"There is hardly any theory which is more elementary than linear algebra, in splite of the fact that generations of professors and textbook writers have obscued its simplicity by peposterous calculations with matrices."

—Jean Dieudonné

尽管一批教授和教科书编者用关于矩阵的荒唐至极的计算内容掩盖了线性代数的简明性，但是鲜有与之(线性代数)更为初等的理论。

—让·迪厄多内

线性代数的学习可以从两个角度入手{数值运算(Numeric operations)
几何直观(Geometric intution)

1.向量究竟是什么？

"The introduction of numbers as coordinates by reference to the particular division scheme of the open one-dimensional continuum is an act of violance."

—Hermann Weyl

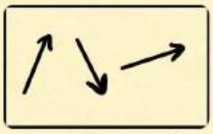
参照一维连续统的特定划分模式来引入一些数作为坐标是一种暴力的行为。

—赫尔曼·外尔

向量是什么？三种角度{物理专业学生角度(Physics student)
计算机专业学生角度(CS student)
数学家角度 (Mathematician)

物理学生角度：向量是空间中的箭头(由其长度和方向决定)，只要长度和方向相同，可以自由移动一个向量而保持它不变。

计算机学生角度：向量是有序的数字列表



物理学生角度

$$\begin{bmatrix}2\\1\end{bmatrix} \quad \begin{bmatrix}5\\0\\3\end{bmatrix} \quad \begin{bmatrix}2.3\\0.1\end{bmatrix}$$

计算机学生角度

数学家角度：尝试概括两种观点，认为向量可以是任何东西，只要保证两个向量相加以及数字与向量相乘是有意义的即可

Note：这种观点相当抽象：一先放一边，后面再提。但这一观点暗示我们：向量加法和向量数乘将贯穿线性代数始终，二者起着很重要的作用。

从物理学生角度入手，向量是箭头，将它放入坐标系中考虑(线性代数中向量通常以原点为起点)

从计算机学生角度入手，向量是有序数字列表，将它视为向量坐标，即也放到坐标系中考虑。

则每一个箭头对应一个有序数字列表，每一个有序数字列表对应一个箭头，它们一一对应。

Note：从物理和计算机学生角度入手，都可以看到向量加法和数乘对原向量的影响。

事实上，无论你怎样看待向量都无所谓(看成箭头或数字列表)，线性代数的效用很少体现在这些观点的其中一个上，而是更多地体现在它们可以在这些观点中相互转化。如：线性代数为数据分析提供了一条将大量数据列表概念化，可视化的渠道，它让数据样式变得非常明晰，并让你大致了解特定运算的意义(数据列表 → 箭头图象)另一方面，它给物理学家和计算机图形程序员提供了一种语言，让他们通过计算机能处理的数字来描述并操纵空间(箭头图象 → 数字列表)

2. 线性组合、张成的空间与基

"Mathematics requires a small dose, not of genius, but of an imaginative freedom which, in a larger dose, would be insanity."

—— Angus K. Rodgers

数学需要的不是天赋，而是少量的自由想象，但想象太过自由又会陷入疯狂。

——安古斯·罗杰斯

当看到一对描述向量的有序数对（向量坐标）时，例如 (3, -2)，将它的每个坐标看作一个标量(scalar)，即它们表示如何拉伸或压缩一个向量

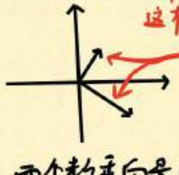
在xy-坐标系中，有两个非常特别的向量：$\hat{i}$, $\hat{j}$（x方向的单位向量与y方向的单位向量）

所以，坐标3表示将$\hat{i}$拉伸3倍，坐标-2表示将$\hat{j}$沿反方向拉伸2倍，得到的向量实际上是两个经过缩放的向量的和：$3\hat{i}-2\hat{j}$

$\hat{i}$和$\hat{j}$合起来被称为坐标系的基（它们实际上是坐标—标量所缩放的对象）

←这不是严格定义

如果我们选择不同的基向量会怎么样？选择不同的基向量，可能可以获得一个合理的新坐标系。



这样新的一组基同样允许我们在一个有序数对和二维向量之间自由转化，这样得到了一个新坐标系（不同坐标系的相互关系见后续）

需注意的是每当我们用数字描述向量时，它都依赖于我们正在使用的基。即，相同的坐标，不同的基向量，可能对应不同的向量

两个数乘向量的和被称为这两个向量的线性组合(linear combination)：$a\vec{v}+b\vec{w}$ ← scalar

向量$\vec{v}$和$\vec{w}$的所有线性组合组成的集合称为$\vec{v}$和$\vec{w}$张成的空间(span)。

Note：对二维空间而言，若给定两个向量共线非零向量，它们的span是一条直线，若不共线，span是整个空间

为了防止过多的向量充斥图形造成视觉上的拥挤，通常我们用向量的终点代表该向量。

Note：考虑单个向量时，把它当箭头，考虑多个向量时，就把它们都当点

两个三维向量张成的空间是什么样的？（不共线非零向量则得到）三维空间上的一个平面。

如果再加上第三个向量，那它们张成的空间会怎么样？两种情况：①若第三个向量恰好落在前两个向量所张成的平面上，张成的空间不变。②若第三个向量不落在平面上，张成的空间变为整个三维空间。

对第三个向量已经落入前两个向量张成的空间中或者两个向量恰好共线的情形，即一组向量中至少有一个是多余的，没有对张成空间做出任何贡献的情形，引入新术语，线性相关

若向量组中，其中一个向量能表示为其它向量的线性组合，则称这个向量组线性相关(linearly dependent)

另一方面，若向量组中的所有向量都为张成空间增添了新维度，即，每个向量都无法表示成其它向量的线性组合，则称这个向量组线性无关(linearly independent)

Note：这里是出于数学目的而引入的较直观的线性无关定义，事实上还有一个等价定义：当且仅当 $a=b=c=0$ 时，$a\vec{v}+b\vec{w}+c\vec{u}=\vec{0}$ 成立，则 $\vec{v}, \vec{w}, \vec{u}$ 是线性无关的。数学家倾向于后一种定义，因为其中所有的向量地位平等。（等价性可以被证明）

有了span，线性无关的定义，基的严格定义是？

基的严格定义：向量空间的一组基是张成该空间的一个线性无关的向量集合。

Note：空间的维度指的是空间对应的基的个数。

3.矩阵与线性变换（本节主题：线性变换的概念以及它和矩阵、矩阵乘法的关系）

"Unfortunately, no one can be told what the Matrix is. You have to see it for yousel."(Surprisingly apt words on the importance of understanding matrix operation visually.)

—Morpheus

很遗憾，矩阵是什么是说不清的，你必须自己亲眼看看（描述直观理解矩阵操作重要性的绝佳台词）

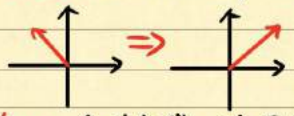
—墨菲斯

本节只集中讨论二维空间的情形（高维可自行推广）

什么是变换？

首先介绍变换：一个集合到自身的映射叫做

在线性代数中，我们考虑的是接收一个向量并且输出一个向量的变换。由于接收向量和输出向量是在同一个集合中的（变换的定义），所以可以认为变换使得输入向量移动到了输出向量。这只是变换作用在某个输入向量上，而变换的整体效果则是作用到空间中的每一个向量（前面说过，在考虑很多向量，可以把向量看成点，所以可以认为作用在每一个点上）。为了获得较好的变换前后对比的视觉效果，考虑的是对有间隔网格上的点作变换，同时保留前后对比。



Note：各式各样对空间的变换所产生的效果很美妙！但相应的变换也可能很复杂。

但幸运的是，线性代数限制在一类特殊类型的变换上：线性变换。

什么是线性变换？直观理解是什么？

直观地定义线性变换：如果一个变换满足①所有直线在变换后仍是直线，没有弯曲，②原点保持固定，则称这个变换是线性变换（也可以看成是保持网格线平行且等距分布的变换）

Note：后面将对线性变换作严格抽象定义。

如何用数值去描述线性变换？

应该如何用数值去描述这些线性变换呢？这个问题的答案有助于你明白应该给计算机什么计算公式，它可以根据你给的向量坐标，告诉你变换后的向量坐标 $\begin{bmatrix} x_{in} \\ y_{in} \end{bmatrix} \xrightarrow{?} \begin{bmatrix} x_{out} \\ y_{out} \end{bmatrix}$

这个问题的答案是，你只需记录两个基向量 $\hat{i}$ 和 $\hat{j}$ 变换后的位置（基向量变化可反映其张成span的变化）

具体地，对某个向量 $\vec{v}=-1\hat{i}+2\hat{j}$，（变换后的 $\vec{v}$）$=-1\times$（变换后的 $\hat{i}$）$+2\times$（变换后的 $\hat{j}$）

即线性变换后，线性关系仍然不变（这是网格线保持平行且等距分布的性质，即线性变换的重要推论）

这意味着，若知道 $\hat{i}$，$\hat{j}$ 在变换后的向量 $\hat{i}$，$\hat{j}$，以及原向量的坐标，就可以算出变换后的向量坐标。

即，若知道一个线性变换使得 $\hat{i}\to\begin{bmatrix}1\\-2\end{bmatrix}$，$\hat{j}\to\begin{bmatrix}3\\0\end{bmatrix}$，则 $\begin{bmatrix}x\\y\end{bmatrix}=x\hat{i}+y\hat{j}\to x\begin{bmatrix}1\\-2\end{bmatrix}+y\begin{bmatrix}3\\0\end{bmatrix}=\begin{bmatrix}x+3y\\-2x\end{bmatrix}$.

Note：这里考察向量坐标，其实是在同一个坐标系，同一组基下考察的，即前后向量坐标都是针对 $\hat{i}$，$\hat{j}$ 这组基而言的，所有的向量包括基向量经过变换确实变化了，但我们仍用之前的基作为基来考虑向量坐标。即 $\begin{bmatrix}x\\y\end{bmatrix}=x\hat{i}+y\hat{j}\Rightarrow x\begin{bmatrix}1\\-2\end{bmatrix}+y\begin{bmatrix}3\\0\end{bmatrix}=x(\hat{i}-2\hat{j})+y(3\hat{i})=(x+3y)\hat{i}-2x\hat{i}$

可见，一个二维线性变换仅由4个数字完全确定，这四个数字即是在变换后的向量坐标，将它们包装在2×2格子中，称为2×2矩阵。于是，一个矩阵 $\begin{bmatrix}a&b\\c&d\end{bmatrix}$ 在这里只是一个记号，它含有描述一个线性变换的全部信息，让这个变换作用于向量 $\begin{bmatrix}x\\y\end{bmatrix}$，结果是 $x\begin{bmatrix}a\\c\end{bmatrix}+y\begin{bmatrix}b\\d\end{bmatrix}=\begin{bmatrix}ax+by\\cx+dy\end{bmatrix}$

其结果也可被定义为 $\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}ax+by\\cx+dy\end{bmatrix}$（矩阵向量乘法）

上面的讨论是将线性变换看成矩阵。事实上，完全也可以把一个给定矩阵看成线性变换，即将它的两列看成是基向量变换后的向量（所以基向量，坐标系要先定好）

矩阵—线性变换转化的例子.

线性变换 → 矩阵：如逆时针转 90° 的线性变换 $\hat{i}=\begin{bmatrix}1\\0\end{bmatrix}\Rightarrow\begin{bmatrix}0\\1\end{bmatrix}$ $\hat{j}=\begin{bmatrix}0\\1\end{bmatrix}\Rightarrow\begin{bmatrix}-1\\0\end{bmatrix}$. 则矩阵为 $\begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$

若想求任何向量在逆时针转 90° 后的位置. 只需把它与矩阵相乘即可. 如 $\begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$

Note：一个特殊的变换叫剪切 (shear, 数学上亦称错切). 对应矩阵 $\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$

反过来. 矩阵 → 线性变换：如 $\begin{bmatrix}1 & 3\\2 & 1\end{bmatrix}$. 如何想象它代表的线性变换的样子：将 $\hat{i}$ 移到 (1,2). $\hat{j}$ 移到 (3,1). 空间其他部分随二者一起移动. 以保持网格线平行且等距分布

Note：若变换后的 $\hat{i}$ 和 $\hat{j}$ 是线性相关的. 线性变换将把整个二维空间压到它们所在的直线上. 即这两个线性相关向量所张成的一维空间上.

总结：线性变换是操纵空间的一种手段. 它保持网格线平行且等距分布, 并且原点保持不动. 而这种变换只需几个数字(或说一个矩阵) 就能描述清楚. 所以矩阵为我们提供了一种描述线性变换的语言. 而矩阵向量乘法就是计算线性变换作用于给定向量的一种途径.

※ 每当你看到一个矩阵时. 你都可以把它解读为对空间的一种特定变换 ※

上面给了线性变换的直观定义. 严格定义如下：

若一个变换 L 满足以下两条性质：① 可加性：$L(\vec{v}+\vec{w})=L(\vec{v})+L(\vec{w})$

② 成比例 (一阶齐次)：$L(c\vec{v})=cL(\vec{v})$. 则称 L 是线性变换.

Note：事实上. 上面的形象理解与这里的两条性质等价.

4. 矩阵乘法与线性变换复合

"It is my experience that proofs involving matrices can be shorted by 50% if one throws the matrices out"

— Emil Artin

据我的经验. 如果丢掉矩阵的话. 那些涉及矩阵的证明可以缩短一半.

— 埃米尔·阿廷

很多时候你想描述这样一个作用：一个变换之后再进行另一个变换. 如先逆转 90°, 再剪切. 从头到尾的总体作用是另一个线性变换. 称为两个独立变换的复合变换

和独立变换可以用矩阵描述一样. 复合变换作为变换. 也可以用单个矩阵描述. 但由于它又是复合的. 所以也可以用那两个独立变换对应的矩阵写出. 即 $\begin{bmatrix}a & b\\c & d\end{bmatrix}\begin{bmatrix}e & f\\g & h\end{bmatrix}=\begin{bmatrix}ae+bg, & \cdots\\\cdots, & cf+dh\end{bmatrix}$ 矩阵乘法

Note：可见. 矩阵乘法的几何意义就是两个线性变换相继作用. 右侧的先作用. 作用顺序从右到左

矩阵相乘时. 先后顺序影响结果吗？即 $M_1M_2 \overset{?}{=} M_2M_1$ 影响. 因为线性变换的复合与顺序有关

证明矩阵结合律：$(M_1M_2)M_3=M_1(M_2M_3)$ 两端都表示先后作用 M_3, M_2, M_1 对应的线性变换

Note：附注1—三维空间中的线性变换被跳过

5. 行列式

"The purpose of computation is insight, not numbers"

——Richard Hamming

—— 理查德·哈明

计算的目的不在于数字本身，而在于洞察其背后的意义。

我们已经知道线性变换可以用矩阵描述，而有的线性变换可以向外拉伸空间，有的则将空间向内挤压。而测量变换究竟对空间有多少拉伸或挤压，即测量一个给定区域面积增大或减小的比例，对理解线性变换而言很有意义

什么是线性变换的行列式？

事实上，对一个变换而言，由"网格线保持平行且等距分布"（线性变换）可以推出任何区域的面积变化比例是相同的。于是我们只需要考虑其中极特殊的，以基向量 $\hat{i}$、$\hat{j}$ 为边的正方形的面积的变化比例，它的缩放比例，即线性变换改变面积的比例被称为线性变换的行列式（determinant of a transformation）。如 $\mathrm{Det}\left(\begin{bmatrix} 0 & 2 \\ -1.5 & 1 \end{bmatrix}\right) = 3$

行列式的绝对值大小有什么意义？

Note：$\mathrm{Det}(A)=3$，表示将一个区域的面积增大为原来的3倍。

$\mathrm{Det}(A)=\frac{1}{2}$，表示将一个区域的面积减小为原来的 $\frac{1}{2}$ 倍。

$\mathrm{Det}(A)=0$，表示将整个空间（平面）压缩到一条线或一个点上（此时列向量线性相关）

从这个式子可以看出，只需要检验一个矩阵的行列式是否为0，即可了解矩阵所代表的变换是否将空间压缩到更小的维度上。

行列式的符号有什么意义？

那 $\mathrm{Det}(A)<0$ 是什么意义？这与定向面积有关，即我们规定从 $\hat{i}$ 到 $\hat{j}$ 为顺时针时，形成的面积是正的。若变换后的 $\hat{i}$ 与变换后的 $\hat{j}$ 为顺时针（以小于180°的那个为准），则 $\mathrm{Det}(A)>0$，否则，逆时针，$\mathrm{Det}(A)<0$。可见，$\mathrm{Det}(A)$ 的正负（符号），可以表示面积的方向，但它的绝对值依然表示区域面的缩放比例。

行列式在三维空间中是什么意义？

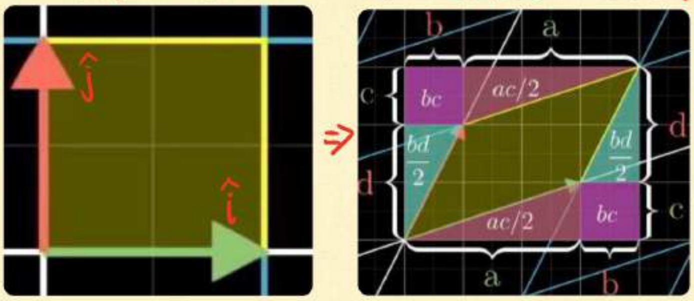
那行列式在三维空间中是什么意义？三维空间中的行列式与定向体积有关，它的大小表示基 $\hat{i}$、$\hat{j}$、$\hat{k}$ 形成的正方体经变换后的缩放比例（这等同于缩放后得到的平行六面体的体积）。行列式的符号仍与定向有关，一般规定符合 $\hat{i}$、$\hat{j}$、$\hat{k}$ 成右手定则的体积为正体积。

怎么计算行列式？

下左图单位正方形经线性变换后成右图，定向面积为 $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)=(a+b)(c+d)-ac-bd-2bc$

$=ad-bc$

Note：徒手计算行列式是必须掌握的，而这只能依赖不断地练习。



证明：$\det(M_1M_2)=\det(M_1)\det(M_2)$ 复合变换的缩放效果等同于独立变换缩放效果的叠加。

6.逆矩阵.列空间与零空间.

"To ask the right question is harder than to answer it."

——Georg Cantor

提出正确的问题比回答它更困难.

——格奥尔格·康托尔

本系列的大部分旨在通过直观的线性变换来理解矩阵与向量运算.
本节旨在通过线性变换来了解逆矩阵.列空间.秩和零空间的概念.
Note:本节不讨论计算部分.计算部分很重要.但有其它优质资源帮助你学习.而且,实践中有软件辅助我们进行计算.

线性代数的有用之处: 它能用来描述对空间的操纵.这对计算机图形学和机器人学有用
它能帮助求解特定方程组.这对几乎所有技术领域有帮助

这里考虑线性方程组(未知数个数与方程个数一致).如:
$$\begin{cases} 2x+5y+3z=-3 \\ 4x+0y+8z=0 \\ 1x+3y+0z=2 \end{cases}$$

结合矩阵向量乘法.线性方程组可以被写成:
$$\underbrace{\begin{bmatrix} 2 & 5 & 3 \\ 4 & 0 & 8 \\ 1 & 3 & 0 \end{bmatrix}}_{系数矩阵A} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}}_{\vec{v}}$$
$\Leftrightarrow A\vec{x}=\vec{v}$.将线性方程组写成矩阵向量乘法的形式.赋予了线性方程组几何直观.A表示一种线性变换.故求$A\vec{x}=\vec{v}$意味着去找一向量$\vec{x}$.使它在变换后与$\vec{v}$重合.

现在,这个方程的解依赖于矩阵A所代表的变换
①是将空间挤压到低维空间.即$det(A)=0$
②还是保持原状态一样的完整空间.即$det(A)\neq0$

①先看$det(A)\neq0$的情形(空间未被压缩为零面积区域).此时有且仅有一个向量在变换后与$\vec{v}$重合.而我们可以通过逆向变换来找到这个向量.逆向变换实际上对应另一个矩阵.称为A的逆.记为A^{-1}.如A是逆时针转90°变换.A^{-1}就是顺时针转90°变换
先作A变换.再作A^{-1}变换.则等于什么也没做.即$A^{-1}A=I$(恒等变换)$=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
一旦有了A^{-1}(可用计算机算).则可由$A^{-1}A\vec{x}=A^{-1}\vec{v}\Leftrightarrow\vec{x}=A^{-1}\vec{v}$来求解方程组
②当$det(A)=0$的情形(变换将空间压缩到低维空间).此时.没有逆变换.但解仍可能存在.例如,变换A将空间压缩成一条直线,而向量$\vec{v}$恰好在那条线上,则解存在.
Note:即使同样使行列式为0.不同的变换下.解存在的可能性也不同.如对三维空间而言.变换A将它压缩成一个二维面.变换B将它压缩成一个一维线.对于各种可能的v.前一种情况解更有可能存在.后一种情况解更不可能存在.为此引入新术语.秩来区分这两种情况.

秩是什么意义.为什么需引入秩

变换后空间的维度称为秩(rank).它是变换(或变换对应的矩阵)的一种属性.例如.对2×2矩阵,它的秩最大为2,秩为2时意味着基向量经变换后仍能张成整个二维空间.且行列式不为0.但对3×3矩阵.秩为2.表示空间被压缩到低维空间,但相比秩为1的变换,压缩没有那么严重.
所有可能的变换的结果的集合.被称为矩阵的"列空间"(column space).
Note:矩阵的列告诉我们基向量变换后的向量.由这些向量张成的空间就是所有可能的变换结果的集合,即矩阵的列空间是矩阵的列向量张成的空间.所以秩更准确的定义是矩阵对应的列空间的维度.当秩达到最大值时,即秩和列数相等.称为满秩(full rank).注意零

4个基本子空间.

向量一定在列空间中(因为其向量坐标仍全为0)

总结：解的结构.

考虑 $A\vec{x}=\vec{0}$ (称为齐次方程)

①若A是满秩的，则唯一能在变换后落在原点的就是零向量，于是 $\vec{x}$ 存在且唯一，是 $\vec{0}$.

②若A不是满秩的，则有一系列向量在变换后成为零向量。如：一个线性变换将二维空间压缩到一条直线(列空间)上，那么沿另一个方向(非列空间方向)上的所有向量都被压缩到了原点.

变换后落在原点的向量的集合被称为矩阵的"零空间"(null space)"或"核"(kernel).

于是，零空间给出的，即是上面的向量方程 $A\vec{x}=\vec{0}$ 所有可能的解.

Note：零空间的一组基，被称为A矩阵对应的齐次方程的一个基础解系.

零向量的维度(即基础解系中解的个数)+列空间维度=原空间维度.

考虑 $A\vec{x}=\vec{v}$ ($\vec{v}\neq\vec{0}$，称为非齐次方程)

①若A是满秩的，前面已经分析过，求逆变换可得唯一解

②若A不是满秩的，则 $\vec{v}$ 在列空间中，有解，$\vec{v}$ 不在列空间中，无解，换句话说，有解 $\Leftrightarrow$ $\vec{v}$ 在列空间中 $\Leftrightarrow$ $\vec{v}$ 的加入，无法使列空间维度提高 $\Leftrightarrow$ $[A;\vec{v}]$ 和A的秩相同(解的存在性定理)

若有解 $\vec{x_0}$，则根据 $A(\vec{x}'+\vec{x_0})=A\vec{x}'+A\vec{x_0}=A\vec{x}'+\vec{v}$，只需令 $A\vec{x}'=\vec{0}$，求 $\vec{x}'$，则 $\vec{x}'+x_0$ 也是 $A\vec{x}=\vec{v}$ 的解，而 $A\vec{x}'=\vec{0}$ 的解即是A的零空间，可见 $A\vec{x}=\vec{v}$ 的解是一个特解 x_0 + 零空间的任一向量

上面对 $A\vec{x}=\vec{0}$，$A\vec{x}=\vec{v}$ 的讨论可以合起来.

考虑 $A\vec{x}=\vec{v}$ → (A与[A;v]同秩?) → 同：有解 → ($\vec{v}=\vec{0}$?) → 等于：求A的零空间

↳ 不同：无解　　　↳ 不等于：求特解+零空间.

Note：在 $A\vec{x}=\vec{v}$ 有解的情况下，解的个数取决于零空间的维度，即原空间维度−列空间维度.

附注2：非方阵.

我们知道，在指定一组基的情况下，线性变换可以用一个方阵表示、这也是我们上面讨论的基础，即矩阵的几何意义。那非方阵的几何意义是什么？答案是：两组基下的线性映射.

Note：线性变换是向量空间V到V的，线性映射是从向量空间V到W的.

例如 $\begin{bmatrix}2&0\\-1&1\\-2&1\end{bmatrix}$ 表示的是分别给定了基的二维空间V到三维空间W的线性映射，将V的基向量2个分别变成了W中坐标为(2,-1,-2)，(0,1,1)的向量，即这个矩阵表示把二维空间整个映射到三维空间的部分(可以是一个面，一条线或原点)

Note：线性变换仍可认为是V中的向量在移动，而线性映射已经不能这么认为了.

对 $\begin{bmatrix}2&0\\-1&1\\-2&1\end{bmatrix}$ 的讨论：它是一个3×2矩阵，它的列空间是三维空间W的一个过原点的二维平面，但它仍是满秩的(更准确的说，是列满秩的)，因为列空间维度与输入空间V维度一样，但不是行满秩的，因为列空间维度与输出空间W的维度不一致.

考虑 $A\vec{x}=\vec{v}$：若A是3×2的，表示是三个方程，二个未知数，A表示将二维空间映射到三维空间，结果可能无解($\vec{v}$ 在列空间外)，可能唯一解(零空间维度为0)，可能无穷解(零空间维度为1或2)。若A是2×3的，二个方程，三个未知数，三维空间映射到二维，可能无解，可能无穷解(零空间维度为1,2,3)，不可能唯一解.

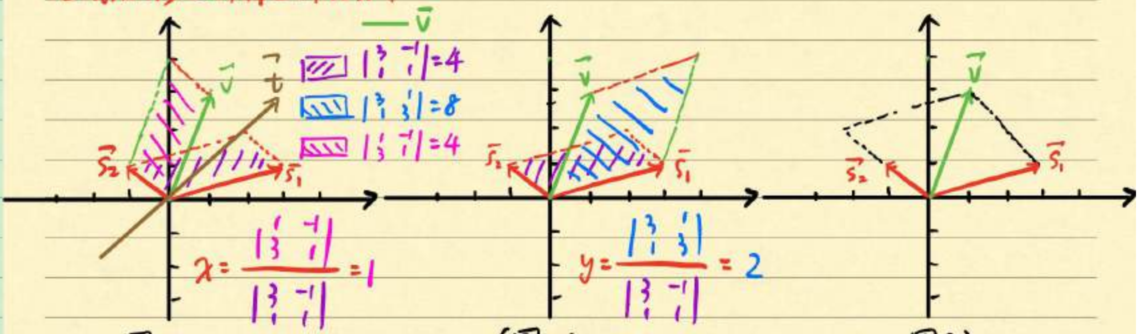
Note：这里零空间维度 = 输入空间V的维度 − 列空间的维度

解线性方程组的克拉姆法则(Cramer's Rule)

例：解线性方程组(矩阵向量形式)：$\begin{bmatrix}3 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}1\\ 3\end{bmatrix}$

解：利用 Cramer's Rule：由 $\begin{vmatrix}3 & -1\\ 1 & 1\end{vmatrix}=4$，则 $x=\begin{vmatrix}1 & -1\\ 3 & 1\end{vmatrix}/4=1$ $y=\begin{vmatrix}3 & 1\\ 1 & 3\end{vmatrix}/4=2$

尝试从几何上解释这种结果：



(图1) (图2) (图3)

这里，姑且称 $\begin{bmatrix}3\\ 1\end{bmatrix}$ 为 $\vec{s_1}$，$\begin{bmatrix}-1\\ 1\end{bmatrix}$ 为 $\vec{s_2}$。

则图1告诉我们 x 应等于($\vec{s_2}$ 与 $\vec{v}$ 组成的面积)比上($\vec{s_1}$ 与 $\vec{s_2}$ 组成的面积)。事实上，这两块面积有相同的底 $\vec{s_2}$，于是面积比即为(沿 $\vec{s_2}$ 垂直方向上)，$\vec{v}$ 与 $\vec{s_1}$ 的投影长度之比，而这个比值恰好说明了 $\vec{s_1}$ 对应的缩放标量应该为1，即不缩放。

图2同理有几何解释。

图3是最终结果。

7. 点积与对偶性：

"Calvin: You know, I don't think math is a science, I think it's a religion.

Hobbes: A religion?

Calvin: Yeah. All these questions are like miracles. You take two numbers and when you add them, they magically become one new number! No one can say how it happens. You either believe it or you don't."

卡尔文：你知道吗，我觉得数学不是一门科学，而是一种宗教。

霍布斯：一种宗教？

卡尔文：是啊，这些公式就像奇迹一样，你取出两个数，把它们相加时，它们神奇地成为了一个全新的数！没人能说清这到底是怎样发生的，你要么完全相信，要么完全不信。

——《卡尔文与霍布斯》连载四格漫画

先从点积的标准定义讲起，再介绍其对应的几何直观

对两个维数相同的向量，如 $\begin{bmatrix}2\\ 7\\ 1\end{bmatrix}$，$\begin{bmatrix}8\\ 2\\ 8\end{bmatrix}$，它们的点积，就是将对应坐标配对，求出每一对坐标的乘积，然后将结果相加，这里即 $2\times8+7\times2+1\times8$

令 $\vec{v}=\begin{bmatrix}4\\ 1\end{bmatrix}$，$\vec{w}=\begin{bmatrix}2\\ -1\end{bmatrix}$，内积 $\vec{v}\cdot\vec{w}=\begin{bmatrix}4\\ 1\end{bmatrix}\cdot\begin{bmatrix}2\\ -1\end{bmatrix}$ 的几何解释是 $\vec{w}$ 朝过原点和 $\vec{v}$ 终点的直线上的正交投影的长度与向量 $\vec{v}$ 的长度的乘积，并带上符号(若 $\vec{w}$ 的投影与 $\vec{v}$ 同向，则符号为正，否则为负。(几何解释也可以从 $\vec{v}$ 向 $\vec{w}$ 投影的角度考虑)

为什么点积，即对应坐标相乘并将结合相加，和投影有所联系？

要一个满意的答案，需引入“对偶性”。

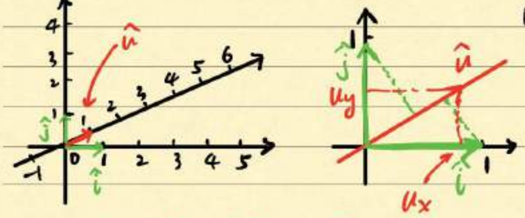
考虑一种特殊的线性映射，它从二维空间映射到一维空间，在这两个空间引入坐标系(基)之后，线性映射可以由 1×2 的矩阵来表示，如 $[1, -2]$，它表示将二维空间中的 $\hat{i}$，$\hat{j}$ 映射到一维空间(数轴)上的 1 和 -2。若这种映射作用在二维向量 $\begin{bmatrix}4\\3\end{bmatrix}$ 上，则由矩阵向量乘法知，对应 $[1, -2]\begin{bmatrix}4\\3\end{bmatrix} = 4\times1+3\times(-2)$，形式上似乎和向量点积很像（$1\times2$ 的矩阵不像一个倾倒的 $\begin{bmatrix}4\\3\end{bmatrix}$ 向量吗？）

事实上，1×2 矩阵与二维向量有着微妙的联系（将向量放倒，则得到与之相关的矩阵，如向量 $\begin{bmatrix}2\\7\end{bmatrix}$，放倒后是 1×2 矩阵 $[2, 7]$，反过来将矩阵立直，则得到与之相关的向量）

举个例子来说明这种联系的重要性（这将回答了第一行提到的点积与投影的问题）

将一个数轴，保持 0 在原点，斜放在空间中，考虑终点落在数轴的 1 上的单位向量 $\hat{u}=\begin{bmatrix}u_x\\u_y\end{bmatrix}$

若将空间中的二维向量投影到这条数轴上，就会得到一个数，于是，这个投影过程对应了一个从二维空间到一维空间（从二维向量到数）的映射(函数)，而这种映射，实际上是一种线性映射，于是可以写成 1×2 矩阵的形式，两列分别表示 $\hat{i}$ 映射后的值和 $\hat{j}$ 映射后的值。利用对偶性可知，这两个值恰为 u_x 和 u_y，于是，将向量 $\begin{bmatrix}x\\y\end{bmatrix}$ 朝着沿 $\hat{u}=\begin{bmatrix}u_x\\u_y\end{bmatrix}$ 方向上的直线投影，就相当于线性映射作用在 $\begin{bmatrix}x\\y\end{bmatrix}$ 上，相当于 1×2 矩阵 $[u_x, u_y]$（可以看成 $\hat{u}$ 的转置）左乘 $\begin{bmatrix}x\\y\end{bmatrix} = xu_x+yu_y$



即 $\vec{w}$ 向单位向量 $\hat{u}$ 投影的结果 $\Leftrightarrow$ $\hat{u}$ 对应的线性映射作用 $\vec{w}$ $\Leftrightarrow$ ($\hat{u}$ 的转置)矩阵左乘 $\vec{w}$

↳ 称投影矩阵

这样，投影和矩阵向量乘法有了联系。

而根据向量点积的定义，xu_x+yu_y 恰可看成 $\begin{bmatrix}u_x\\u_y\end{bmatrix}$ 与 $\begin{bmatrix}x\\y\end{bmatrix}$ 的点积 $\begin{bmatrix}u_x\\u_y\end{bmatrix}\cdot\begin{bmatrix}x\\y\end{bmatrix}$

于是，$\begin{bmatrix}u_x\\u_y\end{bmatrix}\cdot\begin{bmatrix}x\\y\end{bmatrix}$ 点积 $\Leftrightarrow$ $[u_x, u_y]\begin{bmatrix}x\\y\end{bmatrix}$（矩阵向量乘法）$\Leftrightarrow$ 向量 $\begin{bmatrix}x\\y\end{bmatrix}$ 投影到 $\hat{u}$ 方向上

这就是为什么与单位向量 $\hat{u}$ 的点积可解读为将向量投影到单位向量所在直线上所得的投影长度(带符号)。对非单位向量的点积，可以提出其向量长度和单位向量，所以任意两向量 $\vec{v}$ 和 $\vec{w}$ 的内积等于 $\vec{w}$ 在 $\vec{v}$ 上的投影长度(带符号)乘以 $\vec{v}$ 的向量长度

Note：上述过程给我们一个启发：任何一个输出空间为一维空间的线性映射，对应一个 $1\times n$ 矩阵，也对应 n 维输入空间的唯一一个向量 $\vec{v}$（那个矩阵的转置），称为对偶向量。

上述过程也是数学中对偶性（二象性 “duality”）的一个实例。粗略地说，

对偶性 $\Leftrightarrow$ 两种数学事物之间自然而又出乎意料的对应关系

可以说，一个向量的对偶是由它定义的线性映射，是那个 $1\times n$ 矩阵。

而一个 n 维空间到一维空间的线性映射的对偶则是多维空间中的某个特定向量。

8.1 叉积的标准介绍

"Every dimension is special."

——Jeff · Lagarias.

"每一个维度都很特别。"

——杰弗里·拉加里亚斯

向量的叉积是通过两个三维向量 $\vec{v}$ 和 $\vec{w}$ 生成一个新的三维向量 $\vec{p}$，其中 $\vec{p}$ 的长度是 $\vec{v}$ 和 $\vec{w}$ 围成的面积，方向垂直于 $\vec{v}$ 和 $\vec{w}$ 组成的平行四边形（服从 $\vec{v}$，$\vec{w}$，$\vec{p}$ 成右手螺旋）

Note：上面是叉积的几何描述，对叉积的代数描述，可用公式来表示 $\vec{v}$ 和 $\vec{w}$ 的叉积，即

$\vec{v}=\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$，$\vec{w}=\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$，则 $\vec{v}\times\vec{w}=\det\left(\begin{bmatrix} \hat{i} & v_1 & w_1 \\ \hat{j} & v_2 & w_2 \\ \hat{k} & v_3 & w_3 \end{bmatrix}\right)$

8.2 以线性变换的眼光看叉积

"From [Grothendieck], I have also learned not to take glory in the difficulty of a proof: difficulty means we have not understood. The ideal is to be able to paint a landscape in which the proof is obvious."

——Pierre Deligne

"从他（格罗滕迪克）和他的作为中，我还学到了一点：不以证明困难为豪，因为难度高意味着我们还不理解。理想的情况是能够绘出一幅美景，而其中的证明显而易见"

——皮埃尔·德利涅

本节的目的是说明上一节的代数描述和几何描述确实是对应的，即由 $\det\left(\begin{bmatrix} \hat{i} & v_1 & w_1 \\ \hat{j} & v_2 & w_2 \\ \hat{k} & v_3 & w_3 \end{bmatrix}\right)$ 得到的三维向量确实是垂直于 $\vec{v}$ 和 $\vec{w}$ 的，方向符合右手系的长度为 $\vec{v}$ 和 $\vec{w}$ 组成的平行四边形面积的向量。

总体计划：① 根据 $\vec{v}$ 和 $\vec{w}$ 定义一个三维到一维的线性映射 ② 说明它的对偶向量 $=\vec{v}\times\vec{w}$

定义一个三维空间到数轴的函数 $f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)=\det\left(\begin{bmatrix} x & v_1 & w_1 \\ y & v_2 & w_2 \\ z & v_3 & w_3 \end{bmatrix}\right)$，它的几何意义是，对任一输入的三维向量，输出它与 $\vec{v}$ 和 $\vec{w}$ 确定的平行六面体的体积（带符号的体积）

容易验证，上面定义的映射是线性映射（由行列式性质可知）。于是由对偶性可知，这个线性映射对应一个对偶向量 $\vec{p}=\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$ 满足，$\underbrace{\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}\cdot\begin{bmatrix} x \\ y \\ z \end{bmatrix}}=[p_1 \ p_2 \ p_3]\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\underbrace{\det\left(\begin{bmatrix} x & v_1 & w_1 \\ y & v_2 & w_2 \\ z & v_3 & w_3 \end{bmatrix}\right)}$

分别展开"～"处可得 $p_1x+p_2y+p_3z=(v_2w_3-v_3w_2)x+(v_3w_1-v_1w_3)y+(v_1w_2-v_2w_1)z$

于是对应相等可得 $p_1=v_2w_3-v_3w_2$，$p_2=v_3w_1-v_1w_3$，$p_3=v_1w_2-v_2w_1$

即 $\begin{bmatrix} v_2w_3-v_3w_2 \\ v_3w_1-v_1w_3 \\ v_1w_2-v_2w_1 \end{bmatrix}$，而这个结果，恰与由 $\det\left(\begin{bmatrix} \hat{i} & v_1 & w_1 \\ \hat{j} & v_2 & w_2 \\ \hat{k} & v_3 & w_3 \end{bmatrix}\right)$ 导出的 $\vec{v}\times\vec{w}$ 的结果一致。

Note：这样，就将"求由 $\vec{v}$ 和 $\vec{w}$ 定义的线性映射的对偶向量"，与叉积的代数描述联系起来了。

下面将两者与叉积的几何描述联系起来。

利用点积的几何意义和上面的线性映射的定义，我们希望从几何角度认识 $\vec{p}$。

$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}\cdot\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 表示向量 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 在 $\vec{p}$ 上的投影（带符号）乘以 $\vec{p}$ 的长度，而由映射定义，其值为由 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$，$\vec{v}$，$\vec{w}$ 组成的平行六面体的体积 $=$ $\vec{v}$，$\vec{w}$ 组成的平行四边形面积乘以 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 在垂直平行四边形方向上的分量（带符号）。将"～"看作 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 在 $\vec{p}$ 上的投影（带符号），则 $\vec{p}$ 的长度对应 $\vec{v}$ 和 $\vec{w}$ 组成的平行四边形，方向符合右手系，与 $\vec{v}$，$\vec{w}$ 和 $\vec{w}$ 垂直，这恰为 $\vec{v}\times\vec{w}$ 的几何描述（定义）

由定义的映射的对偶向量是唯一的，所以叉积的几何描述和代数描述联系在了一起。

向量 $\vec{v}$ 与 $\vec{w}$ 的叉积有两个定义，一个从几何角度，说明叉积是垂直 $\vec{v}$ 与 $\vec{w}$ 且成右手系，长为 $\vec{v}$ 与 $\vec{w}$ 组成平行四边形面积的向量。另一个从代数角度，给出叉积计算公式 $\det\left(\begin{bmatrix} i & v_1 & w_1 \\ j & v_2 & w_2 \\ k & v_3 & w_3 \end{bmatrix}\right)$

本节通过构造一个从三维空间到数轴的线性映射，得到一个对偶向量 $\vec{p}$，分别从几何和代数角度考察 $\vec{p}$，发现 $\vec{p}$ 与 $\vec{v}\times\vec{w}$ 的几何和代数结果一致，于是说明了叉积的几何、代数定义确实有联系，它们是观察 $\vec{v}\times\vec{w}$ 的两个角度。

即

$$\begin{tikzcd} & \text{几何角度} \arrow[dr] & \\ \text{叉积 } \vec{v}\times\vec{w} \arrow[ur] \arrow[dr] & & \text{构造的线性映射的对偶向量} \\ & \text{代数角度} \arrow[ur] & \end{tikzcd}$$

9.基变换

"Mathematics is the art of giving the same name to different things."

——Henri Poincaré.

数学是一门赋予不同事物相同名称的艺术.

——昂利·庞加莱

一般地，这是一组基，如二维空间中的 $\hat{i}$, $\hat{j}$，才可能给出任意向量的向量坐标，构建出完整的坐标系。所以坐标系是与基上的基密切相关的，没有定义基，那坐标系也就不存在。基定义的不同，那么构建出来的坐标系也不同。在第三节，我们讨论了线性变换，经过线性变换后，确实所有的向量都可能发生了移动，但基向量没有变（坐标系没有变）。

Note：移动的所有向量里确实有形如基向量的，但是它的移动并不代表基向量的改变，基向量仍是移动前的那一组。

发生在向量与一组数之间的任意一种双射，都被称为一个坐标系。如 $\begin{bmatrix}3\\2\end{bmatrix}$ <=> 向量

一个坐标系可以看成一种语言，向量坐标可以看成文字。(coordinate system)

如何在不同坐标系间转化？

使用不同的基来构建不同的坐标系，这些坐标系会有什么联系。例如，你使用 $\hat{i}$, $\hat{j}$ 作为基，而 Jennifer 使用 b_1, b_2 作为基。那么，同一个向量在你俩坐标系下的向量坐标分别是什么，有什么联系。你用向量坐标 $\begin{bmatrix}-1\\2\end{bmatrix}$ 表示的向量，在 Jennifer 的坐标系下的向量坐标是什么？

Note：假如，在你的坐标系下，b_1 是 $\begin{bmatrix}2\\1\end{bmatrix}$，$b_2$ 是 $\begin{bmatrix}-1\\1\end{bmatrix}$（当然，在 Jennifer 那，$b_1$ 是 $\begin{bmatrix}1\\0\end{bmatrix}$，$b_2$ 是 $\begin{bmatrix}0\\1\end{bmatrix}$）。虽然基的选择不同，但你和 Jennifer 的坐标系 (0,0) 点是重合的，即原点重合。

①首先，看 Jennifer 中的 $\begin{bmatrix}-1\\2\end{bmatrix}$ 向量，在你的坐标系中的向量坐标是什么？

由 $\begin{bmatrix}-1\\2\end{bmatrix} \to (-1)\times b_1+2b_2=(-1)\times\begin{bmatrix}2\\1\end{bmatrix}+2\times\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}2&-1\\1&1\end{bmatrix}\begin{bmatrix}-1\\2\end{bmatrix}$

得结果是 Jennifer 中的基在你的坐标系的向量坐标组成的矩阵和 Jennifer 中向量的坐标相乘。

Note：由于 $\begin{bmatrix}2&-1\\1&1\end{bmatrix}$ 可以看成线性变换对应的矩阵，所以这个结果直观上可解释为将基变换为 $\begin{bmatrix}2\\1\end{bmatrix}$，$\begin{bmatrix}-1\\1\end{bmatrix}$ 时，$\begin{bmatrix}-1\\2\end{bmatrix}$ 变换得到的结果（注意，用于定义坐标系的基向量没变，所以还是在你的坐标系下）。这里的 $\begin{bmatrix}2&-1\\1&1\end{bmatrix}$ 被称为基变换矩阵，表示旧基(Jennifer的)在新基(你的)下的向量坐标。

②相反情形会怎样，即你的 $\begin{bmatrix}3\\2\end{bmatrix}$ 向量，在 Jennifer 中的向量坐标是什么？将上面的基变换矩阵求逆即可。$\begin{bmatrix}2&-1\\1&1\end{bmatrix}^{-1}=\begin{bmatrix}1/3&1/3\\-1/3&2/3\end{bmatrix}$（你的基在 Jennifer 的基下的表示）

则 $\begin{bmatrix}1/3&1/3\\-1/3&2/3\end{bmatrix}\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}5/3\\1/3\end{bmatrix}$，表示在 Jennifer 的坐标系下，相同向量的坐标是 $\begin{bmatrix}5/3\\1/3\end{bmatrix}$

以上就是如何在坐标系之间对单个向量的描述进行互相转化（通过基变换矩阵）

③事实上，同一个线性变换，作用在空间上的效果是唯一的，但考虑到将线性变换对应到矩阵的话，其结果就与坐标系、与基有关，而变得不唯一了，那不同坐标系(基)下，同一线性变换对应的矩阵有什么联系呢？例如，你用 $\begin{bmatrix}0&-1\\1&0\end{bmatrix}$ 描述逆时针转90°，那么 Jennifer 如何描述同样空间的逆90°旋转？

正常的过程是，从 Jennifer 中的任一向量 $\vec{v}$ 出发，用基变换矩阵转化为你的语言（坐标系）：$\begin{bmatrix}2&-1\\1&1\end{bmatrix}\vec{v}$，再左乘变换矩阵 $\begin{bmatrix}0&-1\\1&0\end{bmatrix}$（表示逆时针转90°），再用基变换矩阵的逆转化为 Jennifer 的语言：

$\begin{bmatrix}2&-1\\1&1\end{bmatrix}^{-1}\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}2&-1\\1&1\end{bmatrix}\vec{v}$，于是这就是 Jennifer 的视角下 $\vec{v}$ 作逆时针转90°变换后的结果

$\begin{bmatrix}2&-1\\1&1\end{bmatrix}^{-1}\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}2&-1\\1&1\end{bmatrix}$ 就是变换矩阵。

Note：$A^{-1}MA$ 暗示了数学上的转移作用，中间代表你所见的变换，而外侧两个矩阵则表示视角上的转移，矩阵乘积仍代表同一变换，只是从其他人的角度来看。

10. 特征向量与特征值

"Last time, I asked: 'What does mathematics mean to you?', and some people answered: 'The manipulation of numbers, the manipulation of structures.' And if I had asked what music means to you, would you have answered: 'The manipulation of notes?'"

—— Serge Lang

"上一次演讲中我问道：'数学对你来说意味着什么？'有些人回答：'处理数字，处理结构。'那么如果我问音乐对你来说意味着什么，你会回答'处理音符'吗？"

—— 塞尔日·兰

首先考虑二维空间中的某个线性变换，其对应矩阵 $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$，考虑它对特定向量的作用，并考虑这个向量张成的空间（即通过原点和向量终点的直线）。大部分向量在变换中都离开了其张成的空间，但仍有些特殊向量留在所张成的空间里（即矩阵对它的作用仅仅是拉伸或压缩而已，如同一个标量）。对 $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$，可以找到两个这样的特殊向量 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$，对 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$，变换只对它拉伸了3倍，对 $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$，变换对它拉伸了2倍。另外事实上，$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 分别张成的两个空间上的每个向量，都是满足上面性质的特殊向量。

Note：事实上，找到的特殊向量就被称为特征向量，对应的伸缩比（可带符号表方向）被称为特征值。而每找到一个特征向量都有一个对应的特征子空间（那些特征向量张成的空间）。这里提到的特征向量都是非零特征向量

特征值和特征值有用且和值得细究的一个例证是：考虑一个三维空间中的旋转，如果你能找到这个变换的特征向量（它的特征值必为1），那么你找到的就是旋转轴。而把一个三维旋转看成绕某个轴转一定角度，要比考虑 3×3 矩阵直观地多。

本节不涉及特征向量和特征值的具体细节，但会概述一下计算思想。

用符号表述特征向量，特征值的概念：$A\vec{v} = \lambda\vec{v} = \lambda I\vec{v}$（$I$ 为单位阵），A 为具体变换对应矩阵（方阵），$\vec{v}$ 为特征向量，λ 为特征值。即变换作用在特征向量上，只相当于标量 λ 作用在上面。$A\vec{v} = \lambda I\vec{v} \Leftrightarrow (A-\lambda I)\vec{v} = \vec{0}$

Note：这个方程必有零向量解，但我们要的是非零解 $\vec{v}$，所以要求 $\det(A-\lambda I)=0$，解出某个 λ_0，则代入 $(A-\lambda I)\vec{v}=\vec{0}$ 中得 $(A-\lambda_0 I)\vec{v}=\vec{0}$，则 $A-\lambda_0 I$ 的零空间就是 λ_0 对应的特征子空间。另外，按特征向量定义，其依赖于变换，不依赖于矩阵，故与基的选择无关。

有几种特殊情况值得说明：① 变换不一定有特征值，如二维空间逆时针转90°，就没有特征向量，虽然其有特征值 $\pm i$，但没有实特征向量。② 属于单个特征值的特征向量可能不止在一条直线上，如将所有向量变为2倍的变换，则全空间的向量都是特征向量且特征值为2。

如果基向量恰为特征向量，则变换矩阵是对角矩阵（除对角线外其它元素均为0的矩阵被称为对角矩阵），即对对角矩阵，空间的所有基向量都是特征向量，矩阵的对角元是它们所属特征值（对角矩阵在作幂运算时很方便）。但若基向量不全是特征向量，但变换有很多特征向量，多到可以张成全空间，则可变换坐标系，使得这些特征向量就是基向量。例如，原变换 $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$（旧基），(新基—特征基的基变换阵为) $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$，则 $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$，这个特征基视角下的变换的矩阵形式必为对角阵。

一组基向量(同时也是特征向量)构成的集合被称为一组"特征基".

要计算某矩阵的100次幂，一种容易的做法是先变换到特征基，则变换矩阵变为对角阵，在那里计算变换的100次幂（即矩阵的100次幂），然后再转换回标准坐标系。

Note：不是所有变换都能进行这一过程（这个过程实际上是相似对角化）

一道练习：取 $A=\begin{bmatrix}0&1\\1&1\end{bmatrix}$，它的两个特征向量为 $\vec{v_1}=\begin{bmatrix}2\\1+\sqrt5\end{bmatrix}$、$\vec{v_2}=\begin{bmatrix}2\\1-\sqrt5\end{bmatrix}$，试求 A^n，并用 A^2、A^3（手算）检验。

解：$A^2=\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}0&1\\1&1\end{bmatrix}=\begin{bmatrix}1&1\\1&2\end{bmatrix}$ $\quad A^3=\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}1&1\\1&2\end{bmatrix}=\begin{bmatrix}1&2\\2&3\end{bmatrix}$

A对应变换在新基下的矩阵形式 $=\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}^{-1}\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}=\begin{bmatrix}\lambda_1&0\\0&\lambda_2\end{bmatrix}\triangleq B$

其中 λ_1 是 $\begin{bmatrix}2\\1+\sqrt5\end{bmatrix}$ 对应的特征值，λ_2 是 $\begin{bmatrix}2\\1-\sqrt5\end{bmatrix}$ 对应的特征值

即 $\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}2\\1+\sqrt5\end{bmatrix}=\lambda_1\begin{bmatrix}2\\1+\sqrt5\end{bmatrix}$，$\begin{bmatrix}0&1\\1&1\end{bmatrix}\begin{bmatrix}2\\1-\sqrt5\end{bmatrix}=\lambda_2\begin{bmatrix}2\\1-\sqrt5\end{bmatrix}$ $\Rightarrow \begin{cases}\lambda_1=\frac{1+\sqrt5}{2}\\ \lambda_2=\frac{1-\sqrt5}{2}\end{cases}$

于是 $A^n=\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}B^n\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}^{-1}=\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}\begin{bmatrix}\left(\frac{1+\sqrt5}{2}\right)^n&0\\0&\left(\frac{1-\sqrt5}{2}\right)^n\end{bmatrix}\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}^{-1}$

其中由 $\left[\begin{array}{cc|cc}2&2&1&0\\1+\sqrt5&1-\sqrt5&0&1\end{array}\right]\xrightarrow{①\times\frac12}\left[\begin{array}{cc|cc}1&1&\frac12&0\\1+\sqrt5&1-\sqrt5&0&1\end{array}\right]\xrightarrow{②-①\times(1+\sqrt5)}\left[\begin{array}{cc|cc}1&1&\frac12&0\\0&-2\sqrt5&-\frac{1+\sqrt5}{2}&1\end{array}\right]$

$\xrightarrow{②\times\frac{-1}{2\sqrt5}}\left[\begin{array}{cc|cc}1&1&\frac12&0\\0&1&\frac{5+\sqrt5}{20}&-\frac{\sqrt5}{10}\end{array}\right]\xrightarrow{①-②}\left[\begin{array}{cc|cc}1&0&\frac{5-\sqrt5}{20}&\frac{\sqrt5}{10}\\0&1&\frac{5+\sqrt5}{20}&-\frac{\sqrt5}{10}\end{array}\right]$

于是 $\begin{bmatrix}2&2\\1+\sqrt5&1-\sqrt5\end{bmatrix}^{-1}=\begin{bmatrix}\frac{5-\sqrt5}{20}&\frac{\sqrt5}{10}\\\frac{5+\sqrt5}{20}&-\frac{\sqrt5}{10}\end{bmatrix}$

于是 $A^n=\begin{bmatrix}\frac{5-\sqrt5}{10}\left(\frac{1+\sqrt5}{2}\right)^n+\frac{5+\sqrt5}{10}\left(\frac{1-\sqrt5}{2}\right)^n & \frac{\sqrt5}{5}\left(\frac{1+\sqrt5}{2}\right)^n-\frac{\sqrt5}{5}\left(\frac{1-\sqrt5}{2}\right)^n\\ \frac{\sqrt5}{5}\left(\frac{1+\sqrt5}{2}\right)^n-\frac{\sqrt5}{5}\left(\frac{1-\sqrt5}{2}\right)^n & \frac{5+\sqrt5}{10}\left(\frac{1+\sqrt5}{2}\right)^n+\frac{5-\sqrt5}{10}\left(\frac{1-\sqrt5}{2}\right)^n\end{bmatrix}$