为一个内积空间，同时 $\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}$, 那么这种二元运算 f 也被我们被我们计作 $\langle\boldsymbol{v}_1,\boldsymbol{v}_2\rangle$, 且该二元运算满足特定的性质。我们于是给出内积空间的定义：

定义 4-1

定义作用于复数域 $\mathbb{C}$ 上的线性空间 $\mathcal{V}$，若 $\mathcal{V}$ 中存在二元运算 $\langle\ ,\ \rangle$, 且对于任意的 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathcal{V}$ 以及 $\lambda\in\mathbb{C}$, 满足：

① $\langle\boldsymbol{u}+\boldsymbol{v},\boldsymbol{w}\rangle=\langle\boldsymbol{u},\boldsymbol{w}\rangle+\langle\boldsymbol{v},\boldsymbol{w}\rangle$;

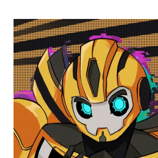
② $\langle\lambda\boldsymbol{u},\boldsymbol{v}\rangle=\lambda\langle\boldsymbol{u},\boldsymbol{v}\rangle$;

③ $\langle\boldsymbol{u},\boldsymbol{v}\rangle=\overline{\langle\boldsymbol{v},\boldsymbol{u}\rangle}$ （交换两个向量的顺序，所得结果互为共轭）;

④ $\langle\boldsymbol{u},\boldsymbol{u}\rangle\geq 0$, 不等式取得“=”当且仅当 $\boldsymbol{u}=\boldsymbol{0}$

此时我们称具备这种运算 f 的线性空间为一个**内积空间** ***(Inner Product Space)***。

在内积空间中，我们同样给出**模 (Norm)** 的定义：



定义 4-2
在内积空间 $\mathcal{V}$ 中，设 $\boldsymbol{v} \in \mathcal{V}$, 我们定义 $\boldsymbol{v}$ 的**模 (Norm)** (记作 $\|\boldsymbol{v}\|$) 为

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

若 $\|\boldsymbol{v}\| = 1$, 我们将其称作**单位向量 (Unit Vector)**。

在定义 4-1 的第三条里面，我们发现内积空间中具有共轭这一性质的存在，通过对该条性质进行深入研究，我们可以得出几条额外的性质：

定理 4-1
定义作用于复数域 $\mathbb{C}$ 上的内积空间 $\mathcal{V}$，则对于任意的 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$ 以及 $\lambda \in \mathbb{C}$, 满足：

① $\langle \boldsymbol{u}\ \boldsymbol{v} + \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle$;

② $\langle \boldsymbol{u}, \lambda \boldsymbol{v} \rangle = \overline{\lambda} \langle \boldsymbol{u}, \boldsymbol{v} \rangle$;

③ $\langle \boldsymbol{0}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{0} \rangle = 0$

④ $\|\lambda \boldsymbol{v}\| = |\lambda| \cdot \|\boldsymbol{v}\|$

该定理的证明将留作练习。

接下来我们来看一下这种二元运算的实际意义：

① 当我们考虑 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}; \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ 时，有

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

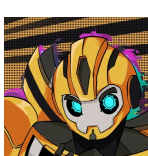
特别地，当 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ 时，这种运算也被我们称作是向量的**数量积 (Dot Product)**，记作 $\boldsymbol{x} \cdot \boldsymbol{y}$ 特别地，在该种情况下如果我们设 $\boldsymbol{x}, \boldsymbol{y}$ 之间的夹角为 θ, 那么有

$$\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos(\theta)$$

② 当我们考虑作用于多项式上的内积空间时，设 $f(x), g(x) \in \mathcal{V}$ 且 $f(x), g(x)$ 为定义在 $[0,1]$ 上的连续光滑函数, 那么我们定义

$$\langle f(x), g(x) \rangle = \int_0^1 f(x) \overline{g(x)} dx$$

或者, 对于定义域为 $\mathbb{R}$ 的连续光滑函数 $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n; g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$



而言，我们还可以定义

$$\langle f(x), g(x)\rangle = \sum_{i=1}^{n} a_i\overline{b_i}$$

③ 设 A, B 为矩阵，我们定义矩阵 A 的共轭矩阵为 $\overline{A} = \overline{(a_{ij})}$，同时定义 A 的**伴随矩阵 (Adjoint Matrix)**$A^* = (\overline{A})^T$。那么在乘法运算有意义时，我们定义

$$\langle A, B\rangle = \operatorname{tr}(B^*A)$$

其中 $\operatorname{tr}(A)$ 表示矩阵的**迹 (Trace)**，其代表矩阵主对角线上全部元素之和。

我们此时很想知道，这种二元运算 $\langle\ ,\ \rangle$ 与模有什么样的关系？我们提出一条非常重要且完美的不等式：柯西不等式。

定理 4-2：(Cauchy-Banyakovski-Schwarz 不等式)
设 $\mathcal{V}$ 为一个内积空间且 $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$，则

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

不等式取得 "=" 当且仅当向量 $\boldsymbol{u} = k\boldsymbol{v}$ (即 $\boldsymbol{u}, \boldsymbol{v}$ 共线。)

柯西不等式的证明我们在此忽略。不过我们可以利用柯西不等式来证明另外一条重要的不等式：**三角不等式 (Triangle Inequality)**。

定理 4-3：(三角不等式)
设 $\mathcal{V}$ 为一个内积空间且 $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$，则

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$

不等式取得 "=" 当且仅当向量 $\boldsymbol{u} = k\boldsymbol{v}$ (即 $\boldsymbol{u}, \boldsymbol{v}$ 共线。)

定理 4-3 的证明：

证明. 我们只需证明

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 \leq (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2$$

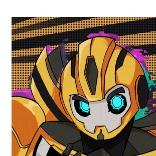
即

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}\rangle \leq \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

即

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle} \leq \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

根据共轭复数的性质，$\|\boldsymbol{u}\| = \|\overline{\boldsymbol{u}}\|$，因此通过柯西不等式即得证。 ■



例 4.1.1

设 $a_1, a_2, \cdots, a_n \in \mathbb{R}$, 证明:

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq n\left(\sum_{i=1}^{n} a_i^2\right)$$

证明. 我们不妨设 $\boldsymbol{x} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}; \boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ a \end{bmatrix}$, 那么我们有

$$(\langle \boldsymbol{x}, \boldsymbol{y} \rangle)^2 = (a_1 + a_2 + \cdots + a_n)^2; \|\boldsymbol{y}\|^2 = n$$

由柯西不等式

$$(\langle \boldsymbol{x}, \boldsymbol{y} \rangle)^2 \leq \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2$$

我们便可以得到

$$(a_1 + a_2 + \cdots + a_n)^2 \leq n(a_1^2 + a_2^2 + \cdots + a_n^2)$$

由此原不等式得证。 ■

在内积空间中, 我们再给出**距离 (Distance)** 的概念:

定义 4-3
在内积空间 $\mathcal{V}$ 中, 设 $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$, 我们定义向量 $\boldsymbol{x}, \boldsymbol{y}$ 之间的距离 $d(\boldsymbol{x}, \boldsymbol{y})$ 为:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$$

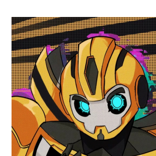
距离还满足以下的性质:

定理 4-4

① $d(\boldsymbol{x}, \boldsymbol{x}) \geq 0$, 不等式取得 “=” 当且仅当 $\boldsymbol{x} = \mathbf{0}$;

② $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$;

③ $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$, 即向量之间的距离满足三角不等式。



若存在一个集合 $\mathcal{S}$, 使得集合 $\mathcal{S}$ 中的元素满足定理 4-4 中的三条内容, 我们也称该集合 $\mathcal{S}$ 为一个**度量空间 (Metric Space)**。在度量空间中，我们同样用模来定义向量的长度，用 $d(\boldsymbol{x}, \boldsymbol{y})$ 来定义向量 $\boldsymbol{x}, \boldsymbol{y}$ 之间的距离 (即 $\boldsymbol{x}-\boldsymbol{y}$) 的模。但是值得注意的是在一般的度量空间中，模的计算有所不同。我们现在考虑向量 $\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, 设 $p\in\mathbb{R}$，我们定义向量 $\boldsymbol{x}$ 的 $\mathcal{L}_p$ **模 ($\mathcal{L}_p$ Norm)** 为

$$\|\boldsymbol{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$$

不难发现，在我们一般的计算中，我们默认 $p=2$。

3.4 练习

---

1. 设 $\mathcal{V}$ 为一个内积空间，且 $\boldsymbol{v}, \boldsymbol{u}\in\mathcal{V}$。

① $2\|\boldsymbol{v}\|^2 + 2\|\boldsymbol{u}\|^2 = \|\boldsymbol{v}+\boldsymbol{u}\|^2 + \|\boldsymbol{v}-\boldsymbol{u}\|^2$ *(该性质也被称作是向量的平行四边形法则)*

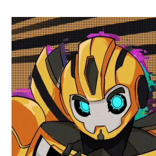
② 在 $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$ 中，我们定义 $\mathbf{Re}(\langle \boldsymbol{v}, \boldsymbol{u}\rangle)$ 为其实部，$\mathbf{Im}(\langle \boldsymbol{v}, \boldsymbol{u}\rangle)$ 为其虚部。证明：

$$\mathbf{Re}(\langle \boldsymbol{v}, \boldsymbol{u}\rangle) = \frac{1}{2}\left(\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2 - \|\boldsymbol{v}-\boldsymbol{u}\|^2\right)$$

以及

$$\mathbf{Im}(\langle \boldsymbol{v}, \boldsymbol{u}\rangle) = \frac{1}{2}\left(\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2 - \|\boldsymbol{v}-i\boldsymbol{u}\|^2\right)$$

2.



4.2 内积空间的正交性

在本节中，我们将讨论内积空间中的一个非常重要的性质：**正交性 (Orthogonality)**。我们首先给出正交的定义：

定义 4-4
在内积空间 $\mathcal{V}$ 中，设 $\boldsymbol{x},\boldsymbol{y}\in\mathcal{V}$，我们定义向量 $\boldsymbol{x},\boldsymbol{y}$ **正交**，记作 $\boldsymbol{x}\perp\boldsymbol{y}$，当且仅当

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{y},\boldsymbol{x}\rangle=0$$

接下来我们定义向量的**正交投影 (Orthogonal Projection)**：

定义 4-4
在内积空间 $\mathcal{V}$ 中，设 $\boldsymbol{x},\boldsymbol{y}\in\mathcal{V}$，我们定义向量 $\boldsymbol{x}$ 在向量 $\boldsymbol{y}$ 方向上的正交投影 $\mathbf{proj}_{\boldsymbol{y}}(\boldsymbol{x})$ 为

$$\mathbf{proj}_{\boldsymbol{y}}(\boldsymbol{x})=\langle\boldsymbol{x},\boldsymbol{y}\rangle\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

不难发现，$\boldsymbol{x}-\mathbf{proj}_{\boldsymbol{y}}(\boldsymbol{x})\perp\boldsymbol{y}$ (读者可以尝试自行证明)。

此时，我们不妨假设内积空间 $\mathcal{V}$ 的一组基底为 $\beta=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$，我们给出**正交基**（**Orthogonal Basis**）的概念：

定义 4-5
设内积空间 $\mathcal{V}$ 的一组基底为 $\beta=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$，我们称 β 为一组**正交基** ***(Orthogonal Basis)***，如果其满足

$$\langle\boldsymbol{e}_i,\boldsymbol{e}_j\rangle=0\ \ (i\neq j)$$

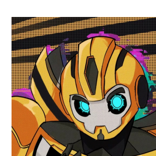
特别地，如果对任意的 $\boldsymbol{e}_k\in\beta$ 而言，$\|\boldsymbol{e}_k\|=1$，我们则称 β 为一组**单位正交基** ***(Orthonormal Basis)***。

任何一个线性空间（内积空间）均存在正交基与单位正交基。比如在 $\mathbb{R}^2$ 中，我们熟知的 $\boldsymbol{x}_1=\begin{bmatrix}1\\0\end{bmatrix};\boldsymbol{x}_2=\begin{bmatrix}0\\1\end{bmatrix}$ 便是一组单位正交基。

那么相较于其他的基底，（单位）正交基有什么优点？

①我们此时不妨设内积空间 $\mathcal{V}$ 的一组单位正交基为 $\beta=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$，那么对于任意的 $\boldsymbol{v}\in\mathcal{V}$，我们设

$$\boldsymbol{v}=c_1\boldsymbol{e}_1+c_2\boldsymbol{e}_2+\cdots+c_n\boldsymbol{e}_n$$



考虑此时 $\boldsymbol{v}$ 在 $\boldsymbol{e}_i$ 方向上的正交投影，我们根据定义便有

$$\begin{aligned}
\mathbf{proj}_{\boldsymbol{e}_i}(\boldsymbol{v}) &= \langle \boldsymbol{v}, \boldsymbol{e}_i \rangle \boldsymbol{e}_i \\
&= \langle c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + \cdots + c_n\boldsymbol{e}_n, \boldsymbol{e}_i \rangle \\
&= c_1\langle \boldsymbol{e}_1, \boldsymbol{e}_i \rangle + c_2\langle \boldsymbol{e}_2, \boldsymbol{e}_i \rangle + \cdots + c_n\langle \boldsymbol{e}_n, \boldsymbol{e}_i \rangle \\
&= c_i\langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle \\
&= c_i
\end{aligned}$$

因此我们有

$$\boldsymbol{v} = \sum_{i=1}^{n} \langle \boldsymbol{v}, \boldsymbol{e}_i \rangle \boldsymbol{e}_i = \sum_{i=1}^{n} \mathbf{proj}_{\boldsymbol{e}_i}(\boldsymbol{v})$$

其中，我们称 $\langle \boldsymbol{v}, \boldsymbol{e}_i \rangle$ 被称作是 $\boldsymbol{v}$ 的 **傅立叶系数 (Fourier Coefficients)**。

② 我们此时不妨设内积空间 $\mathcal{V}$ 的一组单位正交基为 $\beta = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$，我们此时再定义线性变换 $\mathcal{T}: \mathcal{V} \longrightarrow \mathcal{V}$。那么设线性变换在基底 β 下的矩阵形式为 $[\mathcal{T}]_\beta$，其中我们知道该矩阵的第 j 列即为 $[\mathcal{T}(\boldsymbol{e}_j)]_\beta$，由我们在①中的推理可以得到

$$[\mathcal{T}(\boldsymbol{e}_j)]_\beta = \begin{pmatrix} \langle \mathcal{T}(\boldsymbol{e}_j), \boldsymbol{e}_1 \rangle & \langle \mathcal{T}(\boldsymbol{e}_j), \boldsymbol{e}_2 \rangle & \cdots & \langle \mathcal{T}(\boldsymbol{e}_j), \boldsymbol{e}_n \rangle \end{pmatrix}^T$$

由此我们可以直接得出

$$[\mathcal{T}]_\beta = \left(\langle \mathcal{T}(\boldsymbol{e}_j), \boldsymbol{e}_i \rangle\right)_{ij} = \begin{bmatrix} \langle \mathcal{T}(\boldsymbol{e}_1), \boldsymbol{e}_1 \rangle & \cdots & \langle \mathcal{T}(\boldsymbol{e}_n), \boldsymbol{e}_1 \rangle \\ \langle \mathcal{T}(\boldsymbol{e}_1), \boldsymbol{e}_2 \rangle & \cdots & \langle \mathcal{T}(\boldsymbol{e}_n), \boldsymbol{e}_2 \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{T}(\boldsymbol{e}_1), \boldsymbol{e}_n \rangle & \cdots & \langle \mathcal{T}(\boldsymbol{e}_n), \boldsymbol{e}_n \rangle \end{bmatrix}$$

③ 另外，我们还有一条至关重要的定理：

定理 4-5

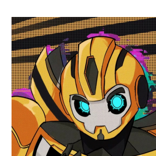
设 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 为内积空间 $\mathcal{V}$ 的一组单位正交基，设矩阵 $\mathcal{P} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$，那么

① 若此时内积空间作用的数域为 $\mathbb{R}$，我们称矩阵 $\mathcal{P}$ 为**正交矩阵** ***(Orthogonal Matrix)***，且

$$\mathcal{P}^T = \mathcal{P}^{-1}$$

② 若此时内积空间作用的数域为 $\mathbb{C}$，我们称矩阵 $\mathcal{P}$ 为**酉矩阵** ***(Unitary Matrix)***，且

$$\mathcal{P}^* = \mathcal{P}^{-1}$$



同样地，定理 4-5 的证明将留作练习供读者思考。这里给出一条证明提示：我们有 $\mathcal{P}\mathcal{P}^{-1} = I$, 依照次性质以及在①,②中提到的性质便可轻松证明。

我们此时已经知道单位正交基有着非常多的性质，那么我们该如何构建一个单位正交基呢? 如果任意给出一组基底 γ，我们该通过怎么样的运算将 γ 转变成为单位正交基呢? 我们所运用的原理即为 $\boldsymbol{x} - \mathbf{proj}_{\boldsymbol{y}}(\boldsymbol{x}) \perp \boldsymbol{y}$ 这一公式，这样一种运算也被我们称作是**施密特算法 (Gram-Schimdt Algorithm)**：

定理 4-6 (Gram-Schimdt 算法)

给定内积空间 $\mathcal{V}$ 的任意一组基底 $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$, 我们通过如下步骤的运算，使得最终得到 $\mathcal{V}$ 的一组单位正交基 $\gamma = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$:

① 我们令 $\boldsymbol{e}_1 = \dfrac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}$ (将 $\boldsymbol{v}_1$ 化为单位向量；)

② 对于随后的 $\boldsymbol{v}_k$ $(2 < k \leq n)$, 我们令

$$\boldsymbol{w}_k = \boldsymbol{v}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{v}_k, \boldsymbol{e}_i \rangle \boldsymbol{e}_i$$

③ 随后我们令 $\boldsymbol{e}_i = \dfrac{\boldsymbol{w}_k}{\|\boldsymbol{w}_k\|}$

通过上述步骤得到的 $\gamma = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n\}$ 即为一组单位正交基。

我们不难发现，$\displaystyle\sum_{i=1}^{k-1} \langle \boldsymbol{v}_k, \boldsymbol{e}_i \rangle \boldsymbol{e}_i$ 一项即为 $\boldsymbol{v}_k$ 在 $\mathbf{Span}(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_{k-1})$ 上的正交投影，那么我们便有

$$w_k \perp \mathbf{Span}(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_{k-1})$$

定义 4-6

在内积空间 $\mathcal{V}$ 中，设 $\mathcal{S} \subset \mathcal{V}$ 为一个字空间，我们定义 $\mathcal{S}$ 的**正交空间** ***(Orthogonal Subspace)*** 为 $\mathcal{S}^{\perp}$, 满足：

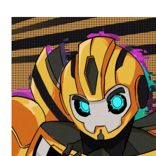
$$\mathcal{S}^{\perp} = \{\boldsymbol{v} \in \mathcal{V} : \boldsymbol{v} \perp \mathcal{S} \quad (\forall \boldsymbol{s} \in \mathcal{S}, \langle \boldsymbol{v}, \boldsymbol{s} \rangle = 0)\}$$

读者可以自行验证 $\mathcal{S}^{\perp}$ 同样为一个字空间。实际上，如果 $\mathcal{S}$ 不为字空间，$\mathcal{S}^{\perp}$ 仍然是一个字空间。

定理 4-7

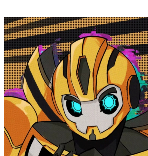
在内积空间 $\mathcal{V}$ 中，设 $\mathcal{S} \subset \mathcal{V}$ 为一个字空间，那么

$$\mathcal{V} = \mathcal{S} \oplus \mathcal{S}^{\perp}$$

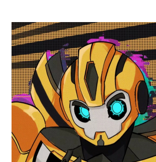


例 4.2.1

在 $\mathbb{R}^3$ 中，给定 $\mathcal{W} = \textbf{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\right)$, 据此求出 $\mathcal{W}^{\perp}$ 的一组单位正交基。



5 线性代数与常微分方程组



6 ＊线性变换的实际应用

本章为选学章节，我主要讲解了线性变换在各个领域的应用，本章各小节之间关系不大，且跳过本章不会影响后面章节的学习。读者可以根据需要选学自己感兴趣的章节。

6.1 矩阵的初等变换

回想在本书的第一节中有关高斯消元法的知识：

定理 1-1 ： 高斯消元定理
在形如 $\mathcal{A}|\boldsymbol{b}$ 的矩阵中，我们可以进行下列变换，使得最后的解不受影响
① 交换任意两行
② 将一行中的所有元素全部乘以同一非零常数 k
③ 将一行中的所有元素乘以同一非零常数 λ，再加到另外一行相对应的元素上

对于高斯消元法则里面的每一步运算，我们能否用线性变换的知识去理解？为了简化运算，我们不妨假设矩阵 $A=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$，然后我们先考虑这样的一个线性变换 $\mathcal{T}_1$，我们记该变换在标准基底下对应的矩阵形式为 E_1，满足：

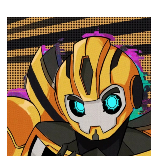
$$E_1=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}$$

此时我们把变换 $\mathcal{T}_1$ 作用在矩阵 A 上，我们得到

$$\mathcal{T}_1(A)=E_1A=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{21}&a_{22}&a_{23}\\a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

此时我们发现，$\mathcal{T}_1(A)$ 便是将矩阵 A 的第一，二行交换。也就是说，变换 $\mathcal{T}_1$ 也就对应了高斯消元里面的第一条法则，即交换任意两行。此时我们还可以给出交换矩阵 A 中第一，第三行的矩阵 E_2 以及交换 A 中第二。第三行的矩阵 E_3，我们设它们所对应的线性变换分别为 $\mathcal{T}_2,\mathcal{T}_3$。则

$$\mathcal{T}_2(A)=E_2A=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{31}&a_{32}&a_{33}\\a_{11}&a_{12}&a_{13}\\a_{11}&a_{12}&a_{13}\end{bmatrix}$$



$$\mathcal{T}_3(A)=E_3A=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\\a_{21}&a_{22}&a_{23}\end{bmatrix}$$

我们不难发现，此时矩阵 E_1,E_2,E_3 均与标准矩阵 I 十分相像，它们都是由标准矩阵 I 经过交换任意两行中的元素从而得到的新矩阵。我们此时再来看另一个线性变换 $\mathcal{S}$, 设该变换在标准基底下的矩阵形式为 E_4，满足:

$$E_4=\begin{bmatrix}\lambda&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

其中 λ 为非零常数。我们此时把该变换作用在矩阵 A 上，那么有

$$\mathcal{S}(A)=E_4A=\begin{bmatrix}\lambda&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}\lambda a_{11}&\lambda a_{12}&\lambda a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

此时，线性变换 $\mathcal{S}$ 便代表将矩阵 A 中第一行元素全部乘以一个非零常数 λ, 参照高斯消元法的定义，该变换也就对应了第二条法则，即把同一行的元素全部乘以同一个非零常数。此时矩阵 E_4 也便是标准矩阵 I 经过把第一行乘以同一常数 λ 之后所得到的新矩阵。当然，我们也可以进行线性变换的复合。考虑复合变换 $\mathcal{S}\circ\mathcal{T}_1$, 根据线性变换复合的定义，我们有

$$(\mathcal{S}\circ\mathcal{T}_1)(A)=\mathcal{S}(\mathcal{T}_1(A))=E_4(E_1A)=(E_4E_1)A$$

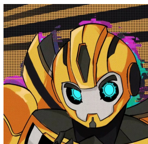
即

$$(\mathcal{S}\circ\mathcal{T}_1)(A)=\left(\begin{bmatrix}\lambda&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\right)\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

$$=\begin{bmatrix}0&\lambda&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}a_{21}&a_{22}&a_{23}\\a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

$$=\begin{bmatrix}\lambda a_{21}&\lambda a_{22}&\lambda a_{23}\\a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

此时我们不难发现，线性变换 $\mathcal{S}\circ\mathcal{T}_1$ 即代表“先 $\mathcal{T}_1$ 后 $\mathcal{S}$”，即为交换矩阵的第一，第二行，然后第一行整体乘以非零常数 λ。通过这个变换，我们是否也能够理解为什么在矩阵乘法中 AB 和 BA 不一定相等？如果是变换 $\mathcal{T}_1\circ\mathcal{S}$, 即为“先 $\mathcal{S}$ 后 $\mathcal{T}_1$, 此时的结果还一样吗？（读者不妨尝试求出此时的矩阵）。在原变换 $\mathcal{S}\circ\mathcal{T}_1$ 中，该复合变换所对应的矩阵 (E_4E_1) 便也可以看作是标准矩阵 I 经过了两步变换（先交换第一，第二行，然后第一行整体元素乘以非零常数 λ）之后所得到的结果。



我们不妨再考虑一个线性变换 $\mathcal{Q}$, 我们设 E_5 为该变换在标准基底下对应的矩阵形式，且满足

$$E_5 = \begin{bmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

那么，当我们把线性变换 $\mathcal{Q}$ 作用在矩阵 A 上时，我们有

$$\mathcal{Q}(A) = E_5 A = \begin{bmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} + \lambda a_{31} & a_{12} + \lambda a_{32} & a_{13} + \lambda a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

该变换 $\mathcal{Q}$ 即代表了将第三行中的元素乘以同一个非零常数 λ，然后加到第一行上。也就是对应了高斯消元法里面的第三条。其中 E_5 也可以看作单位矩阵 I 中，将第三行乘以同一常数 λ，然后加到第一行上所得到的结果。

我们把上文中形如 E_1, E_2, E_3, E_4, E_5 的矩阵称作是**初等矩阵 (Elementary Matrix)**，我们在此给出初等矩阵的定义。我们此时仅针对 2×2 矩阵给出定义：

定义 3-1

我们称矩阵 E 为***初等矩阵 (Elementary Matrix)***, 如果矩阵 E 可以通过对单位矩阵 I 进行**一次**高斯消元法里面的三种变换之一而得到。我们把基本初等矩阵分为三种类型，这三种类型也对应了矩阵的三种初等变换：

类型（*1*）：形如 $E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 的矩阵（对应交换第一，二行）

类型（*2*）：形如 $E = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}$ 的矩阵（对应将第二行乘以非零常数 λ）

类型（*3*）：形如 $E = \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix}$ 的矩阵（对应将第二行乘以非零常数 λ, 然后加到第一行上）

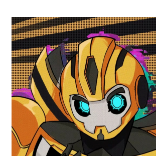
读者可以自行尝试推导剩余的几种变换矩阵，比如用第二行减去第一行的 λ 倍等。

定理 3-1

所有的初等矩阵均可逆，其逆矩阵也为初等矩阵。

该定理的证明将留给读者思考。

由此一来，我们便明白了高斯消元其实也是一个变换 $\mathcal{G} : A \longrightarrow \mathbf{RREF}(A)$，其中 $\mathcal{G}$ 便是若干个初等矩阵复合的结果。我们不妨来看一个简单的例子，我们设矩阵 $B = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$，当我们进行高斯消元的时候，我们先对第一列进行消元，即把第一行的元素全部乘以 2，然后用第二行减去第一行。那么这个变换便可以看作是两



个初等矩阵的复合。我们设 $E_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, 该矩阵即代表将第一行全部元素乘以 2；设 $E_2 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$, 该矩阵即代表用矩阵的第二行减去第一行。那么按照顺序，第一列的消元即可以表示成为

$$\begin{aligned} E_2E_1B &= \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 6 \\ 0 & -5 \end{bmatrix} \end{aligned}$$

这与我们使用高斯消元法所得到的矩阵一致。我们随后来看第二列的消元，为简化处理我们先将第一行同除 2，记 $E_3 = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$ 为该运算所代表的矩阵，那么

$$E_3(E_2E_1B) = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 0 & -5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -5 \end{bmatrix}$$

随后，我们将第一行同乘以 5，将第三行同乘以 3，记 $E_5 = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$, $E_6 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$, 随后我们把第二行加到第一行上去，记 $E_7 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 则有

$$(E_7E_6E_5E_4E_3E_2E_1)A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & -5 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -15 \end{bmatrix}$$

最后，我们将第一行同除以 5，将第三行同除以-15，记 $E_8 = \begin{bmatrix} 0.2 & 0 \\ 0 & 1 \end{bmatrix}$, $E_9 = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{15} \end{bmatrix}$，最终我们即可得到单位矩阵 I，也就是说

$$(E_9E_8E_7E_6E_5E_4E_3E_2E_1)A = I$$

由于此时 $\textbf{RREF}(A) = I$，所以矩阵 A 可逆，同时由于所有的初等矩阵均可逆，则我们进行如下的变换：

$$A = (E_9E_8E_7E_6E_5E_4E_3E_2E_1)^{-1}I$$

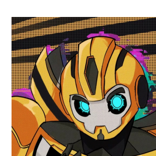
不难发现

$$A = E_1^{-1}E_2^{-1}\cdots E_8^{-1}E_9^{-1}$$

定理 3-2

若矩阵 A 可逆，那么 A 可以写成若干个基本初等矩阵的乘积，当 A 不可逆时，存在可逆矩阵 U，使得 $\textbf{RREF}(A) = UA$, 其中 U 也是若干个基本初等矩阵的乘积。

我们可以再给出一条定理 3-2 的推论：



定理 3-2 (a)

设 A 经由若干次高斯消元之后得到 B, 则:

① 存在可逆矩阵 U，使得 $B = UA$。其中 $U = E_kE_{k-1}\cdots E_2E_1$, 即为若干个初等矩阵的乘积，这些初等矩阵对应了高斯消元里面的变换。

② 对增广矩阵 $\begin{bmatrix} A & \Big| & I \end{bmatrix}$ 进行高斯消元，最终可以得到 $\begin{bmatrix} B & \Big| & U \end{bmatrix}$。

此时，我们不妨将矩阵一般化。我们设 A 为 $m \times n$ 矩阵，且 $\textbf{rank}(A) = r$。我们设 $\textbf{RREF}(A) = R$, 那么根据定理 3-2 可知，存在可逆矩阵 $U \in M_m$, 使得 $R = UA$，并且我们知道存在以下形式的变换:

$$\begin{bmatrix} A & \Big| & I_m \end{bmatrix} \longrightarrow \begin{bmatrix} R & \Big| & U \end{bmatrix}$$

由于 $\textbf{rank}(A) = r$，意味着在矩阵 R 中有 r 个前导变量，意味着 R 中的元素可以写成如下图所示的分块矩阵:

$$R \in M_{mn} = \begin{bmatrix} I_r & Y \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; R^T \in M_{nm} = \begin{bmatrix} I_r & \mathbf{0} \\ Y & \mathbf{0} \end{bmatrix}$$

那么，对 R^T 再次进行变换，我们最终可以得到形如 $\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, 那么此时即存在 $n \times n$ 可逆矩阵 U_1，使得

$$\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = U_1R^T$$

我们设 $V = U_1^T$, 此时我们有

$$UAV = RV = RU_1^T = (U_1R^T)^T = \left(\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n \times m} \right) = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n}$$

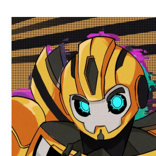
上面的推论即证明了下面这一条定理:

定理 3-3：史密斯标准形（Smith Normal Form)

设 $m \times n$ 矩阵 A 的秩为 r，则存在可逆的 $m \times m$ 矩阵 U 与可逆的 $n \times n$ 矩阵 V，使得

$$UAV = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n}$$

我们称 UAV 为矩阵 A 的 **史密斯标准型 (Smith Normal Form)**。



定理 3-3 (a)

$m \times n$ 矩阵 A 的秩为 r，其史密斯标准型为 UAV，则 U,V 可以通过下列运算得到：

① U 可以通过运算 $\left[A \mid I_m\right] \longrightarrow \left[\mathbf{RREF}(A) \mid U\right]$ 得到

② V 可以通过运算 $\left[(\mathbf{RREF}(A))^T \mid I_n\right] \longrightarrow \left[\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mid V^T\right]$ 得到

3.1 练习

---

1. 我们该如何理解初等矩阵的逆矩阵？尝试从高斯消元法则中去理解，并求出下列初等矩阵的逆矩阵。

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad E_3 = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2. 将下列矩阵表示成若干个基本初等矩阵的乘积：

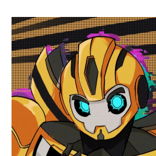
$$A = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 6 \end{bmatrix}$$

3. 若 $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}, C = \begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix}$，求出初等矩阵 E_1, E_2，使得 $C = E_2E_1A$。

4. 在本书第一章第一节中，我便提出过定理 (定理 1-3)：对于任意的矩阵 A 而言，$\mathbf{rank}(A) = \mathbf{rank}(\mathbf{RREF}(A))$。在完成本节的学习之后，尝试证明这一定理。(*HINT:* 结合行列式的知识！)

5. 假设在矩阵变换时我们有 $\left[A \mid I\right] \longrightarrow \left[P \mid Q\right]$，证明：$P = QA$。

6. 设 $A = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$，求出矩阵 A 的史密斯标准形。



6.2 统计学方面的应用：最佳近似与最小二乘法

在统计学等众多领域，我们经常会得到形如 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 的一组数据，那么我们会问一个问题：有没有一个特别的函数 $f(x)$, 使得 $f(x)$ 经过所有的这些数据点？如果我们的数据点非常完美，那么当然可以找到合适的 $f(x)$。但往往我们会发现这些点很多时候并不能全部落在同一个函数上，那么我们会想用一种“近似”的方法，来使得我们选取的 $f(x)$ 是所有可能函数里面的最优解。在高中统计学里面，我们知道回归方程便是求解这个最优解函数的方法，但今天我们将从线性变换的角度去重新审视这个问题。

那么，在给定数据点的情况下，我们该怎样去找到这个最优解 $f(x)$ 呢？这种情况下往往 $f(x)$ 的种类可以给定，比如一次函数，二次函数，对数函数等。我们为了简化模型，不妨先考虑当 $f(x) = mx + n$ 时的情况。我们知道，此时只要求出 m, n，我们便可以得出此时的最优解。设存在 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 等数据点，此时我们理想的情况便是满足下式的一个方程组：

$$\begin{cases} mx_1 + n = y_1 \\ mx_2 + n = y_2 \\ \vdots \\ mx_n + n = y_n \end{cases}$$

显然，在很多情况下这个方程组是无解的, 那么我们能否试着来取一组“最优解”呢？此时我们可以把 m, n 视作未知数，用一个向量 $\boldsymbol{x} = \begin{bmatrix} m \\ n \end{bmatrix}$ 来表示。然后设向量 $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, 矩阵 $\mathcal{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$, 那么我们有这样的形式：

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

此时，我们发现 $\mathcal{A}\boldsymbol{x}$ 便是矩阵 $\mathcal{A}$ 中列向量的线性组合，我们用线性空间 C_A 表示，也即 $C_A = \mathbf{Im}(\mathcal{A})$。由于 $\mathcal{A}$ 为 $n \times 2$ 矩阵，当 n 很大时，我们知道 $\mathbf{Im}(\mathcal{A}) \subset \mathbb{R}^n$，此时就好比空间里面的平面，会存在很多无法表示出来的向量。此时我们能够找到一组“完美”的解的充要条件即为 $\boldsymbol{y} \in \mathbf{Im}(\mathcal{A})$, 否则我们只能尝试去寻找“最优解”来减小误差。那么我们知道，$\mathcal{A}\boldsymbol{x}, \boldsymbol{b}$ 均为向量，“减小误差”这一说法在向量中即为减小向量 $\mathcal{A}\boldsymbol{x} - \boldsymbol{b}$ 的模。我们可以观察下图，发现此时我们可以使得模 $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{b}\|$ 最小。(在下图中 $\boldsymbol{x}$ 被用 $\boldsymbol{z}$ 表示)

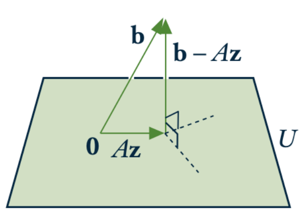
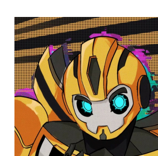


图 11: 此时模 $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{b}\|$ 取得最小值



那么，此时的 $\boldsymbol{x}$ （在图中用 $\boldsymbol{z}$ 表示）便是我们所要寻找的最优解 m, n。那么对任意的 $\mathbf{Im}(\mathcal{A})$ 中的向量 $\boldsymbol{u}$, 根据图 8，我们可以得出这样的推论：

$$(\mathcal{A}\boldsymbol{u}) \cdot (\boldsymbol{b} - \mathcal{A}\boldsymbol{x}) = 0$$

此时 $\mathcal{A}\boldsymbol{u}; \boldsymbol{b} - \mathcal{A}\boldsymbol{x}$ 均为 n 维向量，我们可以根据卷积（参见矩阵乘法一节）的性质，写成如下矩阵和矩阵相乘的形式：

$$(\mathcal{A}\boldsymbol{u})^T(\boldsymbol{b} - \mathcal{A}\boldsymbol{x}) = \mathbf{0}$$

即为

$$(\boldsymbol{u})^T(\mathcal{A})^T(\boldsymbol{b} - \mathcal{A}\boldsymbol{x}) = \mathbf{0}$$

随后我们再化回向量数量积的形式：

$$\boldsymbol{u} \cdot \left[\mathcal{A}^T(\boldsymbol{b} - \mathcal{A}\boldsymbol{x})\right] = 0$$

由于 $\boldsymbol{u}$ 的选取是任意的，因此一定有 $\mathcal{A}^T(\boldsymbol{b} - \mathcal{A}\boldsymbol{x}) = 0$, 即为

$$(\mathcal{A}^T\mathcal{A})\boldsymbol{x} = \mathcal{A}^T\boldsymbol{b}$$

此时 $\boldsymbol{x}$ 便是满足条件的最优解。上面的推论也就证明了关于最佳近似的定理：

定理 3-4：

对于线性方程组 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 而言，设该方程组的最佳近似解为 $\boldsymbol{x}_0$，则 $\boldsymbol{x}_0$ 满足：

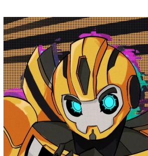
$$(\mathcal{A}^T\mathcal{A})\boldsymbol{x}_0 = \mathcal{A}^T\boldsymbol{b}$$

例 3.2.1：

经统计发现，在一个棒球队中投球手在每一次投球的时候投出好球的概率 p 与两个因素有关：1. 该投球手在球队服役的年数 x_1；2. 该投球手在上一次比赛中使得对方击球手“三振出局”的次数 x_2。现在我们有如下的表格：

投出好球的概率 p	投球手的服役年数 x_1	上局比赛令对手“三振出局”的次数 x_2
0.8	5	3
0.8	3	4
0.6	1	5
0.4	2	1

我们现在知道，p, x_1, x_2 之间满足 $p = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2$ （其中 $\lambda_1, \lambda_2, \lambda_3$ 为常数）的关系，求出此时的最佳近似方程。



例 3.2.1（续）：

由题意，我们在线性方程组 $\mathcal{A}\boldsymbol{x} = \boldsymbol{b}$ 中，有

$$\mathcal{A} = \begin{bmatrix} 1 & 5 & 3 \\ 1 & 3 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 1 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix}$$

在模型 $(\mathcal{A}^T\mathcal{A})\boldsymbol{x}_0 = \mathcal{A}^T\boldsymbol{b}$ 其中经过计算，我们有

$$\mathcal{A}^T\mathcal{A} = \begin{bmatrix} 4 & 11 & 13 \\ 11 & 39 & 34 \\ 13 & 34 & 51 \end{bmatrix} \quad \mathcal{A}^T\boldsymbol{b} = \frac{1}{5}\begin{bmatrix} 13 \\ 39 \\ 45 \end{bmatrix}$$

即我们需要求解线性方程组

$$\begin{bmatrix} 4 & 11 & 13 \\ 11 & 39 & 34 \\ 13 & 34 & 51 \end{bmatrix}\begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 13 \\ 39 \\ 45 \end{bmatrix}$$

我们使用计算器不难得到 $\lambda_0 = 0.14, \lambda_1 = 0.09, \lambda_2 = 0.08$，于是我们得到的 p 关于 x_1, x_2 的最佳近似为

$$p = 0.14 + 0.09x_1 + 0.08x_2$$

此时，我们回到最简单的形式，即给定互不相同的数据点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，我们考虑形如 $f(x) = r_0 + r_1x$ 的近似方程。对于一个可能解 $f(x)$ 而言，我们一起看下面的图 9：

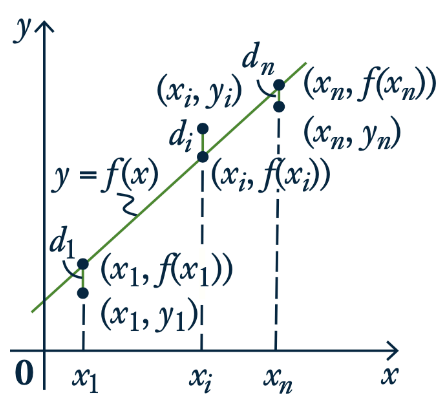
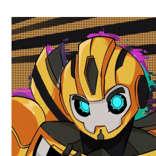


图 12

我们会发现真实的数据点 (x_i, y_i) 与我们所近似的 $(x_i, f(x_i))$ 存在误差，我们记 (x_i, y_i) 与 $(x_i, f(x_i))$ 之间的距离为 d_i，那么我们知道，满足最佳近似的条件就是使得 d_i 尽可能地小。为此，我们给定：

$$S = d_1^2 + d_2^2 + \cdots + d_n^2$$



当 S 取得最小值的时候。我们称此时 $f(x) = r_0 + r_1(x)$ 为该组数据的最佳近似方程。其中对于误差 d_i^2 而言，我们知道 $d_i^2 = [y_i - f(x_i)]^2$, 则

$$S = [y_1 - f(x_1)]^2 + [y_1 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2$$

我们不妨设 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ 以及 $f(\boldsymbol{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} r_0 + r_1x_1 \\ r_0 + r_1x_2 \\ \vdots \\ r_0 + r_nx_n \end{bmatrix}$, 那么此时根据向量的模的定义，我们有

$$S = [y_1 - f(x_1)]^2 + [y_1 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2 = \|\boldsymbol{y} - f(\boldsymbol{x})\|^2$$

于是，我们便得到了熟悉的式子，即为取模 $\|\boldsymbol{y} - f(\boldsymbol{x})\|$ 的最小值。结合我们所知道的有关线性方程组的知识，我们设

$$M = \begin{bmatrix} 1 & x_1 \\ 2 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \boldsymbol{r} = \begin{bmatrix} r_0 \\ r_1 \end{bmatrix}$$

那么不难得到

$$f(\boldsymbol{x}) = M\boldsymbol{r}$$

那么在线性方程组 $M\boldsymbol{r} = \boldsymbol{y}$ 中，为了最大程度地近似，根据定理 3-4，我们知道此时满足条件的最优解 $\boldsymbol{r}_0$ 即为

$$(M^TM)\boldsymbol{r}_0 = M^T\boldsymbol{y}$$

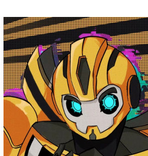
我们要求至少两个点的 x 坐标互不相同，此时 M 的列向量彼此线性无关，那么此时 M^TM 可逆（这一点读者可以尝试自行证明）。因此，我们可以直接得出

$$\boldsymbol{r}_0 = (M^TM)^{-1}M^T\boldsymbol{y}$$

定理 3-5：

对于给定且至少有两个点 x 坐标不同的数据点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，令 $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, 在所有形如 $y = z_0 + z_1x$ 的方程中，满足最佳近似的解 $\boldsymbol{z} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$ 为

$$\boldsymbol{z} = (M^TM)^{-1}M^T\boldsymbol{y}$$



例 3.2.2：

我们设数据点 $(x_1,y_1),(x_2,y_2),\cdots,(x_5,y_5)$ 的值在下表里面给出：

横坐标 x	纵坐标 y
1	1
3	2
4	3
6	4
7	5

求出关于这些数据点的最佳近似方程 $y=z_0+z_1x$。

由定理 *3-5* 我们知道最优解 $\boldsymbol{z}=\begin{bmatrix}z_0\\z_1\end{bmatrix}$ 为

$$\boldsymbol{z}=(M^TM)^{-1}M^T\boldsymbol{y}$$

其中

$$M^TM=\begin{bmatrix}1&1&1&1&1\\1&3&4&6&7\end{bmatrix}\begin{bmatrix}1&1\\1&3\\1&4\\1&6\\1&7\end{bmatrix}=\begin{bmatrix}5&21\\21&111\end{bmatrix}$$

$$M^T\boldsymbol{y}=\begin{bmatrix}1&1&1&1&1\\1&3&4&6&7\end{bmatrix}\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix}=\begin{bmatrix}15\\78\end{bmatrix}$$

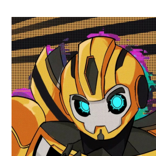
因此

$$\boldsymbol{z}=(M^TM)^{-1}M^T\boldsymbol{y}=\left(\begin{bmatrix}5&21\\21&111\end{bmatrix}\right)^{-1}\begin{bmatrix}15\\78\end{bmatrix}=\frac{1}{38}\begin{bmatrix}9\\25\end{bmatrix}$$

所以最佳近似方程为 $y=0.24+0.66x$。

那么我们现在一定会想，如果一组数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 有着形如 $y=ax^2+bx+c$ 或者其他形式的近似方程，那么我们能否还能像对待一次函数一样使用公式 $\boldsymbol{z}=(M^TM)^{-1}M^T\boldsymbol{y}$ 直接求出最优解 $\boldsymbol{z}$？我们不妨设此时该组数据的最佳近似方程为一个多项式函数：

$$y=f(x)=r_0+r_1x+r_2x^2+\cdots+r_mx^m$$



参照之前的方法，我们设:

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\quad \boldsymbol{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\quad f(\boldsymbol{x})=\begin{bmatrix}f(x_1)\\f(x_2)\\\vdots\\f(x_n)\end{bmatrix}$$

此时我们的误差为

$$S=\|\boldsymbol{y}-f(\boldsymbol{x})^2\|=[y_1-f(x_1)]^2+[y_2-f(x_2)]^2+\cdots+[y_n-f(x_n)]^2$$

我们设 $f(\boldsymbol{x})=M\boldsymbol{r}$，那么我们可以得到

$$M=\begin{bmatrix}1&x_1&x_1^2&\cdots&x_1^m\\1&x_2&x_2^2&\cdots&x_2^m\\\vdots&\vdots&\vdots&&\vdots\\1&x_n&x_n^2&\cdots&x_n^m\end{bmatrix}\quad \boldsymbol{r}=\begin{bmatrix}r_0\\r_1\\\vdots\\r_m\end{bmatrix}$$

因此，在线性方程组 $M\boldsymbol{r}=\boldsymbol{y}$ 中，我们需要求出一组最优解 $\boldsymbol{z}$, $\boldsymbol{z}$ 的求法满足以下的定理:

定理 3-6 :

给定数据点 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$, 我们设

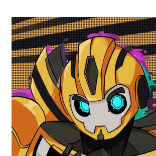
$$\boldsymbol{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\quad M=\begin{bmatrix}1&x_1&x_1^2&\cdots&x_1^m\\1&x_2&x_2^2&\cdots&x_2^m\\\vdots&\vdots&\vdots&&\vdots\\1&x_n&x_n^2&\cdots&x_n^m\end{bmatrix}\quad \boldsymbol{z}=\begin{bmatrix}z_0\\z_1\\\vdots\\z_m\end{bmatrix}$$

若多项式 $f=z_0+z_1x+z_2x^2+\cdots+z_mx^m$ 为该组数据的最佳近似解，那么

$$(M^TM)\boldsymbol{z}=M^T\boldsymbol{y}$$

我们现在所要关注的,便是 M^TM 在什么条件下可逆,一旦 M^TM 可逆,我们就可以写成 $\boldsymbol{z}=(M^TM)^{-1}M^T\boldsymbol{y}$ 的形式从而直接求出满足条件的解。我们假设存在如下的线性组合:

$$r_0\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}+r_1\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\cdots+r_m\begin{bmatrix}x_1^m\\x_2^m\\\vdots\\x_n^m\end{bmatrix}=\mathbf{0}$$



此时我们如果令 $q(x) = r_0 + r_1x + r_2x^2 + \cdots + r_mx^m$, 那么我们便知道在多项式 $q(x)$ 中,

$$q(x_1) = q(x_2) = \cdots = q(x_m) = 0$$

根据多项式基本定理, 得出 $q(x) = 0$ (现阶段我先不做解释, 下一节我们会学到 *LaGrange* 多项式, 读者到时候便可以理解该推论。), 也就意味着 M 的列向量彼此线性无关。我们知道, 若 M^TM 可逆, 则 $M^TM\boldsymbol{x} = \mathbf{0}$ 的唯一解是 $\boldsymbol{x} = \mathbf{0}$，我们考虑 $M\boldsymbol{x}$，则

$$\begin{aligned}\|M\boldsymbol{x}\|^2 &= (M\boldsymbol{x})^T(M\boldsymbol{x}) \\ &= (\boldsymbol{x})^TM^TM\boldsymbol{x}\end{aligned}$$

我们假设 $M^TM\boldsymbol{x} = \mathbf{0}$, 则 $\|M\boldsymbol{x}\|^2 = 0$, 也即 $M\boldsymbol{x} = \mathbf{0}$, 由于 M 的列向量彼此线性无关，因此 $M\boldsymbol{x} = \mathbf{0}$ 的唯一解便是 $\boldsymbol{x} = \mathbf{0}$，也即 $M^TM\boldsymbol{x} = \mathbf{0}$ 的唯一解是 $\boldsymbol{x} = \mathbf{0}$，则 M^TM 可逆。也就是说，只要 $n \geq m+1$, M^TM 均可逆。所以我们可以完善定理 3-6 中的推论：

定理 3-6 (a):

给定数据点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, 我们设

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_m \end{bmatrix}$$

若多项式 $f = z_0 + z_1x + z_2x^2 + \cdots + z_mx^m$ 为该组数据的最佳近似解，那么

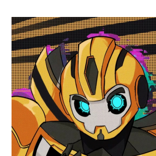
$$(M^TM)\boldsymbol{z} = M^T\boldsymbol{y}$$

同时，若 $n \geq m+1$, 此时 M^TM 可逆，因此有

$$\boldsymbol{z} = (M^TM)^{-1}M^T\boldsymbol{y}$$

我们不妨来看一种最特殊的情况：即 $n = m+1$，此时 M 的行数与列数相等。当我们给定数据点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 且 x_i 互不相同时，根据多项式定理，我们便知道此时存在唯一的 $p(x) = r_0 + r_1x + r_2x^2 + \cdots + r_{n-1}x^{n-1}$，使得 $p(x_i) = y_i$，此时的矩阵 M 为

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & \cdots & x_{n-1}^{n-1} \end{bmatrix}$$



我们把满足条件的这种 M 的行列式称作为**范德蒙特行列式 (Vandermonde Determinant)**

定理 3-7:

设 $x_1, x_2, \cdots, x_n$ 为互不相等的实数，$n \geq 2$, 那么

$$\det\left(\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & \cdots & x_{n-1}^{n-1} \end{bmatrix}\right) = \prod_{1\leq j<i\leq n}(x_i - x_j)$$

关于范德蒙德行列式的其他知识我将不做展开，有兴趣的读者可以自行查阅书籍。

在定理 3-6 中，我们得出了有关多项式的最佳近似方程，其中，多项式里面的 $1, x, x^2, \cdots, x^m$ 便可以看作是一组基底，任何 $p(x) \in \mathbb{P}_m(x)$ 便是这些基底向量的线性组合。那么对于基底的选取能否任意？我们能不能取若干个函数 $f_0(x), f_1(x), f_2(x), \cdots$ 作为基底呢？我们一起来看定理 3-8：

定理 3-8:

给定数据点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, 我们设函数 $f_0(x), f_1(x), \cdots, f_m(x)$ 已知，设

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad M = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_m(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_m(x_n) \end{bmatrix} \qquad \boldsymbol{z} = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_m \end{bmatrix}$$

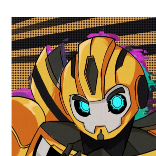
若 $g = z_0 f_0(x) + z_1 f_1(x) + \cdots + z_m f_m(x)$ 为该组数据的最佳近似解，那么

$$(M^T M)\boldsymbol{z} = M^T \boldsymbol{y}$$

同时，若 $\textbf{rank}(M) = m + 1$, 此时 $M^T M$ 可逆，因此有

$$\boldsymbol{z} = (M^T M)^{-1} M^T \boldsymbol{y}$$

因此，定理 3-8 是一般情况下的最佳近似，但此时 $M^T M$ 是否可逆就没有明显的方法进行判断了，$M^T M$ 是否可逆将很大程度上取决于函数 $f_0(x), \cdots, f_m(x)$ 的取值。



例 3.2.3：
设 $(-1,0),(0,1),(1,4)$ 为平面直角坐标系 xOy 内的点，用 $y=r_0x+r_12^x$ 为模型，求出 y 在该组数据下的最佳近似。

我们设 $f_0(x)=x$,$f_1(x)=2^x$, 那么有

$$M=\begin{bmatrix} f_0(x_1) & f_1(x_1) \\ f_0(x_2) & f_1(x_2) \\ f_0(x_3) & f_1(x_3) \end{bmatrix}=\frac{1}{2}\begin{bmatrix} -2 & 1 \\ 0 & 2 \\ 2 & 4 \end{bmatrix}$$

此时我们求得 $M^TM=\frac{1}{4}\begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix}$, 其中 M^TM 可逆，则

$$\boldsymbol{z}=\left(\frac{1}{4}\begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix}\right)^{-1}M^T\boldsymbol{y}$$

不难解出 $\boldsymbol{z}=\begin{bmatrix} \frac{10}{11} \\ \frac{16}{11} \end{bmatrix}$, 也就是说满足最佳近似的解为 $g(x)=\frac{10}{11}x+\frac{16}{11}2^x$。我们注意到 $g(-1)=-\frac{2}{11},g(0)=\frac{16}{11},g(1)=\frac{42}{11}$, 这已经与原题目中的数据点非常接近。

3.2 练习

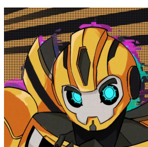
---

1. 在 $\mathbb{R}^2$ 中给出数据点 $M_1=(0,1);M_2=(-1,1);M_3=(-1,0)$, 求出形如 $ax^2+by^2-(x+y)-1=0$ 的方程在该组数据下的最佳近似。

2. 在 $\mathbb{R}^2$ 中给出数据点 $M_1=(1,95);M_2=(2,80);M_3=(3,56)$, 求出形如 $y=a_0+a_1t+a_2t^2$ 的方程在该组数据下的最佳近似。

3. 在 $\mathbb{R}^2$ 中给出数据点 $M_1=(0,3);M_2=(1,0);M_3=(1,-1);M_4=(-1,2)$, 求出形如 $y=r_0+r_1x^2+r_2\sin(\dfrac{\pi x}{2})$ 的方程在该组数据下的最佳近似。

4. 在 $\mathbb{R}^2$ 中给出数据点 $M_1=(1,1);M_2=(3,2);M_3=(4,3);M_4=(6,4)$, 求出形如 $y=z_0+z_1x$ 的方程在该组数据下的最佳近似。



6.3 算法方面的应用：FFT 与有向图

6.3.1 FFT 算法

FFT 算法也被称作**快速傅里叶变换 (Fast Fourier Transform)**, 由 Cooley 与 Tukey 于 1965 年发明。该算法的出现具有里程碑的意义，在众多领域均有着非常广泛的应用。我们今天为了方便读者理解，将讨论 FFT 的一种特殊应用：计算两个多项式的乘积。假设我们有两个次数相同多项式：

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$

$$g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_{n-1}x^{n-1}$$

我们想要计算 $f(x)g(x)$。如果按照“常规方法”的话，我们一项一项地展开，那么我们一共要做 n^2 次乘法运算，运算的次数与多项式中的项数有关。那么在数据结构中，这种算法的时间复杂度即为 $\mathcal{O}(n^2)$(*2-4 一节里面，我简单介绍了时间复杂度的概念，我们现阶段就可以简单理解为 $\mathcal{O}(k)$ 里面的 k 越小，算法越高效简便*)。要知道 $\mathcal{O}(n^2)$ 并不是一个完美的结果，尤其是在大数据模型里面往往 n 会趋于无穷大，那么 $\mathcal{O}(n^2)$ 的算法便会非常缓慢。所以我们尝试引出一个新的算法，这个新的算法便被称为 FFT, 该算法的时间复杂度为 $\mathcal{O}(n\log n)$, 不难看出这已经是一个质的飞跃了，因此 FFT 算法也是在我看来最优美的算法之一。那么我将从线性变换的角度去讲解 FFT，读者不需要任何有关数据结构方面的知识便可以大致理解。

我们知道，对于上面提到的多项式而 $f(x), g(x)$ 而言，我们有很多种不同的表示方法。其中形如 $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$ 的函数式表达便是众多表示方法里面的一种。我们在从线性代数的角度去理解：即选定一组基底（通常取标准基底）$\beta = \{1, x, x^2, \cdots, x^{n-1}\}$, 那么我们知道 $f(x)$ 也可以表示成坐标的形式，即

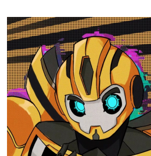
$$[f(x)]_\beta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

除此之外，我们还由代数基本定理得知，我们可以选取一组数据点：$(x_0, y_0); (x_1, y_1); \cdots (x_{n-1}, y_{n-1})$，如果我们假设 x_i 互不相同，那么我们知道存在唯一的一个多项式 $f(x)$，使得 $f(x_i) = y_i$，因此一个多项式还可以用若干个点的坐标来表示。那么对于我们上面提到的 $f(x), g(x)$ 而言，我们也可以通过选取特定的 $x_0, x_1, \cdots, x_{n-1}$ (其中 x_i 互不相同)，然后 $f(x), g(x)$ 便可以表示为：

$$f(x) \simeq \{(x_0, f(x_0)), (x_1, f(x_1)), \cdots, (x_{n-1}, f(x_{n-1}))\}$$

$$g(x) \simeq \{(x_0, g(x_0)), (x_1, g(x_1)), \cdots, (x_{n-1}, g(x_{n-1}))\}$$

那么 $f(x)g(x)$ 便可以由 $(x_i, f(x_i)g(x_i))$ 这些坐标构成。我来举出一个很简单的例子：假设 $f(x) = x^3 + x$, 那么由于 $f(x)$ 的次数为 3，我们至少需要四组坐标才能精确地表示 $f(x)$，那么此时 $f(x)f(x)$ 由于是六次多项式，便需要至少七组坐标表示。但是 $f(x)$ 本身有着非常好的性质，即 $f(x)$ 为奇函数，我们便存在一个天然的等式 $f(-x) = -f(x)$, 也就是说我们并不用真正地找到四组或七组坐标，一旦我们找到 $(x_0, f(x_0))$, 我们



也就自然锁定了另一组坐标 $(-x_0, -f(x_0))$。对于偶函数而言也是如此。出于简化考虑，我们设 n 为偶数，在

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1}, g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_{n-1}x^{n-1}$$

中，为了更好地确定 $f(x)g(x)$，我们需要至少在 $f(x), g(x)$ 上面各取 $2n-1$ 组点。令 $N = 2n-1$，我们按照上面的推理，选取这样的一组点：

$$x_1, x_2, x_3, \cdots, x_N$$

使得在这些点中有着像奇函数和偶函数那样彼此“对称”的关系。当我们提到 N 个点的对称性的时候，不难想到复平面内的“单位圆”。我们令 $\omega^N = 1$, 那么便有模均为 1 的复数 $\omega^0, \omega^1, \cdots, \omega^N$。根据复数的性质：

$$\omega_1 = \cos(\theta_1) + i\sin(\theta_1) = e^{i\theta_1} \quad \omega_2 = \cos(\theta_2) + i\sin(\theta_2) = e^{i\theta_2}$$

且

$$\omega_1\omega_2 = e^{i(\theta_1+\theta_2)}$$

那么我们得到

$$\omega^k = e^{\frac{2\pi i}{N}k}$$

下图便是一个很好的直观例子：

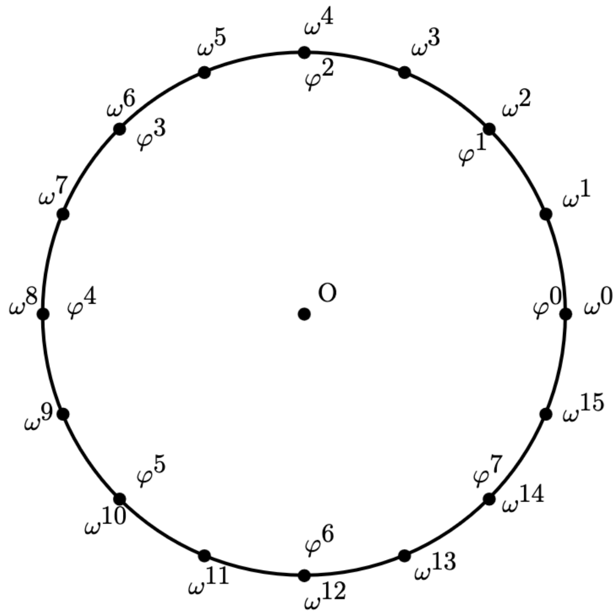


图 13: 在本图中，我们便定义了两组复数：ω, φ, 其中 $\omega^{16} = 1; \varphi^8 = 1$, 这些复数在复平面内的单位圆上的位置分布大致如图

因此，我们此时不妨令

$$f(x)g(x) = P(x) = p_0 + p_1 x + \cdots + p_{N-1}x^{N-1}$$

同时令 $x_i = \omega^i$，那么根据 3.2 一节里面讲到的最佳近似方法（不熟悉的读者请先阅读 *3.2* 小节），我们

有：

$$\begin{bmatrix} P(\omega^0) \\ P(\omega^1) \\ P(\omega^2) \\ \vdots \\ P(\omega^{N-1}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & \omega^1 & \omega^2 & \cdot & \cdot & \cdot & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdot & \cdot & \cdot & \omega^{2(N-1)} \\ \vdots & & & & \ddots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdot & \cdot & \cdot & \omega^{(N-1)^2} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{N-1} \end{bmatrix}$$

我们把其中的 $N \times N$ 矩阵记作 V，该矩阵也被称作是一个**范德蒙特矩阵 (Vandermonde Matrix)**。我们现在记

$$P(x) = P'(x) + xP''(x)$$

其中 $P'(x)$ 含有原多项式 $P(x)$ 里的所有偶数项，$xP''(x)$ 含有原多项式 $P(x)$ 里的所有奇数项，那么

$$P'(x) = p_0 + p_2x^2 + \cdots + p_{N-2}x^{N-2}$$

$$P''(x) = p_1 + p_3x^2 + \cdots + p_{N-1}x^{N-2}$$

令

$$Q'(x) = p_0 + p_2x + \cdots + p_{N-2}x^{\frac{N-2}{2}}$$

$$Q''(x) = p_1 + p_3x + \cdots + p_{N-1}x^{\frac{N-2}{2}}$$

此时，我们再设 $\varphi = \omega^2$, 那么不难发现

$$P'(\omega^j) = Q'(\omega^{2j}) = Q'(\varphi^j) \quad P''(\omega^j) = Q''(\omega^{2j}) = Q''(\varphi^j)$$

那么

$$P(\omega^j) = P'(\omega^j) + \omega^j P''(\omega^j) = Q'(\varphi^j) + \omega^j Q''(\varphi^j), \ \ (0 \leq j \leq N-1)$$

注意到此时 $(\varphi^0, \varphi^1, \varphi^2, \cdots, \varphi^{N-1})$ 实际上只表示了 $N/2$ 组不同的点，因为我们有 $\varphi = \omega^2$，且 $\varphi^k = \varphi^{k \bmod (N/2)}$。读者也可以借助图 10 来增强理解。这样一来，通过对 $P(x)$ 的拆分，我们便 "减少" 了任务量。如果我们此时再设 $\psi = \varphi^2$, 那么我们的任务量会进一步拆分。像这样的算法被称作**递归算法 (Recursion Algorithm)**, 该算法在数据结构里面十分常见，FFT 便是一个典型的递归算法。假设 T_N 代表计算全部 $P(\omega^i)$ 所消耗的时间，那么我们发现该时间一定程度上等价于两倍的计算全部 $Q(\varphi^i)$ 的时间。由于我们知道对于 ϕ 而言，数据组的数量减半，因此计算全部 $Q(\varphi^i)$ 的时间也可被记作 $T_{N/2}$。事实上，存在如下的递归关系：

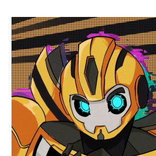
$$T_N = N + 2T_{N/2}$$

通过 **Master 定理.**[1]，我们知道，满足该算法的时间复杂度为 $\mathcal{O}(n \log n)$，这样以来我们便发现相较于我们最开始提出的 $\mathcal{O}(n^2)$ 算法而言，这些许改变已经是质的飞跃。我们现在通过该变换，便得到了 $N-1$ 个点的坐标：

$$(\omega^0, P(\omega^0)), (\omega^1, P(\omega^1)), \cdots, (\omega^{N-1}, P(\omega^{N-1}))$$

那么我们现在想做的便是找出一个唯一的多项式，使得该多项式经过上面已知的点。此时熟悉 3.2 小节

[1] 该定理由于涉及过多的数据结构和算法的知识，我将不做涉及，该定理的大致内容就是给出形如 $T_N = a \cdot T_{N/b} + f(n)$ 的递归关系的时间复杂度的通式。



的读者便会知道我们此时采用的方法。我们设 $f(x)g(x)=P(x)$ 中各次项前的系数为 p_i, 那么我们便有

$$\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{N-1} \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & \omega^1 & \omega^2 & \cdot & \cdot & \cdot & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdot & \cdot & \cdot & \omega^{2(N-1)} \\ \vdots & & & & \ddots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdot & \cdot & \cdot & \omega^{(N-1)^2} \end{bmatrix} \right)^{-1} \begin{bmatrix} P(\omega^0) \\ P(\omega^1) \\ P(\omega^2) \\ \vdots \\ P(\omega^{N-1}) \end{bmatrix}$$

其中该矩阵便为我们之前提到过的范德蒙特矩阵 V 的逆矩阵 V^{-1}。我将不过多说明为何 V 可逆，有兴趣的读者可以参考定理 3-7 然后自行尝试证明。其中，$(V^{-1})_{ij}=\omega^{-ij/N}$。我们该结论进行一个快速的证明：

我们注意到 $V_{ij}=\omega^{ij}$，其中 $0\le i,j\le N-1$, 并且我们知道 $VV^{-1}=I$，那么

$$\sum_{k=0}^{N-1}\frac{\omega^{ik}\omega^{-kj}}{N}=\frac{1}{N}\sum_{k=1}^{N-1}\omega^{k(i-j)}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$$

随后，我们引出一条性质：

$$(1+\omega^l+\cdots+\omega^{l(N-1)})(\omega^l-1)=\omega^{lN}-1=0$$

因此得证。

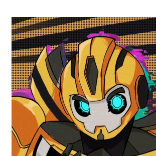
我们现在定义

$$Q(x)=\frac{P(\omega^0)}{N}+\frac{P(\omega^1)}{N}x+\cdots+\frac{P(\omega^{N-1})}{N}x^{N-1}$$

则

$$p_i=\sum_{j=0}^{N-1}\frac{P(\omega^j)}{N}\omega^{-ij}=Q(\omega^{-i})$$

随后，我们便可以通过相似的方法算出 $Q(\omega^{-i})$, 也即最后算出了 p_i，$P(x)=f(x)g(x)$ 因此确定。该算法最终的时间复杂度为 $\mathcal{O}(n\log n)$。



6.3.2 有向图

有向图（**Directed Graph**）是一种简单的数据结构，有向图主要由数个**顶点**（**Vertices**）以及连结顶点之间的**有向线段**（**Edges**）构成。我们定义 $G(V,E)$ 为一个顶点数为 V；有向线段数为 E 的有向图。如图 11 便是一个简单的有向图 $G(3,4)$。

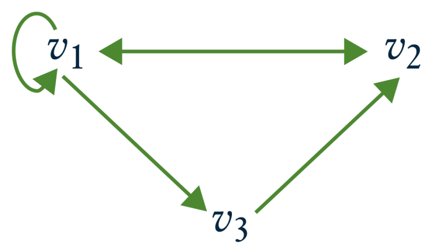


图 14: 该图即为由 3 个顶点和 4 条有向线段组成的有向图

在有向图中，我们定义由顶点 v_i 到 v_j 的一条路径的长度为 r，如果存在一条由 v_i 出发，且依次经由 r 条有向线段后能够到达 v_j。那么在图 11 中，我们便可以定义一条从 v_1 到 v_3，且长度为 4 的路径 r，这条路径 r 可以表示为

$$r = v_1 \longrightarrow v_3 \longrightarrow v_2 \longrightarrow v_1 \longrightarrow v_3$$

所以，我们发现从 v_i 到 v_j 的路径可能不止一条，长度也不一定相同。为了更好地研究这些路径，我们定义一个有向图的**邻接矩阵**（**Adjacency Matrix**）A 为一个 $V \times V$ 矩阵，其中我们定义：

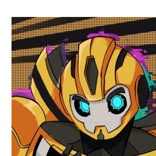
$$A_{ij} = \begin{cases} 1 & \text{存在一条由 } v_j \text{ 指向 } v_i \text{ 的有向线段} \\ 0 & \text{不存在这样的一条有向线段} \end{cases}$$

这样一来，在图 11 中的有向图 $G(3,4)$ 中，我们便能够得到这样的邻接矩阵：

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

定理 3-9:

记有向图 $G(V,E)$ 的邻接矩阵为 A，其顶点分别为 $v_1.v_2,\cdots,v_V$。则对于任意的正整数 k 而言，在矩阵 A^k 中，$(A^k)_{ij}$ 即表示所有的由 v_j 到 v_i，且长度为 k 的路径的条数。



定理 3-9 的证明:

证明. 我们将对 k 采用数学归纳法。当 $k=1$ 时，那么根据定义，定理 *3-9* 当然成立。我们现在假设定理 *3-9* 对大于 *1* 的整数 n 成立，随后我们只需要证明该定理对整数 $n+1$ 同样成立即可。此时我们知道，任何长度为 $n+1$ 的路径 r_0 均可以看作是一条长度为 n 的路径 $r_1 : v_j \longrightarrow v_k$ 与一条长度为 1 的路径 $r_2 : v_k \longrightarrow v_i$ 的复合。令

$$A=(a_{ij}), A^n=(b_{ij})$$

此时我们根据定义，我们知道一共有 a_{ik} 种 $v_k \longrightarrow v_i$ 且长度为 1 的路径；同时有 b_{kj} 种 $v_j \longrightarrow v_k$ 且长度为 n 的路径。则当我们取定 k 时，满足 $v_j \longrightarrow v_k \longrightarrow v_i$ 且长度为 $n+1$ 的路径共有 $a_{ik}b_{kj}$ 条，我们取遍所有的 k，则一共有

$$a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{in}b_{nj}$$

条 $v_j \longrightarrow v_i$ 且长度为 $n+1$ 的路径。注意到

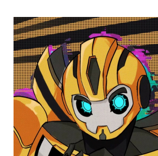
$$a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{in}b_{nj}=\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}^T \cdot \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} = (A^nA)_{ij} = A_{ij}^{n+1}$$

因此定理 *3-9* 得证。 ■

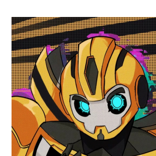
在图 11 种，我们通过计算

$$A=\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad A^2=\begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad A^3=\begin{bmatrix} 4 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

此时我们便可以直观地看出从顶点 v_i 到顶点 v_j 且路径长度为 k 的条数。



6.4　概率论方面的应用：Markov 链



7 References

1. Linear Algebra With Applications (7e) By *W. Keith Nicholson*

2. Advanced Linear Algebra With Applications By *Mohammad Ashraf, Vincenzo De Filippis, Mohammad Aslam Siddeeque*

3. Linear Algebra (Algebra 2) By *Dr. Eyal Goren*, McGill University

4. Linear Algebra (Fifth Edition) By *Stephen H. Friedberg*

5.

6. An Introduction To The FFT Algorithm (Winter 2024 Comp 252 Course Notes) By *Luc Devroye*, McGill University