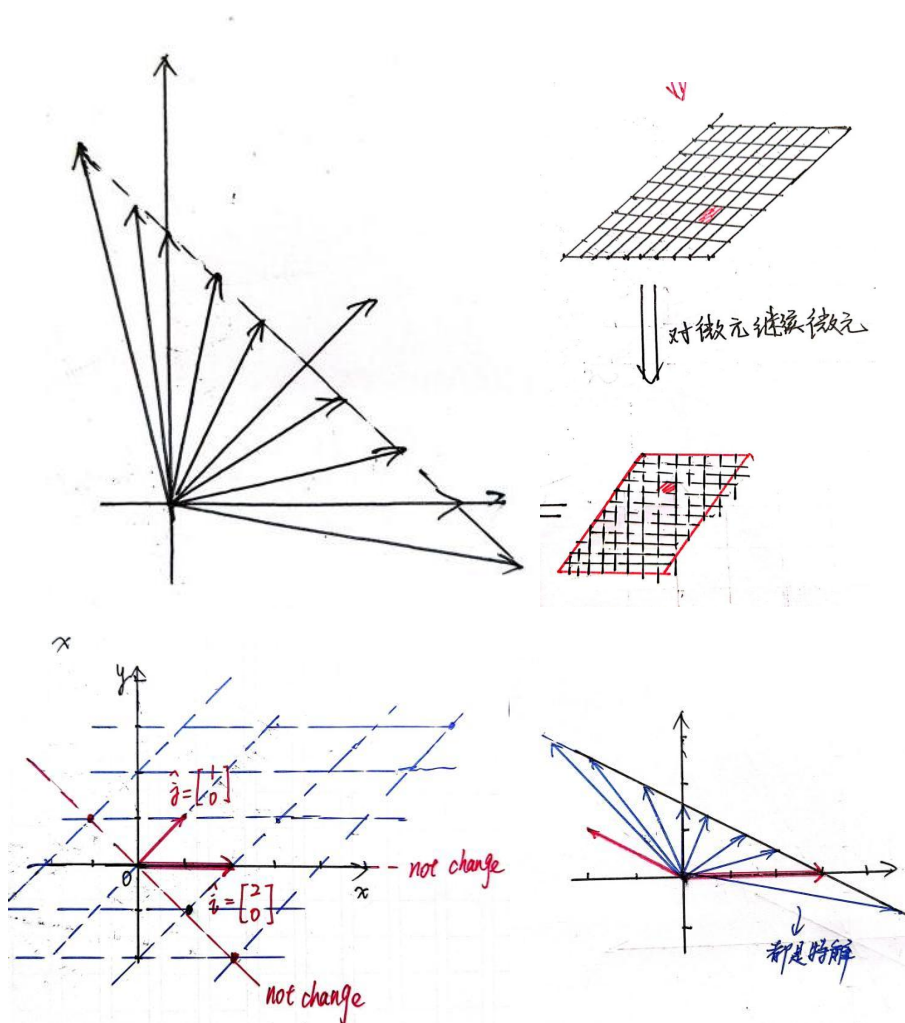


线性代数的本质



作者：徐文阳



北京邮电大学出版社
www.buptpress.com

线性代数的本质

Chap1——线性变换

- 1.1 向量与变换
- 1.2 矩阵与线性变换
- 1.3 矩阵乘法的本质
- 1.4 矩阵乘法——交换与结合的本质
- 1.5 矩阵乘法——非方阵相乘的本质
- 1.6 点乘的本质
- 1.7 矩阵乘法的性质——五种计算方法的本质

Chap2——线性方程组

- 2.1 用向量的眼光看待多元一次方程
- 2.2 消元法初步
- 2.3 消元法与矩阵的关系
- 2.4 逆的本质
- 2.5 可逆与不可逆的本质
- 2.6 矩阵的 LU 分解初步
- 2.7 转置的本质
- 2.8 线性相关的本质
- 2.9 零空间与通解的本质
- 2.10 通解与特解的本质

Chap3——线性空间

- 3.1 秩的本质

3.2 行空间，列空间，零空间的关联

3.3 投影矩阵的本质

3.4 施密特正交化的本质

3.5 行列式的本质

3.6 克拉默法则的本质

3.7 叉乘的本质

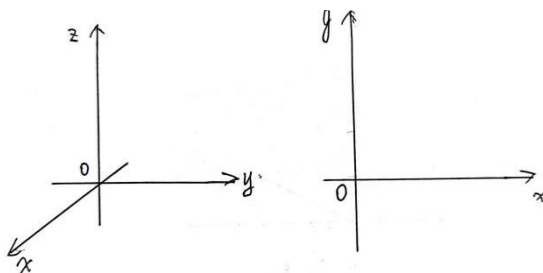
3.8 基变换的本质

3.9 特征向量与特征值的本质

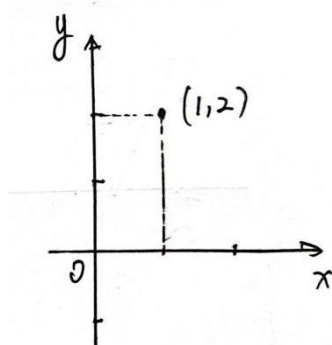
3.10 对称矩阵

Chap1.1: 向量与变换

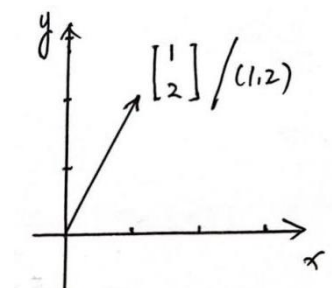
先来给出我们的舞台，坐标系：



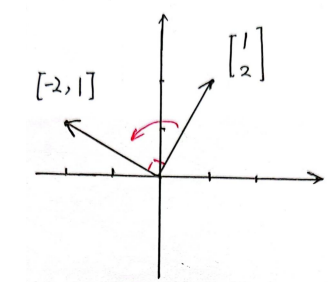
为了简单起见，我们先从平面直角坐标系开始讨论。我们可以用坐标的形式，来描述平面直角坐标系上的点，像这样：



但是，换一个角度，我们可以用“向量”来描述每一个点，像这样：

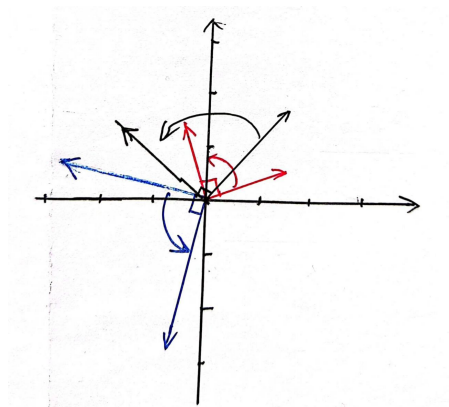


下面提出一个问题：对于 $(2, 1)$ 逆时针旋转 90° 后，得到的是什么向量？

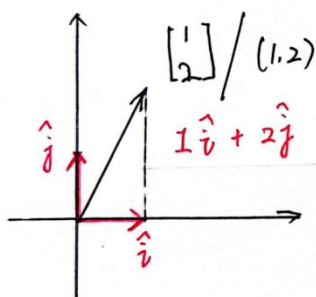


这是显然的，因为一眼可以看出，答案是 $(-1, 2)$

那么，如果对于任意一个 (x, y) ，其旋转 90° 后得到什么向量，是否可以想象呢？像这样：

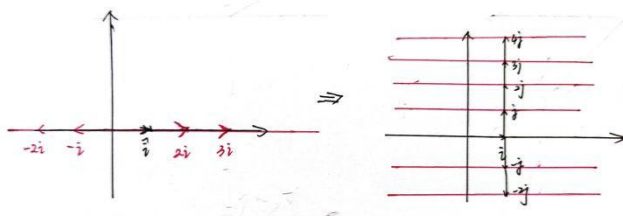


为了方便地解决这个问题，需要引入一个新的视角——“基向量”。比如上面的向量，或者 $(2,1)$ ，可以写成 $2\hat{i}+1\hat{j}$ ，其中 \hat{i} 和 \hat{j} 是 x 轴和 y 轴上的单位向量，像这样：

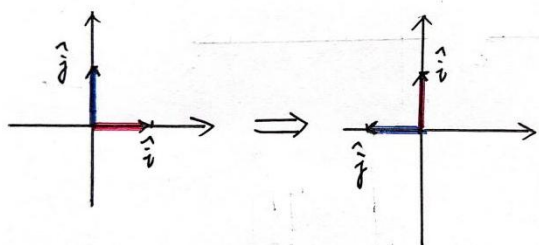


自然地，任何一个向量 (x,y) 都可以写作 $x\hat{i}+y\hat{j}$ ，注意，其中 x, y 是标量，而 \hat{i}, \hat{j} 才是真的向量（单位向量）。

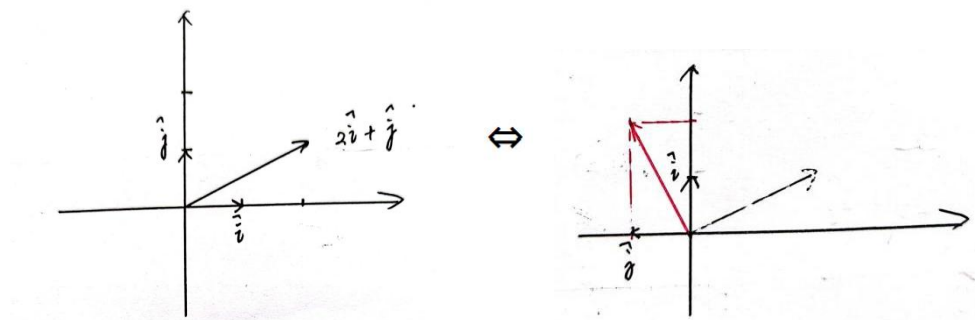
基向量产生 $x\hat{i}+y\hat{j}$ ，通过改变 x, y ，就可以表示平面中的所有向量。这是为什么呢？首先一个向量可以决定一条直线，例如 \hat{i} $(1,0)$ ，通过数乘一个数 a ，则 $a\hat{i}$ 可以代表 x 轴，而 $a\hat{i}+b\hat{j}$ ，引入 \hat{j} 后，意味着 x 轴可以平移，即“线动成面”。



既然基向量可以表示平面内的所有向量，那么如果我将基向量旋转 90° ，是否也就代表着平面中的所有向量都旋转了 90° 呢？像这样，先把基向量旋转：



那么基向量所代表的向量也进行了旋转，像这样：



于是，对于 $xi + yj$ ，我们只需要将 $i = (1,0)$ 改为 $(0,1)$ ， $j = (0,1)$ 改为 $(-1,0)$ ，如此，旋转后的向量就变成了：

$$p_0 = \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Leftrightarrow p_{\text{转}} = x \begin{bmatrix} 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 - y \\ x + 0 \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

这也就意味着， (x, y) 旋转之后，就变成了 $(-y, x)$ 。

观察一下，我们换一种写法：

$$p_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \Leftrightarrow p_{\text{转}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

是的，这就是矩阵，我们下一章再继续探索。

矩阵正如 $f(x)$ 中的 f ，提供了向量“变换”的方式。所以请记住：

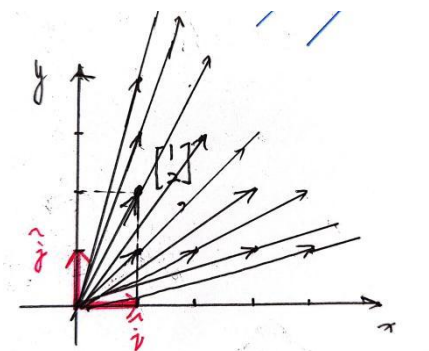
矩阵的本质就是操纵线性空间变换的工具

希望读者能从几何角度窥探矩阵的本质，而非沉醉于复杂的计算，这些事情交给计算机去做即可。正如数学家埃米尔所说：“根据我的经验，如果丢掉矩阵的话，那些涉及矩阵的证明可以缩短一半。”

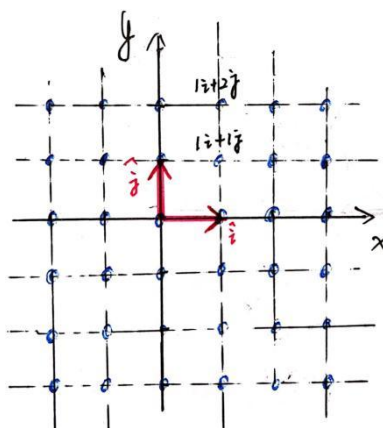
Chap1.2: 矩阵与线性变换

上一节，已经提到了矩阵是什么，并谈到了它可以旋转坐标系中的每一个向量。那么，矩阵还可以产生其他变换吗？

首先，为了体现出空间中所有向量的变换，我们需要先把空间中所有的向量画出来，像这样：

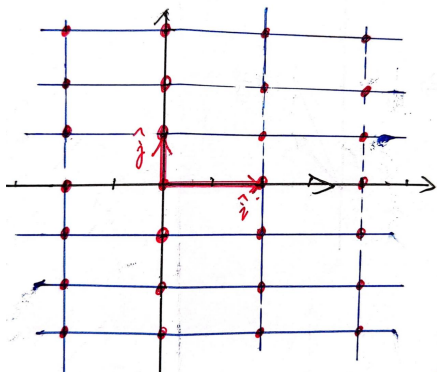


但是，这太乱了，每一个向量都是一根根线，这容易影响我们看到它的变换，于是我们使用点来代替向量，像这样：

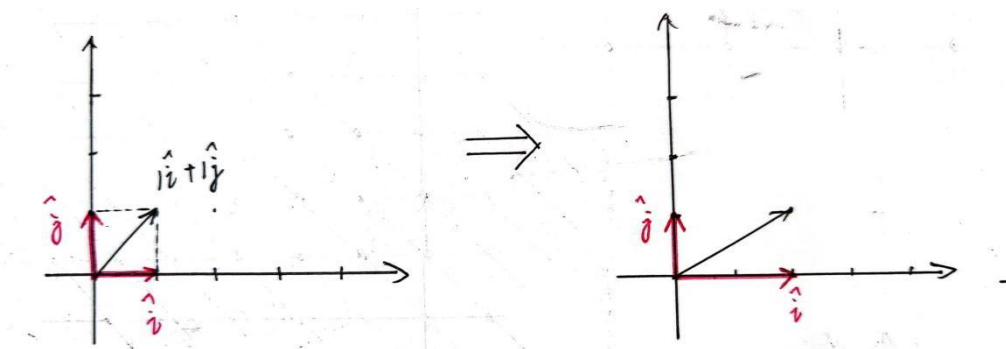


空间中的每一个点，就代表着一个向量。

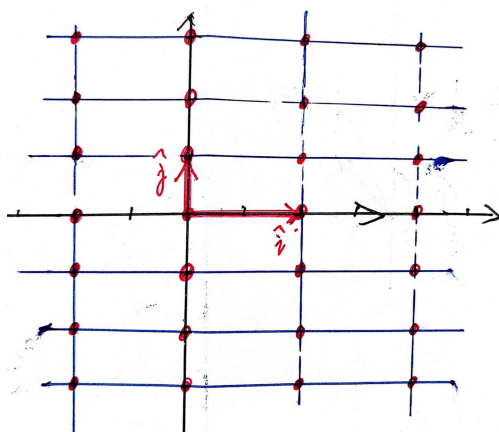
前面我们提到，空间中点的变换的本质就是“基向量”的变换，那我们不妨先找一个最简单的基向量变换， j 保持不变，而 i 伸长 2 倍：



那么比如空间中的一个向量 $1i+1j$ ，将会从坐标系的 $(1,1)$ 变成 $(1,2)$



而对于空间中的每一个点而言，将会变成这样：

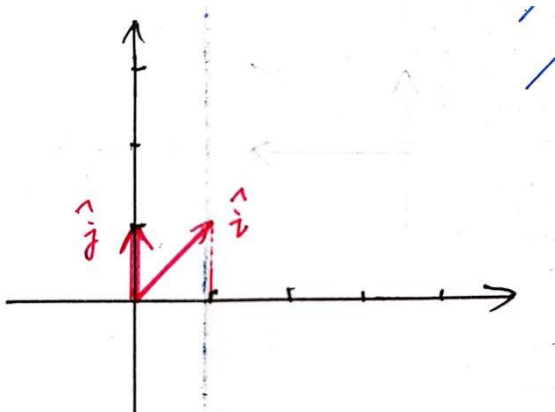


我们回顾一下上一节我们提到，矩阵的本质就是描述空间中向量的线性变换，那么在这个问题中如果 $\begin{bmatrix} x \\ y \end{bmatrix}$ 是变换前的向量，向量 p 是变换后向量，那么应当有：

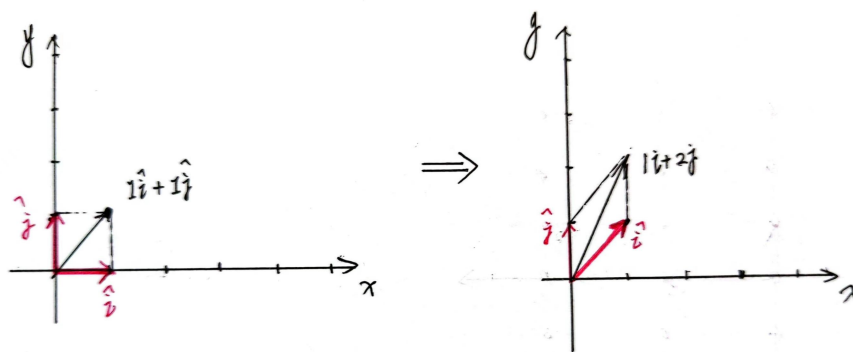
$$p = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y = \begin{bmatrix} 2x \\ y \end{bmatrix}$$

这代表着我们如果随意地输入一个向量，比如 $(2, 1)$ ，这个向量都将通过矩阵，输出一个向量 $(4, 1)$ ，这也正是矩阵的本质，只是一个线性变换的工具。

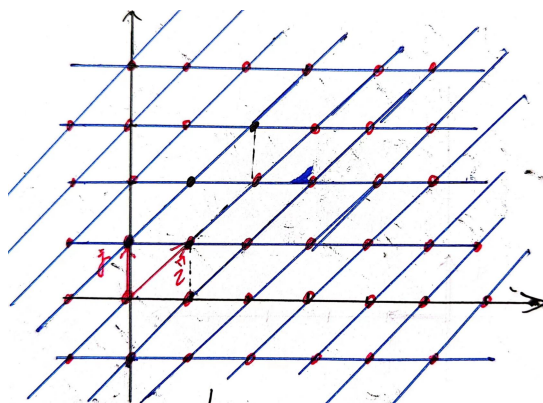
或许上面这个变换太过于直观，我们换一个稍微复杂一些的变换，将 i 从 $(1, 0)$ 变换到 $(1, 1)$ ，像这样：



那么，坐标系中的一个向量 $1i+1j$ ，将会从 $(1, 1)$ 变换到 $(1, 2)$ ，似乎提升了一个单位长度，像这样：



那么对于坐标系中的每一个向量，将会变成这样：

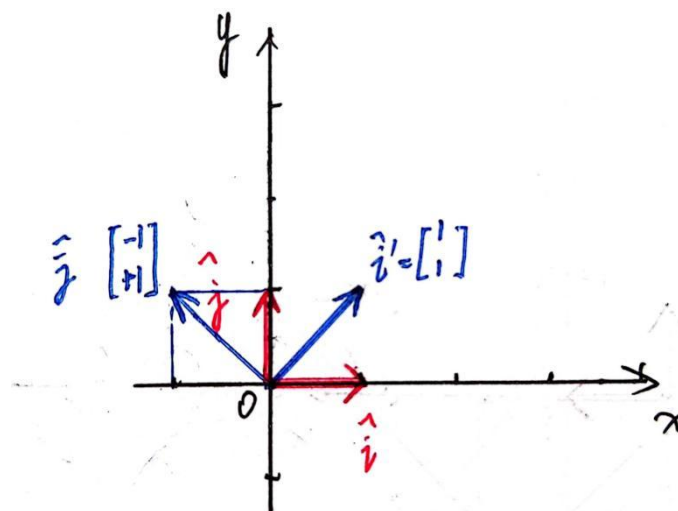


那么在这个问题中如果 $\begin{bmatrix} x \\ y \end{bmatrix}$ 是变换前的向量，向量 p 是变换后向量，那么应当有：

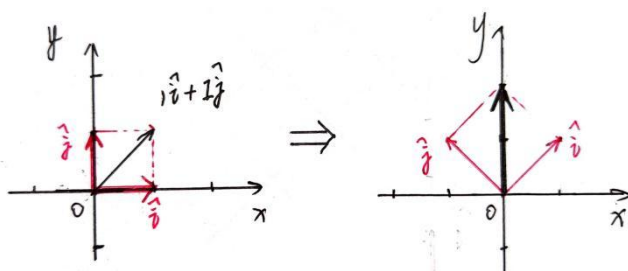
$$p = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y = \begin{bmatrix} x \\ x+y \end{bmatrix}$$

这代表着我们如果随意地输入一个向量，比如 $(2, 1)$ ，这个向量都将通过矩阵，输出一个向量 $(2, 3)$

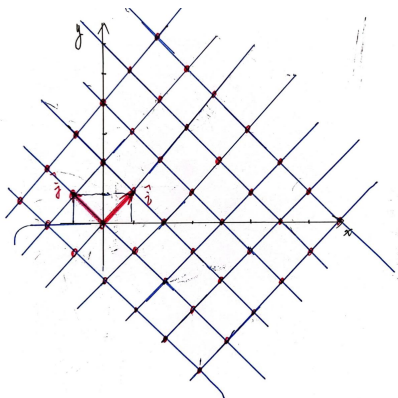
接着，我们继续加大难度，如果变换成为了 i 从 $(1, 0)$ 变换为了 $(1, 1)$ ， j 从 $(0, 1)$ ，变换到了 $(-1, 1)$ ，那么这样的变换如何呢？



比如说一个向量 $i\hat{i}+j\hat{j}$ 的变换，就像这样：



那么对于空间中每一个向量，它们的变换将会变成这样：



我们用矩阵来描述，也就变成了这样：如果 $\begin{bmatrix} x \\ y \end{bmatrix}$ 是变换前的向量，向量 p 是变换后向量，

那么应当有：

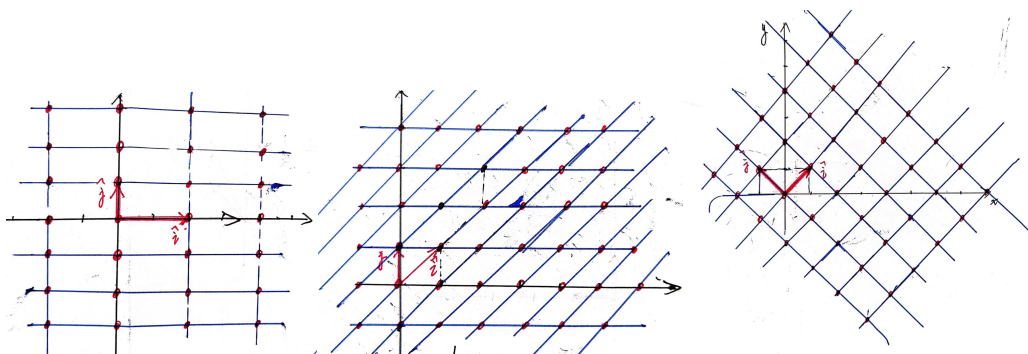
$$p = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} -1 \\ 1 \end{bmatrix} y = \begin{bmatrix} x-y \\ x+y \end{bmatrix}$$

这代表着我们如果随意地输入一个向量，比如 $(1, 1)$ ，这个向量都将通过矩阵，输出一个向量 $(0, 2)$ 。

事实上，例子已经够多了，我们可以通过矩阵给出向量的任意一种变换，像这样：

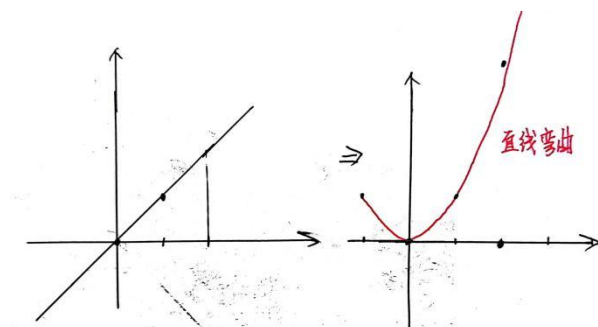
$$p = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} x + \begin{bmatrix} c \\ d \end{bmatrix} y = \begin{bmatrix} ax+cy \\ bx+dy \end{bmatrix}$$

于是，下面我们观察一下以上这三个变换，发现都有两个共同点：



- 1、由于基向量的起点都是原点，所以变换后，所有向量的起点也仍然是原点
- 2、由于都是基向量的加减和数乘，所以变换后，原坐标系中直线仍旧保持直线，像这

样，如果不是通过矩阵的一个随便的变换，将会长成这样：



其中输入 $\begin{bmatrix} x \\ y \end{bmatrix}$ 而输出 $\begin{bmatrix} x \\ y^2 \end{bmatrix}$ ，这一变换无法通过矩阵或者基向量完成，直线会弯曲。

但是，如果是一个通过矩阵的变换，一条直线变换后将仍然是一条直线。从直观上讲，其本质在于，矩阵的变换只是对基向量进行旋转，伸缩，所以向量仍然保持平直，并且不改变基向量位于原点的位置，所以整个空间仍然处于原点。

如果从直观上不好理解，依旧可以从代数角度来理解，比如说我们在原来最开始的基坐标下，一条直线的解析式为 $y = kx + m$ ，那么变换后的每一个点，都可以写作：

$$p = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x \\ kx + m \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} x + \begin{bmatrix} c \\ d \end{bmatrix} (kx + m) = \begin{bmatrix} ax + c(kx + m) \\ bx + d(kx + m) \end{bmatrix} = \begin{bmatrix} (a + ck)x + cm \\ (b + dk)x + dm \end{bmatrix}$$

于是，观察最后的坐标，其中的每一个 x 都是一次的，我们设新坐标系下的坐标为 (X, Y) ，那么原来的 x （小写），就可以看作是参数方程。

$$\begin{cases} X = (a+ck)x + cm \\ Y = (b+dk)x + dm \end{cases} \Rightarrow \frac{X - cm}{Y - dm} = \frac{a+ck}{b+dk} \Rightarrow X = \frac{a+ck}{b+dk} Y - dm \frac{a+ck}{b+dk} + cm$$

可见，仍然是一条直线，所以通过矩阵或者基向量产生的变换都保持着两条性质，这正是线性变换的定义：

1、原点保持不动。

2、直线在变换后仍然是直线。

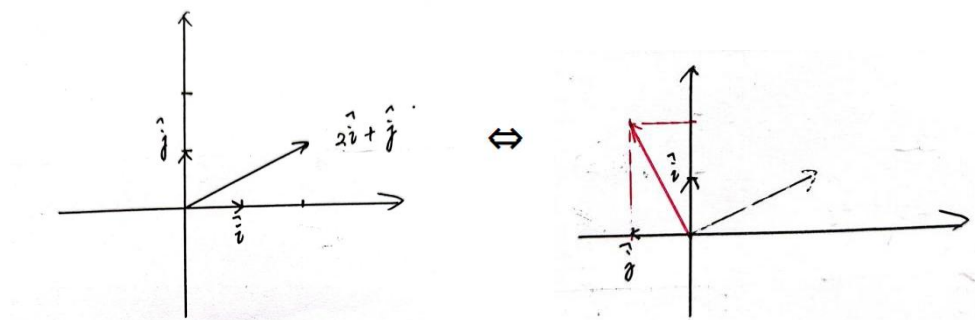
这也叫做“线性变换”，这正是《线性代数》这门课的核心内容。

Chap1.3 矩阵乘法的本质

上一节，我们谈到，矩阵的坐标运算长成这个样子：

$$p = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} x + \begin{bmatrix} c \\ d \end{bmatrix} y = \begin{bmatrix} ax + cy \\ bx + dy \end{bmatrix}$$

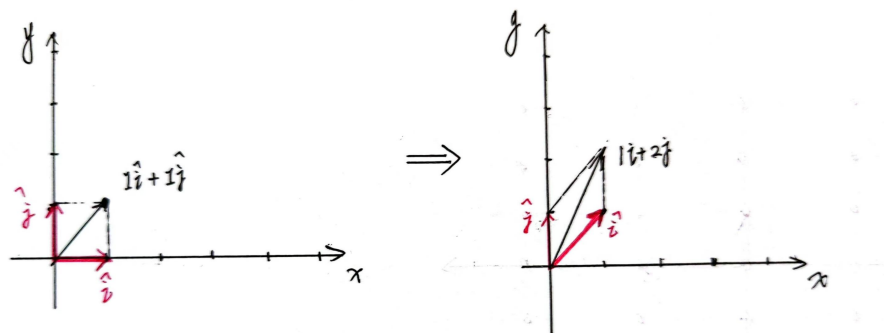
但是，在上一节，我们只考虑了“一次变换”，如果我们像对基同时进行两次变换呢？举个例子，在 Chap1.1，我们谈到，如何将一个向量进行旋转，我们称之为“旋转变换”。



$$p_{\text{转}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

其中参与“线性变换”的矩阵为 $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ，我们称之为旋转矩阵。

在 Chap1.2 中，我们介绍了三种变换，其中的一种长成了这样：



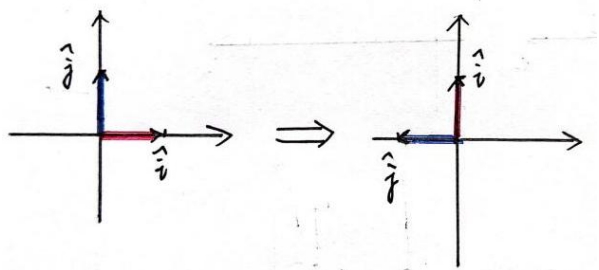
$$p_{\text{拉}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y = \begin{bmatrix} x \\ x + y \end{bmatrix}$$

我们发现基向量和空间中所有的向量似乎都向上拉伸了一些，于是我们称之为“拉伸变换”，其中参 $2 \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ ，我们称之为拉伸矩阵

此时，我们不禁思考，能不能对空间中的所有向量，先进行旋转变换，再进行拉伸变换呢？

首先，我们依旧回到“基向量”，为了让空间中所有的向量“先进行旋转变换，再进行拉伸变换”，我们只需要对基向量“先进行旋转变换，再进行拉伸变换”。

首先，我们不妨先进行一个旋转变换，像这样：



如果空间中的每一个向量都是 $\begin{bmatrix} x \\ y \end{bmatrix}$ ，那么也就意味着，现在所有的 $\begin{bmatrix} x \\ y \end{bmatrix}$ 已经变换到了

$\begin{bmatrix} -y \\ x \end{bmatrix}$ ，像这样：

$$P_{\text{转}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

那么，如果再进行一次“拉伸变换”呢？

$\begin{bmatrix} x \\ y \end{bmatrix}$ 乘上旋转矩阵实现了旋转的效果，于是很自然地，我们再乘一个拉伸矩阵，就再次

实现了拉伸的效果：

$$\text{即， } P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

但是问题是，应该如何计算两个矩阵的相乘呢？

我们不妨再次回到线性变换这一本质。所有的变换，本质都是基向量的变换，所以，我们只需要再对新产生的基向量 (i, j) 再次进行一次拉伸变换就可以了。在“旋转变换”之后，新的基向量分别是 (0,1) 和 (-1,0)，这一基向量隐藏在“旋转矩阵中”。

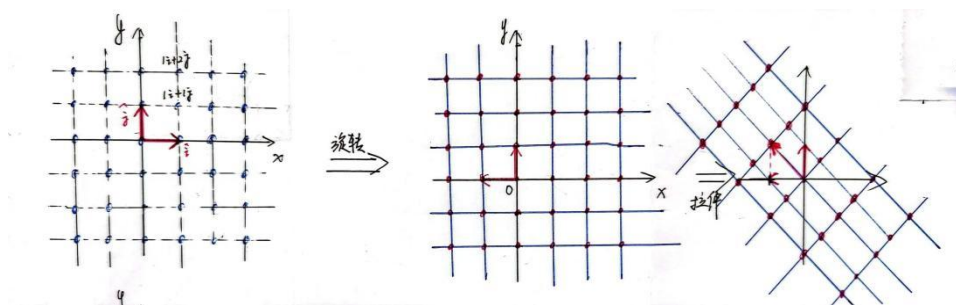
在 $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 中，基向量为 i (0,1)，j (-1,0)，如果将这两个向量分别作用于矩阵 $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ ，

那么就会得到如下式子：

$$\text{基向量} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

换句话说，我们通过这样的运算得到了基向量究竟是什么，我们设新产生的 i 和 j，那么就像这样：

$$\begin{aligned} i &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ j &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \end{aligned}$$



是的基向量中的 i 就是 $(-1,0)$ ，而 j 就是 $(0, -1)$ 。是的

如果我们这两基向量再次合并在一起，那么就获得了一个新的矩阵，综合了“旋转和拉伸”两个作用，我们称之为“复合矩阵”。

$$A = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$$

那么，回顾我们之前所谈到的这个式子，其实他也可以这样看

$$p = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \Leftrightarrow p = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

于是，很自然地，我们就定义了矩阵的乘法，即：

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

于是，如果我们想先用矩阵 $A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ ，完成一个变换，再用一个矩阵 $B = \begin{bmatrix} e & g \\ f & h \end{bmatrix}$

完成一个变换，这两个变换的叠加就是 AB ，这样正是矩阵的乘法。

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix}$$

下面我们来求解 AB 究竟是什么？

仿照之前的思路，我们认为 A 是在对 B 中的两个基向量进行线性变换，那么：

$$i = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} e + \begin{bmatrix} c \\ d \end{bmatrix} f = \begin{bmatrix} ae + cf \\ be + df \end{bmatrix}$$

$$j = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} g + \begin{bmatrix} c \\ d \end{bmatrix} h = \begin{bmatrix} ag + ch \\ bg + dh \end{bmatrix}$$

如果我们将两个基向量写在一起，于是将会得到：

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae + cf & ag + ch \\ be + df & bg + dh \end{bmatrix}$$

请记住：

矩阵的本质，在于通过操纵基向量，实现对线性空间的线性变换

矩阵的乘法，其本质在于对基向量进行两次变化，从而对线性空间进行两次线性变换；

Chap1.4 矩阵乘法的性质——交换与结合的本质

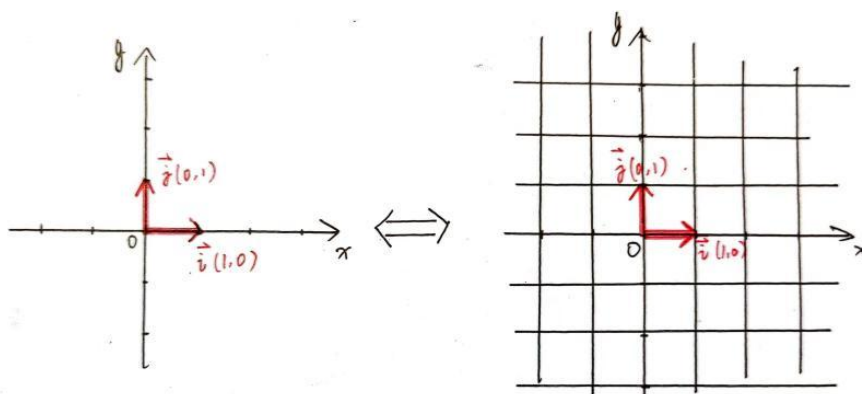
前面，我们已经得到了矩阵乘法的计算公式，

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae+cf & ag+ch \\ be+df & bg+dh \end{bmatrix}$$

一、矩阵运算满足交换律吗？

首先，我们提出一个问题：AB 等于 BA 吗？回归矩阵乘法的本质，是对用空间的变换。AB 的意思是：对基 $i(1,0)$ 和 $j(0,1)$ 先做 B 变换，再做 A 变换。BA 的意思是：对基 $i(1,0)$ 和 $j(0,1)$ 先做 A 变换，再做 B 变换。

请看，这就是我们初始的基。

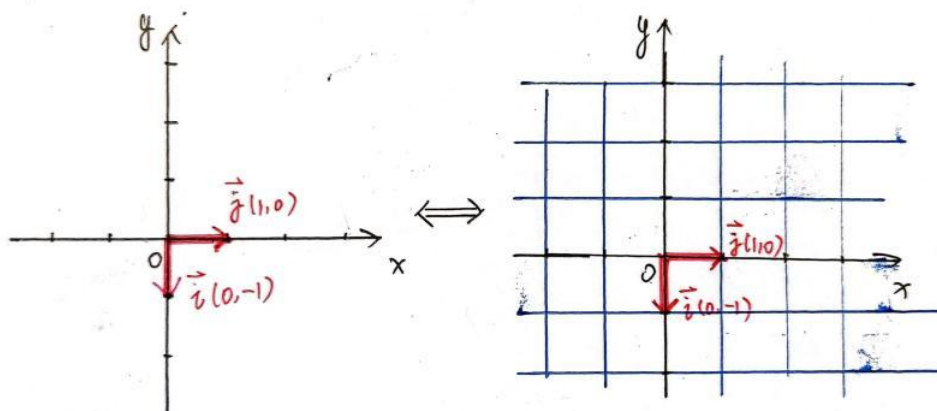


于是，最简单的想法是，假设 A 矩阵代表着将空间中的所有向量先顺时针旋转 90 度（即 $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ ），B 代表对空间中的所有向量向上拉伸（即 $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ ），那么，我们仍然

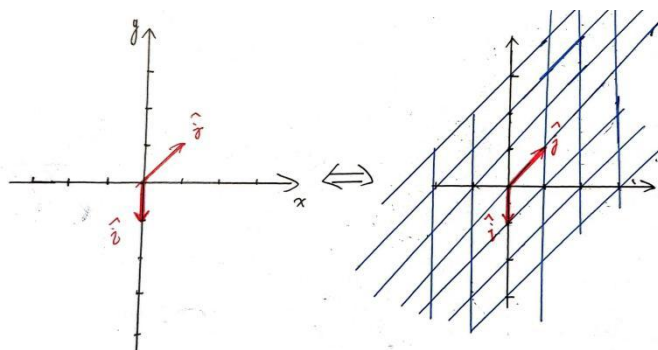
研究基的变换，像这样：

(1) BA:

先进行 A 变换（顺时针旋转 90°），空间将变成这样：

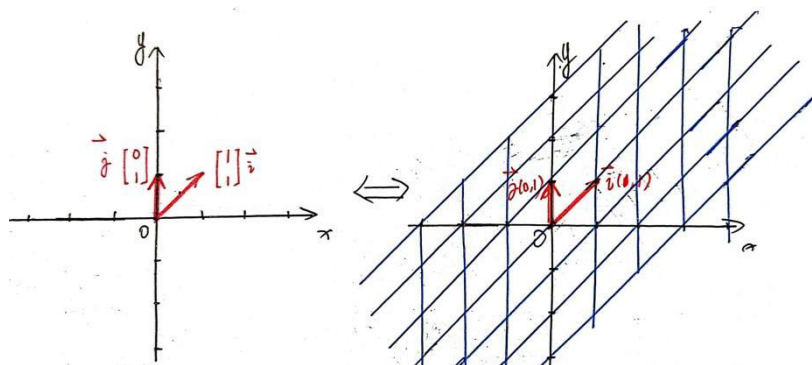


然后再进行 B 变换，即拉伸变换，空间将变成这样：

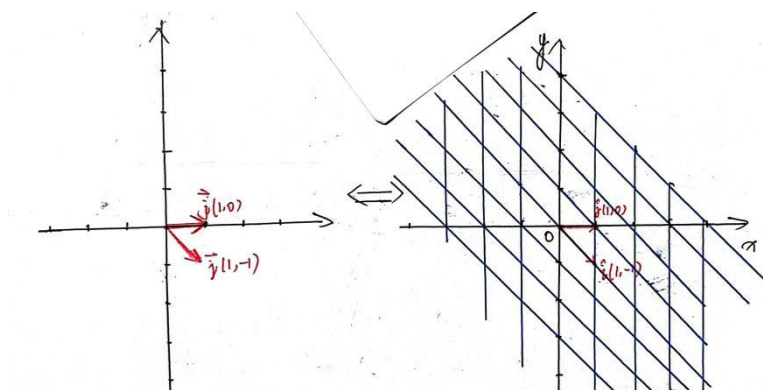


(2) AB:

先进行 B 变换，即向上拉伸，那么空间将会变成这样：



然后再进行 A 变换，即顺时针旋转 90° ，那么空间将变成这样：



从上面的例子中，我们可以看出，AB 和 BA 完全是对空间的两个变换！这是从几何角度的直观认识，我们再从代数角度进行验证：

$$AB = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

而 AB 新产生的两个基向量为：

$$\begin{aligned} i &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} 1 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} 1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ j &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} 0 + \begin{bmatrix} 1 \\ 0 \end{bmatrix} 1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned}$$

所以，AB 就算出来了，得到：

$$AB = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$$

是的

$$BA = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

而 AB 新产生的两个基向量为：

$$\begin{aligned} i &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 0 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (-1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ j &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} 0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

所以，AB 就算出来了，得到：

$$BA = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$$

所以，这也就得到了我们的第一条结论：

$$AB \neq BA$$

二、矩阵运算满足结合律吗？

首先，结合律意味着讨论 ABC 是否等于 (AB) C。

ABC 意味着先对基向量 i (1, 0)，j (0, 1) 进行 C 变换，再进行 B 变换，然后进行 A 变换。

(AB) C 同样意味着先对基向量 i (1, 0)，j (0, 1) 进行 C 变换，然后通过 AB，找到与 AB 效果相同的矩阵，再作用于经过 C 变换后的 i 和 j。

我们前面在讨论矩阵乘法的时候就已经明白 AB 的意思就是对一个基向量先进行 B，再进行 A，所以 (AB) C 和前面的 ABC 是同样的意思。

从几何直观上，我们便得到了结论：

$$ABC = (AB) C$$

而从代数上的验证比较复杂，计算量很大，最后算出来的确是正确的，我把这个证明放在了附录中，在文档的最后面！（注：这真是个愚蠢的证明！）

请记住：

矩阵的本质，在于通过基向量操纵线性空间

矩阵乘法的本质就是寻找到两次线性变换的等效矩阵

不符合交换律，本质是操纵线性空间的方式取决于顺序

符合结合律，本质是操纵线性空间的顺序压根没有改变，甚至没有任何变化

Chap1.5: 矩阵乘法的性质——非方阵相乘的本质

前面，我们提到，两个向量的相乘可以表示成这样：

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae+cf & ag+ch \\ be+df & bg+dh \end{bmatrix}$$

而如果将这一结论拓展到三维，思考问题的方式依旧是相同的，即看作是三维空间内的线性变换，只不过多增加了一个基向量，变成了 $i(1,0,0)$ ， $j(0,1,0)$ ， $k(0,0,1)$ 。像这样：

$$\text{例如，矩阵 } A = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & I \end{bmatrix}, \text{ 矩阵 } B = \begin{bmatrix} j & m & p \\ k & n & q \\ l & o & r \end{bmatrix}, \text{ 这两个矩阵相乘，得到：}$$

$$A = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & I \end{bmatrix} \begin{bmatrix} j & m & p \\ k & n & q \\ l & o & r \end{bmatrix}$$

我们计算一下在这个三维空间内新产生的三个基向量，得：

$$\begin{aligned} i &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} j \\ k \\ l \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} j + \begin{bmatrix} d \\ e \\ f \end{bmatrix} k + \begin{bmatrix} g \\ h \\ i \end{bmatrix} l = \begin{bmatrix} aj+dk+gl \\ bj+ek+hl \\ cj+fk+il \end{bmatrix} \\ j &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} m \\ n \\ o \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} m + \begin{bmatrix} d \\ e \\ f \end{bmatrix} n + \begin{bmatrix} g \\ h \\ i \end{bmatrix} o = \begin{bmatrix} am+dn+go \\ bm+en+ho \\ cm+fn+io \end{bmatrix} \\ k &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} p + \begin{bmatrix} d \\ e \\ f \end{bmatrix} q + \begin{bmatrix} g \\ h \\ i \end{bmatrix} r = \begin{bmatrix} ap+dq+gr \\ bp+eq+hr \\ cp+fq+ir \end{bmatrix} \end{aligned}$$

如果我们将两个基向量写在一起，于是将会得到：

$$AB = \begin{bmatrix} aj+dk+gl & am+dn+go & ap+dq+gr \\ bj+ek+hl & bm+en+ho & bp+eq+hr \\ cj+fk+il & cm+fn+io & cp+fq+ir \end{bmatrix}$$

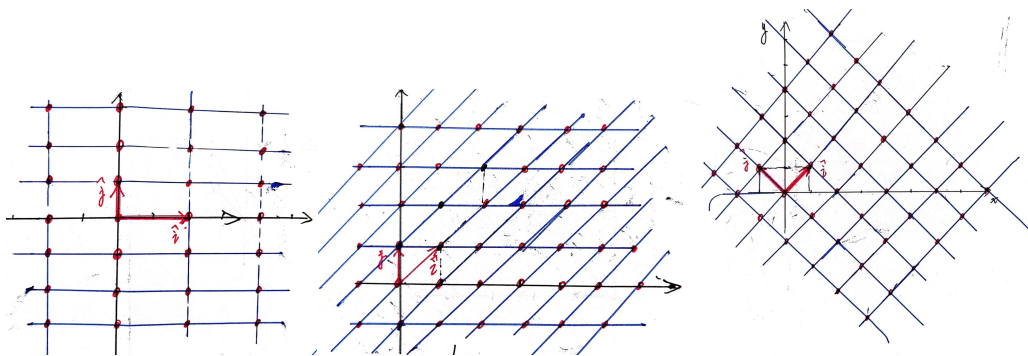
所以我们就获得了三维空间内的一个新的基向量，同时经过了 B 和 A 两个线性变换。

下面，我们继续回归到几何直观的内容，

在 Chap1.1 中，我们提到，如果确定了 $i(1,0)$ ， $j(0,1)$ ，那么 $xi+yj$ 可以代表平面直角坐标系中的全部向量。因为假设我们先不考虑 j ，只让 i 做伸长和压缩，即 xi 。那么 xi 就代表着整条数轴，从而“点动成线”。而加上一个 yj ，就意味着将数轴 (xi) 上下平移，从而“线动成面”。

而经过“矩阵”线性变换的新产生的基向量似乎也可以布满整个平面，像这样：

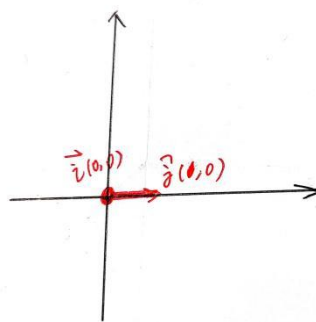
其中的每一个新产生的基向量所覆盖的空间都是整个二维空间。



但是，新产生的基向量一定能布满整个二维空间吗？我们不妨回顾为什么“两个向量可以布满二维空间”，因为“点动成线，线动成面”，所以，如果能“点动不成线”或者“线动不成面”，那么这两个基向量将无法张成整个二维空间。

(1) “点动不成线”

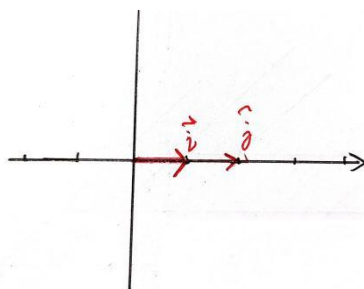
比如说，我们让 i 为 $(0, 0)$ ，即零向量，像这样：



那么 i 和 j 的线性组合就永远只能是 x 轴，只能是一条线，因为 xi 永远是 0，点动不能成线。

(2) “线动不成面”

这也就意味着两个向量都可以产生一条线，但是这条线却无法移动。是的，这是 i 和 j 是共线的，自然不能“线动成面”，像这样：



比如 i 是 $(1, 0)$ 而 j 是 $(2, 0)$ ，那么， $xi+yj$ 也只能是 x 轴。

于是，我们自然地想知道，是否存在一种矩阵，使得 $i(1, 0)$ 与 $j(0, 1)$ 经过变换之后，产生的新的基向量为 $i(1, 0)$ 与 $j(2, 0)$ 。 $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$

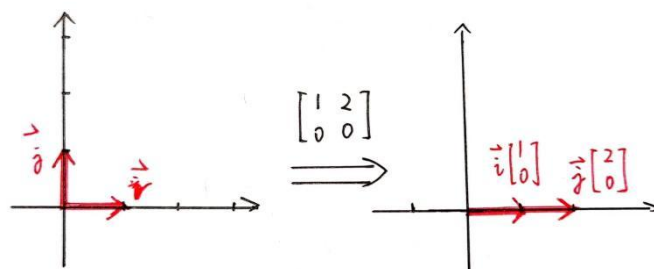
如果我们把这件事写成矩阵的形式也就变成了这样

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

显然，我们看出，只要让 (a, b) 为 $(1, 0)$ ，而 (c, d) 为 $(0, 1)$ 即可。也就意味着

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

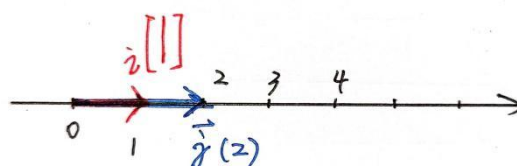
也就意味着如果我们直接将 $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ 与 $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 相乘，那么将会得到基向量 i 为 $(1, 0)$ ，并且基向量 j 为 $(2, 0)$ ，像这样：



事实上，其本质就在于仍然是一条处于二维空间内的直线，在计算的过程中， $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ 中的两个“0”，似乎没有起到任何作用，于是我们不妨定义这样一种很新的矩阵—— $A = \begin{bmatrix} 1 & 2 \end{bmatrix}$ ，这是什么？以及我们可否直接用 A 与 B 相乘呢？

前面的讨论中，我们都在讨论长得一样的矩阵，即 2×2 的矩阵与 2×2 的矩阵相乘，以及 3×3 的矩阵与 3×3 的矩阵相乘，那么 1×2 的矩阵是否能用与 2×2 的矩阵相乘呢？

我们继续来从几何含义分析，首先 $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 的两列分别代表着两个基向量，即 $(1, 0)$ 和 $(0, 1)$ ，而 $A = \begin{bmatrix} 1 & 2 \end{bmatrix}$ ，这是一个矩阵，当然也可以将它的列理解为基向量，只不过这是一维空间中的两个向量 $i = \begin{bmatrix} 1 \end{bmatrix}$ 和 $j = \begin{bmatrix} 2 \end{bmatrix}$ ，像这样：



首先，我们再次回归 Chap1.1 中矩阵究竟是什么的讨论： $\begin{bmatrix} x \\ y \end{bmatrix}$ 是二维空间中的任意一个向量，如果经过一个矩阵的作用 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ，那么它将会变成这个样子：

$$p_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y, \text{ 也就是说, 任意输入一个 } (x, y) \text{ 都会输出一个向量,}$$

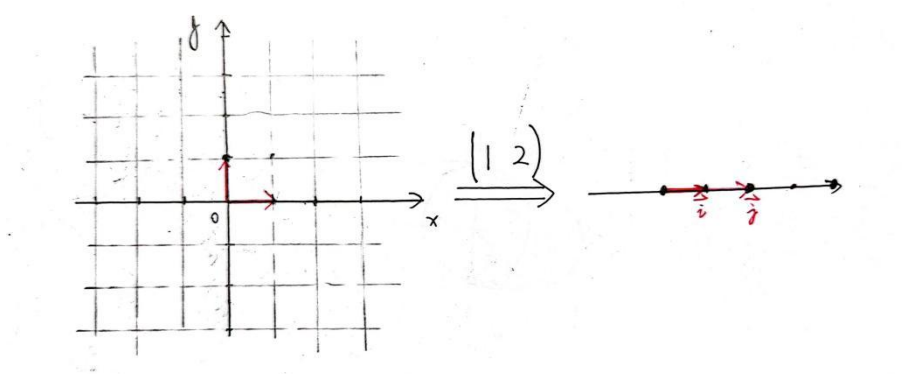
输出的向量就是 $\begin{bmatrix} x \\ y \end{bmatrix}$ 。

那么同样的道理，对于 $A = \begin{bmatrix} 1 & 2 \end{bmatrix}$ ，而言，同样地我们用刚刚的思路，也就变成了这样：

$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ ，代表着向着矩阵 $A = \begin{bmatrix} 1 & 2 \end{bmatrix}$ 输入一个向量 $\begin{bmatrix} x \\ y \end{bmatrix}$ ，而输出的向量将会是

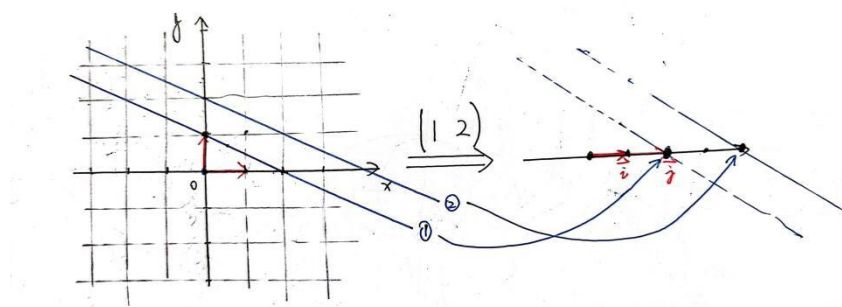
$[x + 2y]$ 。不知道你是否发现， $\begin{bmatrix} x \\ y \end{bmatrix}$ 本来是二维空间中的一个向量，在 A 的作用下变成了一维。

在几何上，这代表着将二维空间压缩到了一条线，像这样：



如果你不能接受这么草率的解释，我可以告诉你，在左侧的二维空间中的每一条直线，被压缩到了一维空间的一个点。

像这样：



由于 $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x + 2y]$ ，所以我们只需要找出来 $x + 2y$ 为定值的直线，就可以找到

对应右侧一维空间上的向量。

这也正是不同维度矩阵相乘的直观体现，即降维。

请记住：

矩阵的本质在于通过对基向量的操纵，改变线性空间
而非方阵矩阵的本质在于通过将基向量重合，使得线性空间压缩至更低的维度

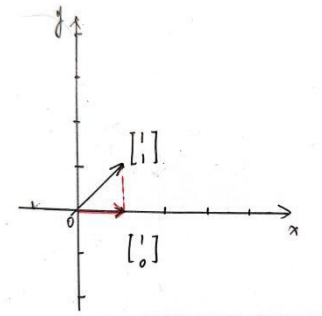
Chap1.6: 点乘的本质

初中的时候，我们明白垂直不做功。

高中的时候，我们明白了点乘的计算公式，似乎明白了垂直为什么不做功。

$$\text{点乘} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \times 1 + 1 \times 0 = 1$$

像这样：



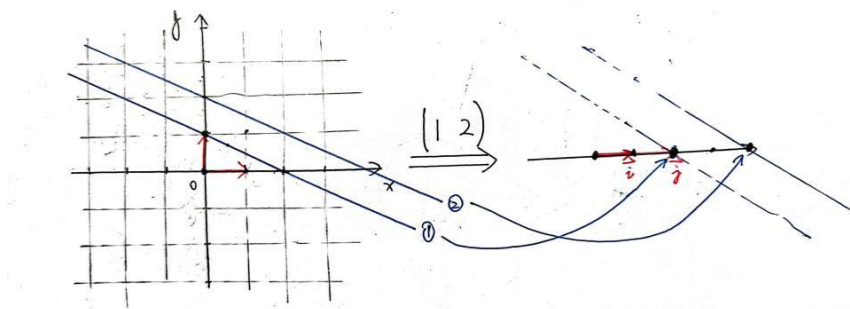
并且，关注到，向量 $(1, 1)$ 在 $(1, 0)$ 的投影也是 1，也就是说：

$$\text{点乘} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = (1, 0) \text{ 的长度} \times (1, 1) \text{ 的投影向量的长度} = 1$$

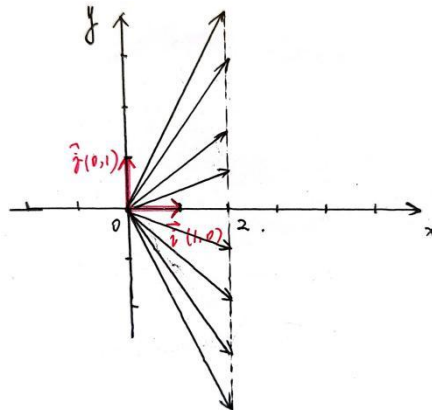
但是，我们是否想过为什么“点乘”会和“投影”有关？

结合上一节，是否有一些想法呢？

上一节中，我们通过矩阵 $\begin{bmatrix} 1 & 2 \end{bmatrix}$ ，将二维空间“斜着”投影到了 x 轴上，像这样：



那么如何“垂直着”投影呢？像这样：



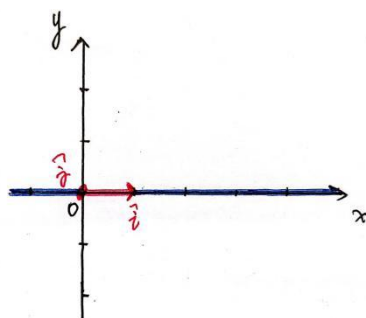
我们发现，当投影到 x 轴时，向量的横坐标没有发生变换，而纵坐标无论是什么，都变成了 0，这也就意味着，对于 (x, y) 向量，实现了下面的变换：

$$xi + yj \Leftrightarrow xi + 0j = xi$$

如果写成矩阵的形式，也就意味着 $i(1, 0)$ 和 $j(0, 1)$ ，经过了这个矩阵的作用：

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

事实上，如果依然把 i 看作 $(1, 0)$ ，把 j 看作 $(0, 0)$ 在几何上含义是这样的——仍然处在二维空间当中，但是知识二维空间中的一条直线，像这样：



现在仍然是在二维空间中的一条直线，因为两个基向量仍然有两个坐标，如果我们将矩阵 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ 改成矩阵 $\begin{bmatrix} 1 & 0 \end{bmatrix}$ ，那么将会将二维空间完全地压缩至一个一维空间。

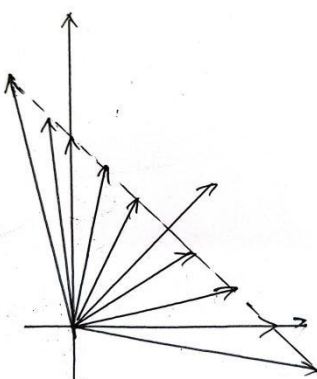
但实际上，和上一节说的一样，可以把 i 看作 (1) ，把 j 看作 (0) ，那么其将只处于一维空间当中，像这样：



如果写成向量来描述这件事，那么将会是这样：

$$x[1] + y[0] = x$$

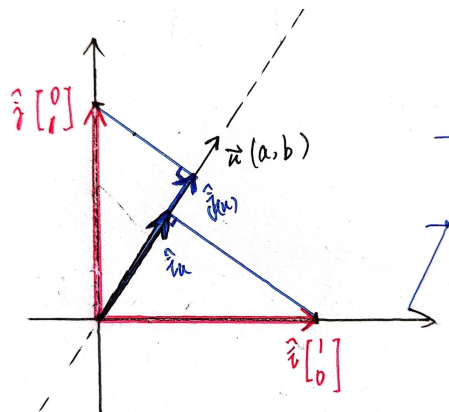
那么，如果不是投影在 x 轴上，而是投影在空间中的任意一个向量上呢？像这样：



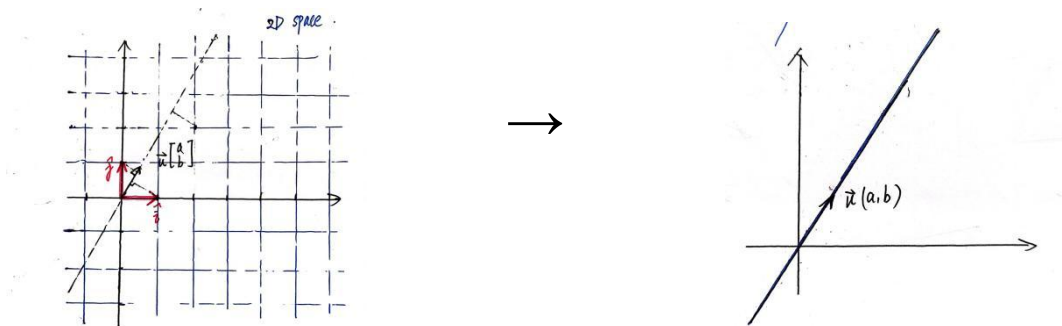
首先，我们明白，对于基向量的变换，决定了线性空间的变换。前面，我们将基向量 j

$(0,1)$ 垂直投影到了 $(0,0)$ ，如此，整个线性空间就垂直投影到了 x 轴。

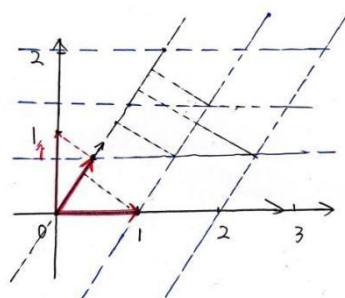
那么，我们是否可以将 $i(1,0)$ 和 $j(0,1)$ 垂直投影到任意一个倾斜的直线上，从而使得整个线性空间都投影到这条直线上呢？像这样：



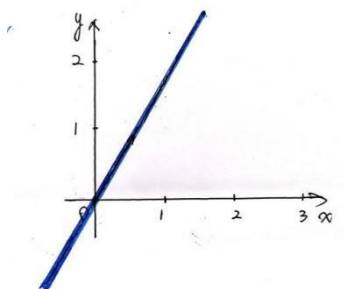
这是否意味着线性空间也发生了相同的变化呢？



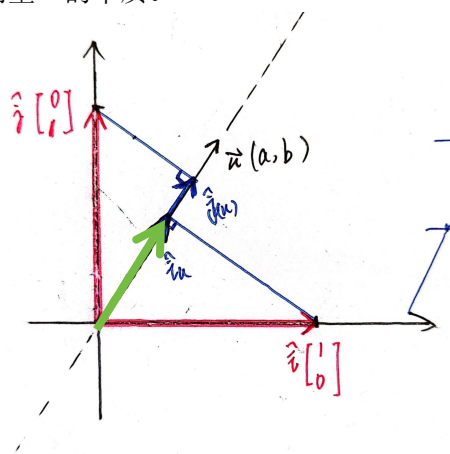
我们先来试一试，首先，如果先让 j 投影到 u 所在的直线上，于是线性空间将会像这样：



然后，再将 i 投影时，空间中的所有向量似乎都垂直地投影再来 j' 所在的直线上像这样：



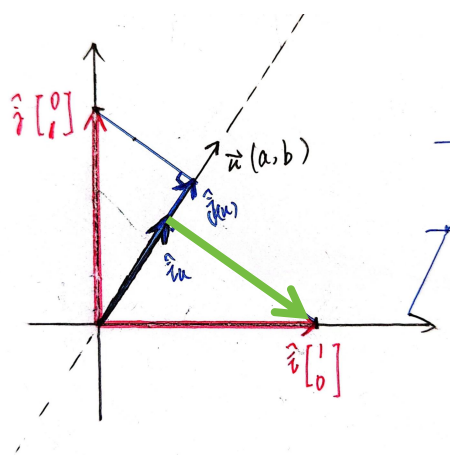
但是这仍然无法确定，是否在原来的空间中满足“垂直的变换”。
我们先来再次回顾“向量”的本质。



首先，我们将 $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 分解为沿着 u 直线方向和垂直于 u 直线方向，那么首先关注绿色

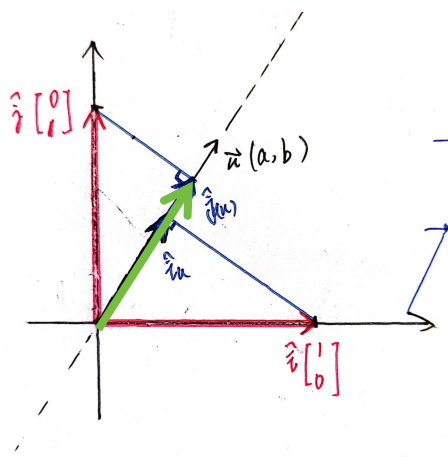
的箭头，它的意思是先从 $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 运动到那个垂点。

接着，再是第二段向量，即垂直于 u 直线的方向运动到 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ，像这样：

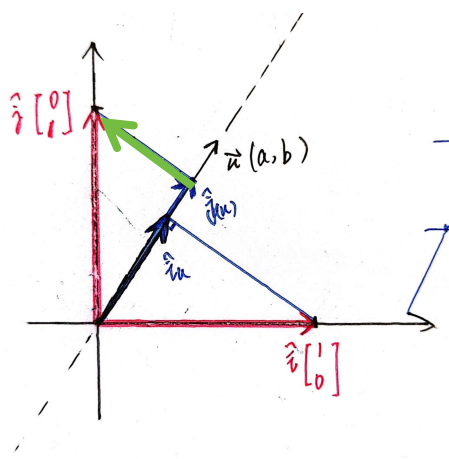


而将 $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 只看作沿着 u 直线方向的 i_u ，意味着，不再“走”垂直的那一段路程。

同理，我们可以用同样的方法研究 $j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ，如果将它分解到沿着 u 的方向的向量，和垂直方向的向量，那么它的第一步将会是，像这样：



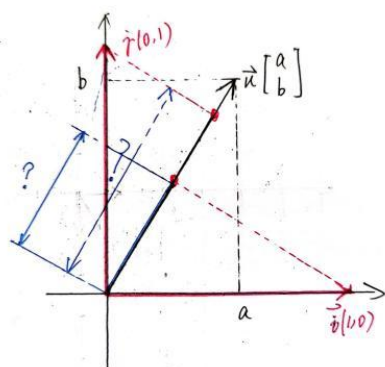
是的，这意味着先沿着 u 前进，到垂点处，然后再垂直着 u 前进，到 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ，像这样：



是的，如果我们只看在 u 方向的 j_u 的话，那么垂直方向的运动将会被抹去。是的，不知道你是否发现了这一规律。如果我们只看 i_u 和 j_u 的话，也就意味着只是沿着 u 这条直线移动，而不垂直于它移动。如果这一路径倒过来，不就意味着垂直地走回来吗？是的，这就意味着所有的向量将会全部垂直地投影到 u 这条直线之上！

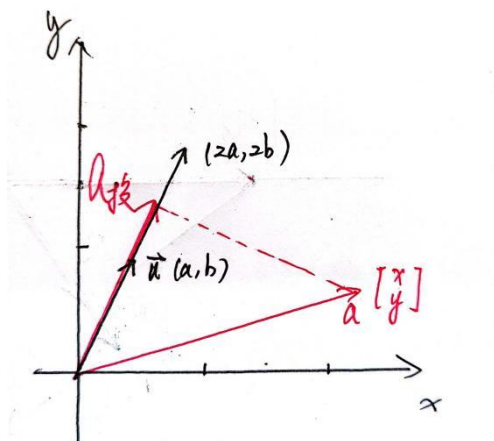
牛逼吗？抓住向量的本质思考问题，跳过了多少复杂的计算啊！少年，你感受到向量的力量了吗？

下面，我们的问题来到，如何求解两个基向量投影在直线后的长度，也就是解决这幅图中的两个问号，究竟是几？



首先我们要主要到的是矩阵 $\begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix}$ ，即代表一个矩阵，又是图中的 $\mathbf{u} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$ ，即单位向量。这也就意味着，矩阵和向量之间似乎可以相互转化。

自然地，我们先用 $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ 在矩阵的 $\begin{bmatrix} \mathbf{a} & \mathbf{b} \end{bmatrix}$ 作用下，投影在 \mathbf{u} 这条直线上，像这样：


$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [ax + by]$$
$$2[ax + by] = [2ax + 2by]$$

如果我们将上述过程合在一起，将会变成这样：

$$\begin{bmatrix} 2a & 2b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [2ax + 2by]$$

是的，你是否已经发现了什么？而 $2ax+2by$ 代表着什么？是的没错，就代表着投影向量的乘积。因为我们先将 (x, y) 投影到了 u 所在直线上，然后与 u 的长度相乘，从而得到了投影向量长度的乘积。

下面我们把这个问题再一般化，假设有一个向量 $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ 以及另一个向量 $\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ ，那么这两个

向量投影的乘积又是多少呢？

是的，我们可以先将其中一个向量看作矩阵，像这样：

$$\begin{bmatrix} x_1 & y_1 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

那么，我们就可以将矩阵 $\begin{bmatrix} x_1 & y_1 \end{bmatrix}$ 的作用理解成将向量 $\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ ，投影到向量 $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ 所在直

线上，并且与向量 $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ 的长度相乘。

而如果我们将上述讨论结合在一起，于是得到：

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = x_1x_2 + y_1y_2$$

这也正是点乘的本质。

请记住：

矩阵的本质是通过操纵基向量，从而操纵线性空间

向量的本质是“点的运动”

向量点积的本质是通过矩阵将线性空间垂直地投影到一个向量所在的直线上

Chap1.7 矩阵乘法的性质——五种计算方法的本质

前面，我们利用矩阵的本质，即进行“线性变换”的工具这一想法，得出来了矩阵的计算的两种方法。

(1) 第一种算法——计算机的算法，机械而愚蠢

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae+cf & ag+ch \\ be+df & bg+dh \end{bmatrix}$$

(2) 第二种算法——列向量的想法，看成对基向量的操作

任意矩阵 $A = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$, $B = \begin{bmatrix} j & m & p \\ k & n & q \\ l & o & r \end{bmatrix}$, 这两个矩阵相乘，得到：

$$AB = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} j & m & p \\ k & n & q \\ l & o & r \end{bmatrix}$$

我们计算一下在这个三维空间内新产生的三个基向量，得：

$$\begin{aligned} i &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} j \\ k \\ l \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} j + \begin{bmatrix} d \\ e \\ f \end{bmatrix} k + \begin{bmatrix} g \\ h \\ i \end{bmatrix} l = \begin{bmatrix} aj+dk+gl \\ bj+ek+hl \\ cj+fk+il \end{bmatrix} \\ j &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} m \\ n \\ o \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} m + \begin{bmatrix} d \\ e \\ f \end{bmatrix} n + \begin{bmatrix} g \\ h \\ i \end{bmatrix} o = \begin{bmatrix} am+dn+go \\ bm+en+ho \\ cm+fn+io \end{bmatrix} \\ k &= \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} p + \begin{bmatrix} d \\ e \\ f \end{bmatrix} q + \begin{bmatrix} g \\ h \\ i \end{bmatrix} r = \begin{bmatrix} ap+dq+gr \\ bp+eq+hr \\ cp+fq+ir \end{bmatrix} \end{aligned}$$

如果我们将两个基向量写在一起，于是将会得到：

$$AB = \begin{bmatrix} aj+dk+gl & am+dn+go & ap+dq+gr \\ bj+ek+hl & bm+en+ho & bp+eq+hr \\ cj+fk+il & cm+fn+io & cp+fq+ir \end{bmatrix}$$

注意：前两种算法是从“本质”上思考，而后面三种算法是本质算法的拓展，它们不具

有良好的几何直观，知识从代数的角度上，恰好和上面的计算结果相同。

(3) 第三种算法——行向量的想法

可以说，矩阵的算法符合这个规则，而为什么符合这个规则，并不存在直观的证明。前面的计算机算法是这样的：

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae+cf & ag+ch \\ be+df & bg+dh \end{bmatrix}$$

我们可以这样看：

对于 $ae+cf$ 和 $ag+ch$ （即第一行）而言，其可以看做：

$$\begin{aligned} & a \begin{bmatrix} e & g \end{bmatrix} \\ & + \\ & c \begin{bmatrix} f & h \end{bmatrix} \\ & = \\ & \begin{bmatrix} ae+cf & ag+ch \end{bmatrix} \end{aligned}$$

换句话说，可以这样理解，我的这一行，需要 a 倍的 $(e \ g)$ 并且需要 c 倍的 $(f \ h)$ ，这在后面理解初等行变换会相当有用。

（4）第四种算法——点乘的想法

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} ae+cf & ag+ch \\ be+df & bg+dh \end{bmatrix}$$

而它的基向量的运算是这样的：

$$\begin{aligned} i &= \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} e + \begin{bmatrix} c \\ d \end{bmatrix} f = \begin{bmatrix} ae+cf \\ be+df \end{bmatrix} \\ j &= \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} g + \begin{bmatrix} c \\ d \end{bmatrix} h = \begin{bmatrix} ag+ch \\ bg+dh \end{bmatrix} \end{aligned}$$

前面，我们都是基于基向量进行的推导，而这一次，我们不再写出基向量，而是先分析一下这些基向量在运算时的特点：

例如 i 的运算，我们只关注 $ae+cf$ 是如何计算出来的。

$$i = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} e + \begin{bmatrix} c \\ d \end{bmatrix} f = \begin{bmatrix} ae+cf \\ be+df \end{bmatrix}$$

可以看出，是左侧的 (a, b) 中的 a 与 e 相乘，而右侧的 (c, d) 中的 d 与 f 相乘，那么我是否可以将 i 的计算写成这个样子？

$$i = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} e + \begin{bmatrix} c \\ d \end{bmatrix} f = \begin{bmatrix} \begin{bmatrix} a & c \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} \\ \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} \end{bmatrix} = \begin{bmatrix} ae+cf \\ be+df \end{bmatrix}$$

也就意味着，从原本的把“向量”拆开，变成了将“矩阵”拆开。于是矩阵的上下两个位置就变成了向量的点乘（这和上一节所说是一样的）。

这也正是第三种计算方法，如果我们将它展开写就会变成这样：

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a & c \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} & \begin{bmatrix} a & c \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \\ \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} & \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \end{bmatrix}$$

请注意，我们这里并没有讨论：“为什么能够这样计算”，只不过矩阵的乘法恰好符合这种计算规则。

于是我们得出了第三种计算方法，将矩阵乘法，看作是行向量与列向量的点乘。

(5) 第五种算法——矩阵分块的计算方法

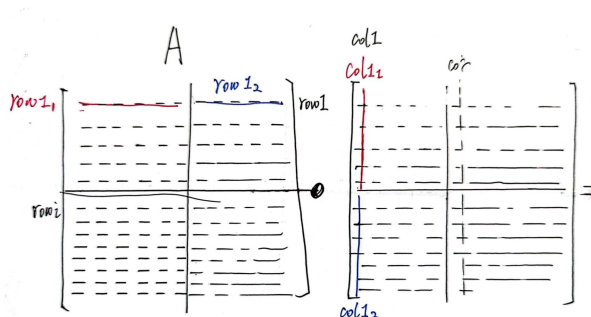
通过前面（3）例子的讨论，你是否已经看出一些“矩阵”相互组合的例子？

我们似乎已经将矩阵进行了分割，分割出了小矩阵，像这样：

$$AB = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a & c \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} & \begin{bmatrix} a & c \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \\ \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} & \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \end{bmatrix}$$

这的确是成立的，而这是不存在几何上的直观证明，只能作为矩阵运算产生的一个特殊性质。

我们可以通过这样的代数角度来理解：



例如，左侧矩阵的 row1 与右侧矩阵的 col1 可以组合，产生结果矩阵的 (1,1) 位置，像这样：

$$\begin{bmatrix} \text{row1} \cdot \text{col1} \\ \vdots \\ \text{rowi} \cdot \text{coli} \end{bmatrix}$$

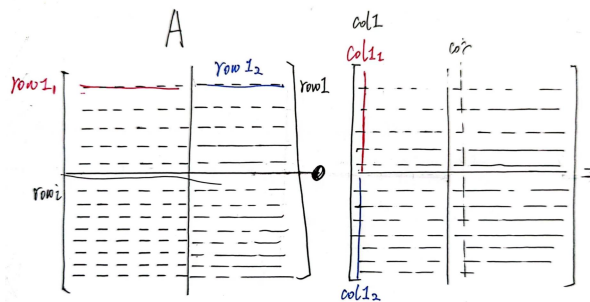
← (1,1) of this matrix

同理，对于结果中的矩阵 (i, j) 的位置，那么就可以得到其结果应当是左侧矩阵中的 rowi 乘右侧矩阵的 colj。

于是，我们这个结论便成立了

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \end{bmatrix} = \begin{bmatrix} A_1 A_2 & A_1 B_2 \\ B_1 A_2 & B_1 B_2 \end{bmatrix}$$

现在我们将结论扩展，考虑将一行拆开，变成 row11, row12 和 col11 和 col12，像这样：



那么对于结果中的矩阵，其 (1,1) 的位置就应当是 $row11 \cdot col11 + row12 \cdot col12$ ，像这样：

$$\begin{bmatrix} row1_1 \cdot col1_1 + row1_2 \cdot col1_2 \\ row1_1 \cdot col2_1 + row1_2 \cdot col2_2 \end{bmatrix}$$

(1,1) of this matrix

而这似乎可以写成这样：

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} = \begin{bmatrix} A_1 B_1 + A_2 B_3 & A_1 B_2 + A_2 B_4 \\ A_3 B_1 + A_4 B_3 & A_3 B_2 + A_4 B_4 \end{bmatrix}$$

这也正是矩阵乘法的第四种算法，即将矩阵拆成了向量。

以上这四种方式需要灵活运用，在不同的场合运用不同的方法。

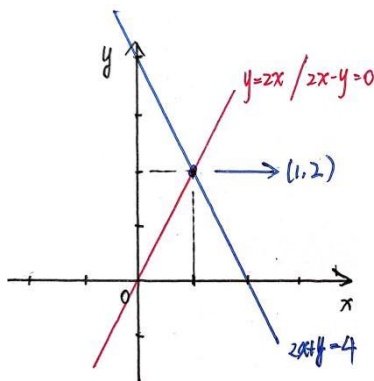
Chap2.1 用向量的眼光看待多元一次方程

先来举个例子，你能在 5 秒钟之内看出下面这个方程的解吗？

$$2x - y = 0$$

$$2x + y = 4$$

或许你可以一眼看出这个方程组的解，但还是从图像中观察一下：

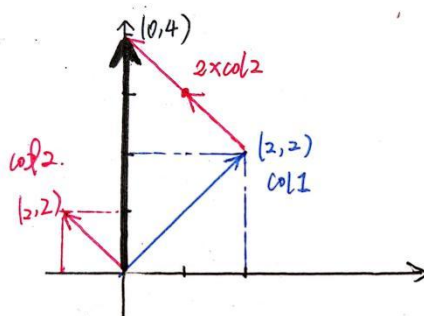


从图像中，我们观察到，解就是 $x=1$, $y=2$ ，这是一种观点，我们换一个视角再来看待这个问题：

我们看将这个方程组看成向量的形式：

$$\begin{aligned} 2x - y &= 0 \\ 2x + y &= 4 \end{aligned} \Leftrightarrow \begin{bmatrix} 2 \\ 2 \end{bmatrix} x + \begin{bmatrix} -1 \\ 1 \end{bmatrix} y = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

所以上述问题就转化为，寻找一个 x , y ，使得向量 $(2, 2)$ 和 $(-1, 1)$ 通过平行四边形的组合，产生 $(0, 4)$ ，我们来再画一个图像，称之为“Col Picture”



或许，并没有从这个例子中看出，这两个图的区别，我们再换一个例子：

$$2x + y + 4z = 2$$

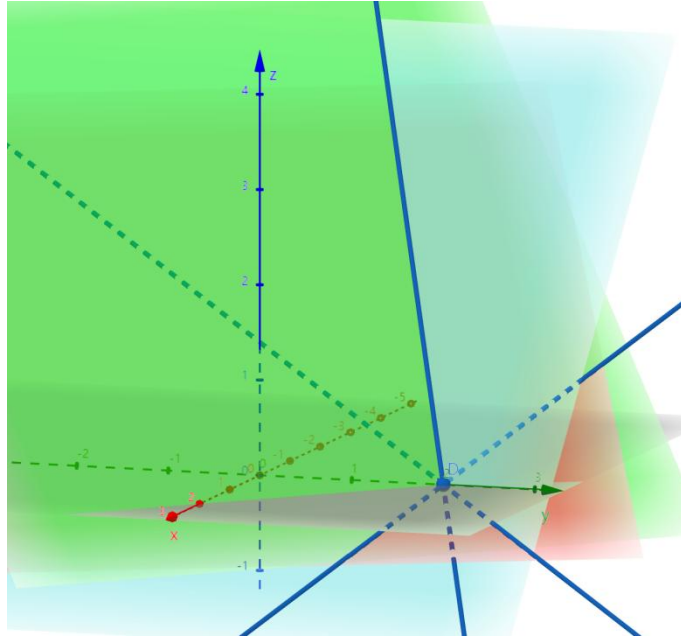
$$x + z = 0$$

$$2x + 2y + 3z = 4$$

你能 5s 之内看出来这个方程组的解吗？

同样地，我们使用刚才的思路，绘制这个方程组的图像：

注意， $2x+2y+3z=4$ ，可以想象，其在空间中是一个平面，从图中，我们看出，三个平面交于 D 点 $(0, 2, 0)$ ，这也正是这个方程的解。

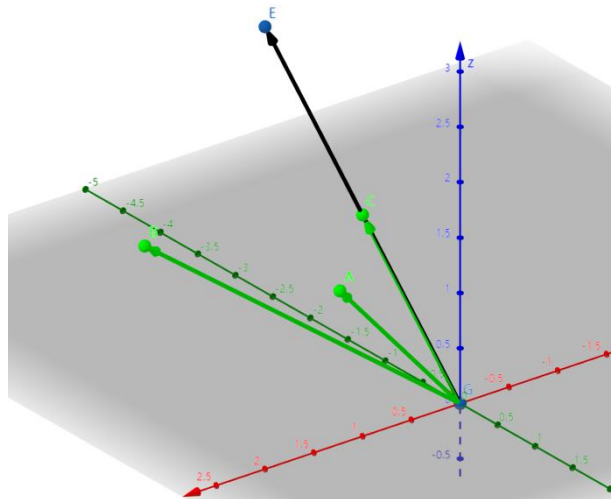


用一下 Col Picture 如何呢？

$$\begin{aligned} 2x + 1y + 4z &= 2 \\ x + z &= 0 \\ 2x + 2y + 3z &= 4 \end{aligned} \Leftrightarrow \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} y + \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} z = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$$

现在，你是否可以一眼看出来这个方程组的解呢？

是的，只需要让 $x=0$, $y=2$, $z=0$ ，因为注意到方程左侧 y 对应的向量是 $(1, 0, 2)$ ，而方程右侧是 $(2, 0, 4)$ ，自然地，让 x 和 z 是 0，让 y 是 2，这样就得到了正确的组合结果。



在 Col Picture 中，A 代表向量 $(2, 1, 2)$ ，B 代表向量 $(4, 1, 3)$ ，C 代表向量 $(1, 0, 2)$ ，而最终的解是这个：

其中黑色的 OE 向量，代表等式右侧的解： $(2, 0, 4)$

而在这个图像中，可以轻松地看出，只需要 $OC \times 2$ 即可得到 OE，而剩下两个 OA，OB 无论怎么组合，也不可能组合出来 OE，这也就得出来了方程组的解就是 $(0, 2, 0)$ 。

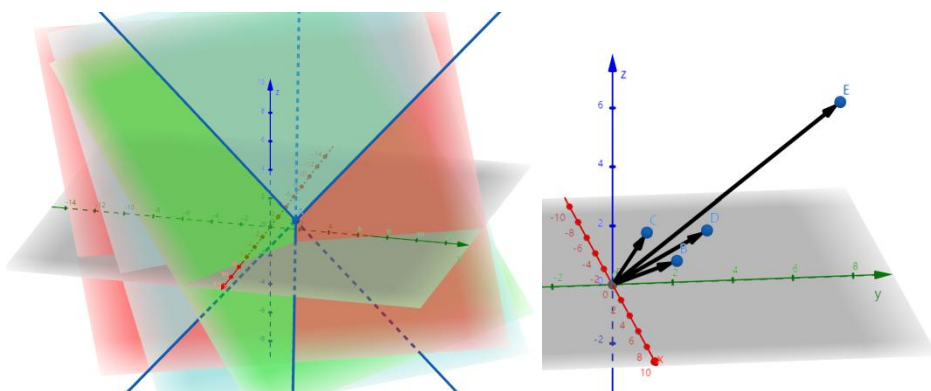
Chap2.2 消元法初步

前面提到用向量的视角审视多元一次方程组，下面讨论如何求解的问题
例如：

$$\begin{aligned}x + y + z &= 4 \\2x + y + 3z &= 7 \\x + 2y + 2z &= 7\end{aligned}$$

如何求解这个方程组的解？

当然我们可以像上一节绘制两种图像，如下



此时我们还没有求出这个解，应当采用消元法来求出，这一想法很自然，因为解不出来的本质就是变量太多，故把变量减少就可以了。

$$\begin{aligned}x + y + z &= 4 & x + y + z &= 4 & x + y + z &= 4 \\2x + y + 3z &= 7 & \Leftrightarrow 0x - y + z &= -1 & \Leftrightarrow 0x - y + z &= -1 \\x + 2y + 2z &= 7 & 0x + y + z &= 3 & 0x + 0y + 2z &= 2\end{aligned}$$

如此，我们由第三个方程，得出来 $z=1$ ，从而往前面的方程代入，得到 $y=2$ ， $x=1$ ，此时如果回到那两个图，也可以看出，左图的交点为 $(1,2,1)$ 。

但是，并不是所有的方程组都可以通过“消元法”得出解。

首先，我们来讨论“有没有解”的问题

我们先不研究三元一次方程组，而是研究二元一次方程组

$$\begin{aligned}2x - y &= 0 \\2x + y &= 4\end{aligned}$$

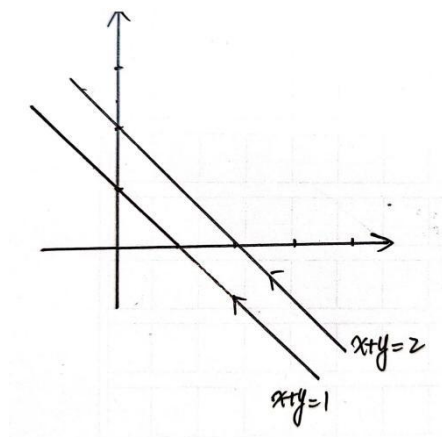
这个方程组是有解的，因为从集合上，可以理解两条直线的交点，或者两个向量的组合。但是，两条直线一定就会有一个交点吗？两个向量一定可以组合出另一个向量吗？

并不是，因为存在直线平行或者共线的情况。

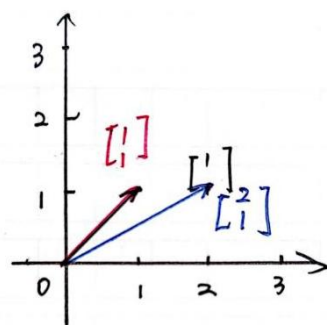
(1) 直线平行

$$\begin{aligned}x + y &= 2 \\x + y &= 1\end{aligned}$$

如果从直线的视角，我们可以看出：



直线平行，无交点，故不可能有解，如果从向量的视角，

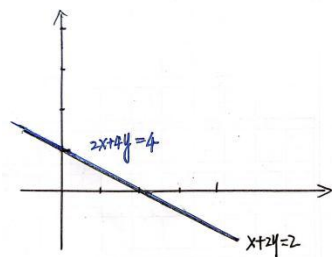


两个 $(1,1)$ 向量不可能通过平行四边形法则产生 $(2,1)$ ，故同样没有解。
(2) 直线共线

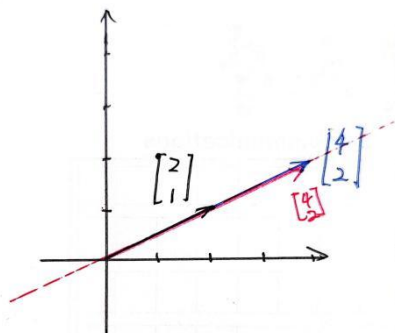
$$2x + 4y = 4$$

$$x + 2y = 2$$

从直线的角度，此时直线重合，有无数多个解



从向量的角度：



有无数种方法，使得向量（2,1）和（4,2）组合成（4,2），所以也是无数组解。
所以，从上面的讨论得到，

解的三种情况：一个解，无解，无穷多个解

现在我们将视野推展的三维，

三维中，一个多元一次方程代表的是一个平面。

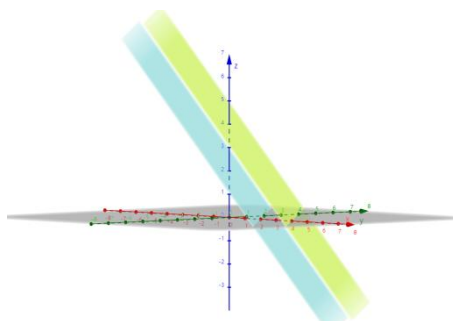
（1）平面平行

$$x + y + z = 2$$

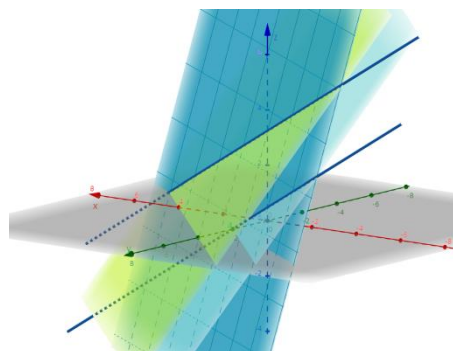
$$x + y + z = 4$$

$$2x + 3y + z = 6$$

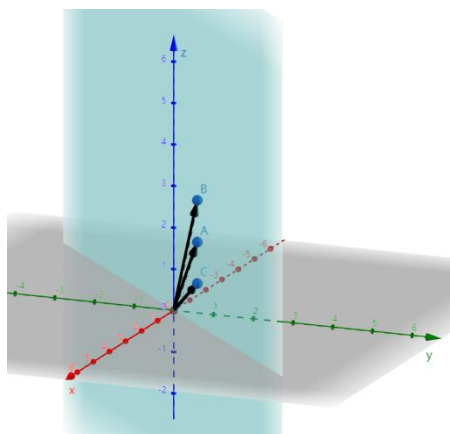
把前两个方程代表的平面画出来，可以得到：



此时，无论第三个平面如何放置，都不可能存在某一点同时位于三个平面上。



可以看出，第三个曲面与上述平行的两个曲线交出两条平行的直线，不可能存在交点。
如果从向量的角度：



这三个向量是共面的，故不可能产生向量（2,4,6）。

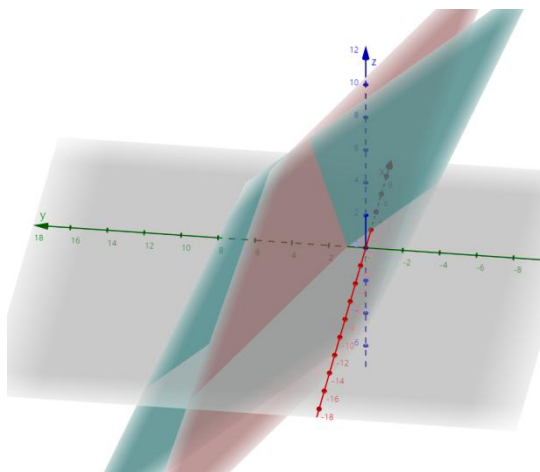
(2) 平面重合

$$x + 2y + z = 2$$

$$2x + 4y + 2z = 4$$

$$x + 3y + z = 3$$

同样地，我们还是先把前两个平面画出来：

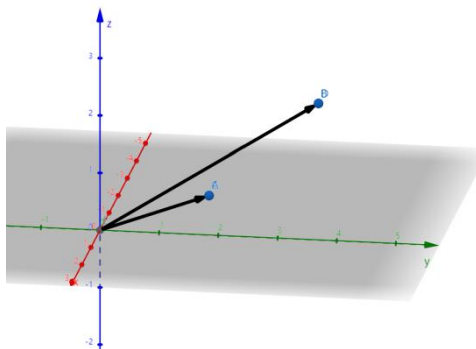


此时两个平面重合，故第三个平面有两种情况，

第一种情况是第三个平面重合或者平行于前两个平面，这时的效果和前面讨论的一致

第二种情况就是不重合且不平行，此时，第三个平面就会与前两个平面交出来一条直线，此时将会有无穷多个解。

而从向量角度，有两个向量是相同的，故等式左侧的三个向量形成一个平面，而恰好，右侧的向量又位于这个平面之上，故此时便有了无穷多组解。



上面，我们通过考虑平面的交线，交点，向量的重合，共面，来讨论解的存在性问题。

下面我们从代数视角说明为什么会存在“无解”，“无穷多组解”的情况

(1) 有解的情况：

$$\begin{array}{lll} x + y + z = 4 & x + y + z = 4 & x + y + z = 4 \\ 2x + y + 3z = 7 & \Leftrightarrow 0x - y + z = -1 & \Leftrightarrow 0x - y + z = -1 \\ x + 2y + 2z = 7 & 0x + y + z = 3 & 0x + 0y + 2z = 2 \end{array}$$

此时先求解 z ，然后通过回代求解 y , z

观察我们的消元法，每一次只是对 x, y, z 系数的运算，以及对等式右侧的运算，故我们

把这些数字都拿出来即可，不需要每一次都写 x, y, z ，于是我们引入了“矩阵”。
这就是矩阵的形式，本质就是方程组的系数罢了。

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 2 & 1 & 3 & 7 \\ 1 & 2 & 2 & 7 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & -1 & 1 & -1 \\ 0 & 1 & 1 & 3 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 2 & 2 \end{pmatrix}$$

(2) 无解的情况

$$\begin{aligned} x+y+z &= 2 & x+y+z &= 2 \\ x+y+z &= 4 & \Leftrightarrow & 0x+0y+0z=2 \\ 2x+3y+z &= 6 & 0x+y-z &= 2 \end{aligned}$$

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 4 \\ 2 & 3 & 1 & 6 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & -1 & 2 \end{pmatrix}$$

我们观察到了一个结构： $0x+0y+0z=2$ ，而 0 乘任何数都应当是 0，故不可能有解
于是我们得到一个不严谨的结论

通过换元法产生 $0x+0y+0z=c$ ($c \neq 0$) 结构，则无解

(3) 无穷多组解的情况

$$\begin{aligned} x+2y+z &= 2 & x+2y+z &= 2 \\ 2x+4y+2z &= 4 & \Leftrightarrow & 0x+0y+0z=0 \\ x+3y+z &= 3 & 0x+y+0z &= 1 \end{aligned}$$

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 4 & 2 & 4 \\ 1 & 3 & 1 & 3 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

我们观察到一个结构， $0x+0y+0z=0$ ，此时无论什么 x, y, z 都满足这个方程。接下来我们开始一般化的讨论，即从特殊到一般。

对于一个任意的三元一次方程，我们把它写作：

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{pmatrix}$$

如果在通过换元法后，产生下面的结构（◆代表了非 0 的数字，*代表任何数字）

(1) 那么则有唯一解：

像是一个台阶的结构，这样的矩阵称为阶梯型矩阵。

$$\begin{pmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{pmatrix}$$

(2) 那么无解:

$$\begin{pmatrix} \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & * \end{pmatrix}$$

(3) 那么有无穷多组解:

$$\begin{pmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ 或 } \begin{pmatrix} \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

这一结论可以拓展到 n 维的矩阵，和上面都是一样的。
请记住:

向量的眼光看待线性方程组

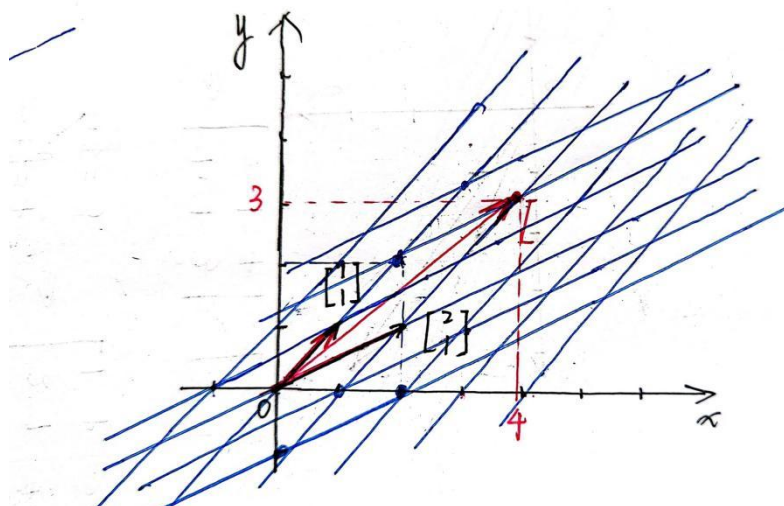
Chap2.3 消元法与矩阵的关系

前面，我们介绍了矩阵的乘法，介绍了矩阵的消元法，接下来我们来建立二者之间的关联。

首先，举个例子：

$$\begin{cases} 2x + y = 4 \\ x + y = 3 \end{cases} \Leftrightarrow \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

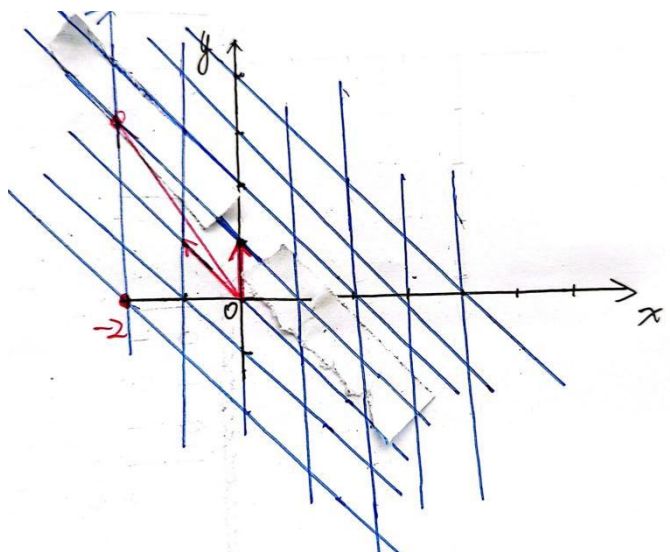
如果从几何上看，将会是这样：



接着，我们利用消元法的规则对其进行化简，于是有：

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

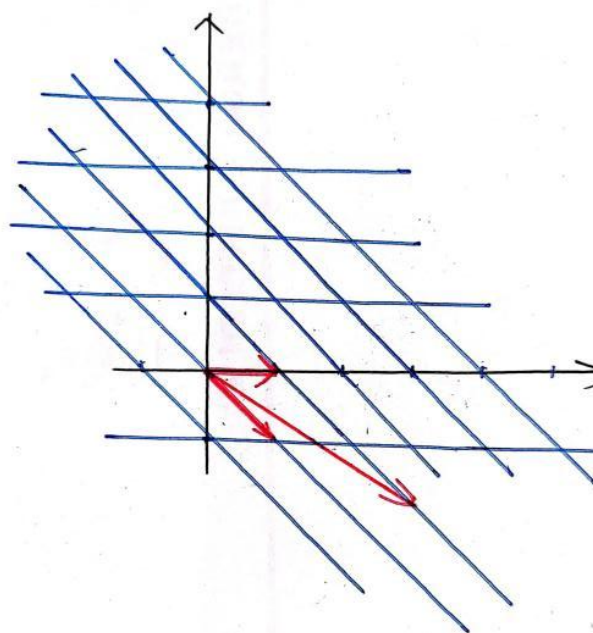
如果从几何上看，线性空间向着 y 轴被推了一下：



接着，我们觉得 0 在有些蠢，于是做一个行变换，像这样：

$$\begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

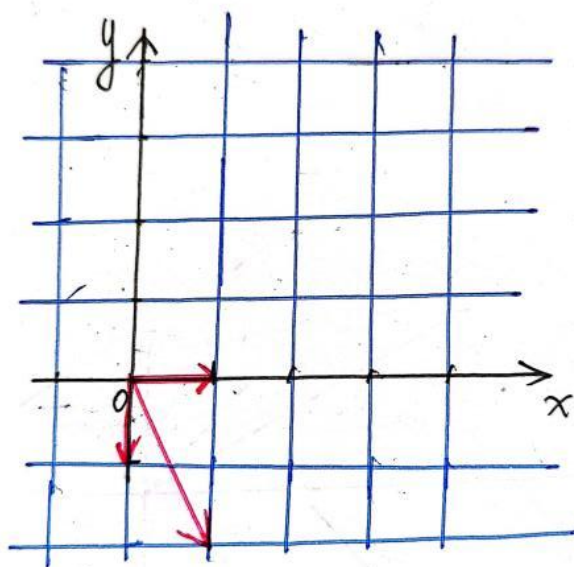
如果从几何上看，像是把整个空间关于 $y=x$ 翻转了一下，这是合理的，因为基向量的横坐标和纵坐标翻转了，这也导致了整个线性空间的横坐标和纵坐标翻转，体现为关于 45° 的 $y=x$ 对称，像这样：



下面，我们再进行一步，于是得到：

$$\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

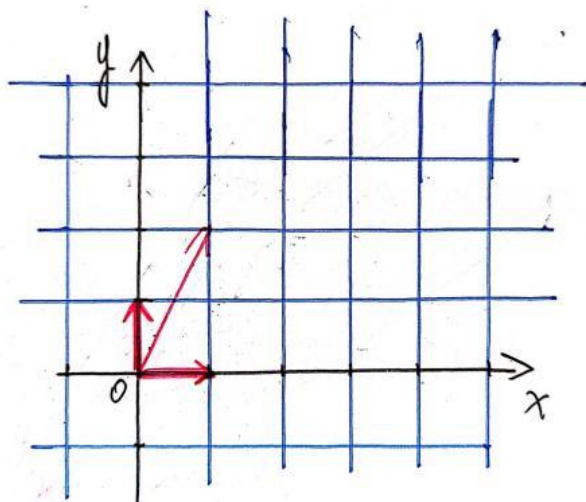
如果从几何上看，这将会变得异常舒适，像这样：



最后一步：

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

在几何上，就变成了这样舒适的线性空间：



观察上面“线性空间”的变换，我们可以联想到，线性空间的变换是与矩阵有关系的，矩阵正是告诉线性空间如何变换的工具。

例如在这一步中

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

能否找到一个矩阵，实现下面的变换呢？

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} ; \quad \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

我们先从简单的角度说起，如果需要一个矩阵保持不变，那最好给它乘的矩阵就是这个：

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

我们来分析一下这个矩阵的含义：

如果用前面“矩阵乘法的五种算法”中“行向量”的观点来看的话，这就会变得异常清晰，由于电脑 word 文档不容易编辑，故采用手写描述。

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1[2 \ 1] \\ + \\ 0[1 \ 1] \\ 0[2 \ 1] \\ + \\ 1[1 \ 1] \end{bmatrix}$$

这代表着，输出的矩阵的第一行为 1 个 (2,1) + 0 个 (1,1)；而输出矩阵的第二行为 0 个 (2,1) 和 1 个 (1,1)，如果有这个观点，那么假若我们需要方程相减的操作，便有：

$$\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1[2 \ 1] \\ + \\ -2[1 \ 1] \\ 0[2 \ 1] \\ + \\ 1[1 \ 1] \end{bmatrix}$$

同理，可以获得其他对应的矩阵。

所以例如上面的每一步操作，我们都可以寻找到一个矩阵，如此，便可以简洁地写出整个消元的过程，像这样：

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

请记住：

可以通过矩阵，来描述线性方程组消元的方式

Chap2.4 逆的本质

前面，我们研究了矩阵的消元法与矩阵的关联，像这样：

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

但实际上这是不太理想的情况，因为涉及到了行交换（线性方程的交换）。我们如果定义这样一个 3×3 的线性方程组，举个例子，这样的矩阵：

$$A = \begin{bmatrix} 1 & 2 & 1 & 6 \\ 1 & 1 & 2 & 5 \\ 2 & 1 & 1 & 5 \end{bmatrix}$$

如果我们将它分解，将会变成这样：

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

其中最后化简完成的矩阵 $\begin{bmatrix} 1 & 2 & 1 & 6 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -4 & -4 \end{bmatrix}$ ，被称为 U 。

我们采用记号 $E_{(2,1)}$ ，代表一个矩阵，使得其作用等效为初等行变换使得第二行，第一列的元素为 0。那么以此类推，刚刚的 3×3 的矩阵的每一个 E ，将会是这样：

$$E_{(2,1)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{(3,1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \quad E_{(3,2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

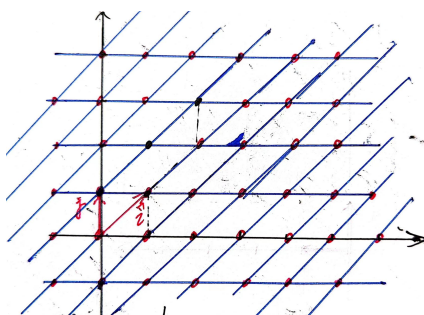
如果我们简化地写，将会是这样：

$$E_{(3,1)} E_{(3,2)} E_{(2,1)} A = U$$

下面，我们考虑矩阵的另一个概念——逆，这个概念需要先引出，才能引出下一节的 LU 分解。

前面我们提到，矩阵是操纵线性空间的工具，这也就意味着，一个矩阵可以伸缩和压缩

一个空间。比如，通过这样一个矩阵 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ ，可以将空间向上拉伸一些，像这样：



那么我们有没有一个矩阵，可以让空间“压缩回去”呢？

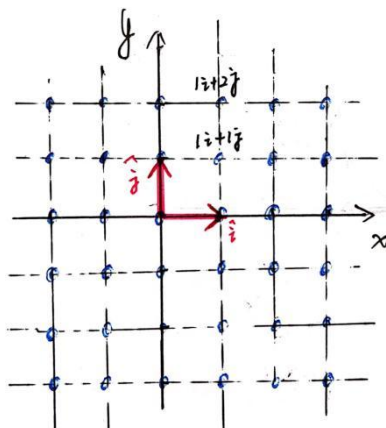
是存在的，这个矩阵就是 $A^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ ，这个矩阵就成为 A 的“逆”矩阵。我们来验

证一下，如果让这个线性空间先进行 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ 变换，再进行 $A^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ 变换，那么

理论上，整个线性空间应该像没变化一样。我们用矩阵的乘法来验证我们的直观认识：

$$A^{-1}A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

可见，我们得到了基向量 $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 的平直的网格线性空间了。

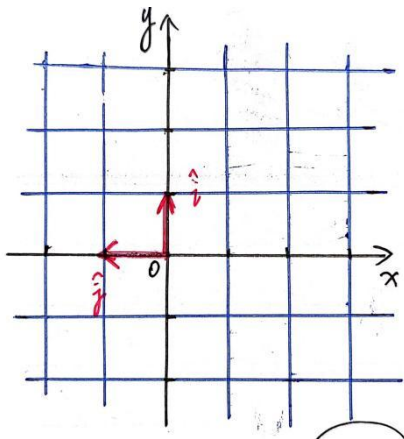


我们称这样的矩阵为单位矩阵，记作 $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ，这在后面会经常用到。

或许，你對自己还不能独立求解矩阵的“逆”而感到悲哀，下面是两个可以一眼看出逆的矩阵。

第一个例子，是 $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ，首先，我们从几何上，即线性空间的角度认识这个矩阵

的作用，也就是将整个空间逆时针旋转了 90° 。像这样：



所以它的逆就代表着顺时针旋转 90° ，是的，我们可以闭着眼写出来顺时针旋转 90°

的矩阵 $\mathbf{A}^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ 。我们来检验一下：

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

所以，检验成功。

第二个例子，是 $E_{(2,1)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ，这意味着将线性方程组的第二行减第一行，自然

地，它的逆就是第二行加上第一行，于是自然地写出 $E_{(2,1)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 。我们再来检查一

下。

$$E_{(2,1)}E_{(2,1)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

所以，对应地，我们可以写出最开始例子的 \mathbf{E} 矩阵的所有逆矩阵。

$$\begin{aligned} E_{(2,1)} &= \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & E_{(3,1)} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} & E_{(3,2)} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} \\ E_{(2,1)}^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & E_{(2,1)}^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} & E_{(3,2)}^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} \end{aligned}$$

在直观理解了矩阵的逆的含义之后，我们给出矩阵的逆的定义：

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

那么就称， \mathbf{A}^{-1} 为 \mathbf{A} 的逆矩阵。

下面我们讨论求解逆矩阵的思路。首先，举个例子，就举那个一开始让你沮丧的例子吧，因为逆没有一眼将这个矩阵的逆看出来：

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

如果求解这个矩阵的逆，回归定义，也就意味着要去找一个 \mathbf{A}^{-1} ，满足

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}。$$

于是，我们不妨先把这个 A^{-1} 给设出来，像这样：

$$A^{-1} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

于是

求解 A^{-1} 的方程将会是：

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

其实，翻译过来，就是这样的一个方程组：

$$\begin{aligned} a &= 1 \\ c &= 0 \\ a + b &= 0 \\ c + d &= 1 \end{aligned}$$

我们可以通过普通的消元法求解出来对应地解，于是可以求解出：

$$A^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

我们发现，这实际上等价于在求解两个线性方程组，像这样：

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

接着，每一个方程进行矩阵的消元法进行消元即可，各自对应的增广矩阵是这样的：

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

但是，似乎有些麻烦。为何不把他写在一起呢？消元的方式都是一样的呀，像这样：

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

这是高斯-若尔当方法，很自然不是吗？如果我们早出生一些或许也能发现。

如果我们将这个矩阵继续向下化简，那么将会得到：

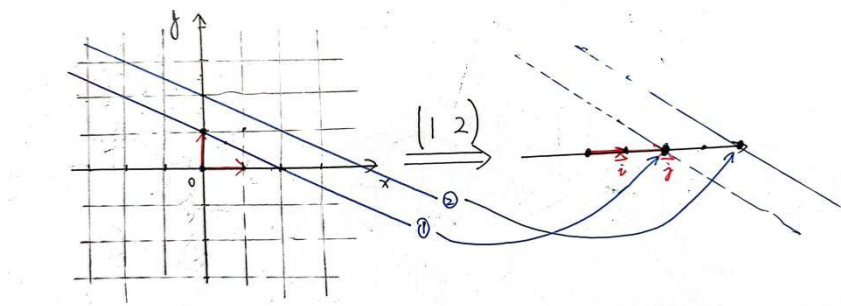
$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

其中右侧的部分就是我们的逆。

Chap2.5 可逆与不可逆的本质

前面，我们提到了矩阵的逆如何计算，但下面的问题恐怕会让你头疼。

在前几个章节中，我们提到了这样的一个变换，将空间压缩到了更低的维度上，像这样：



由于 $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x+2y]$ ，所以我们只需要找出来 $x+2y$ 为定值的直线，就可以找到

对应右侧一维空间上的向量。

但是，能不能有“升维”的矩阵，将这条直线在展开成一个二维的线性空间？首先我们直观地认识一下，如果有逆的话，也就意味着这个逆满足以下这个式子：

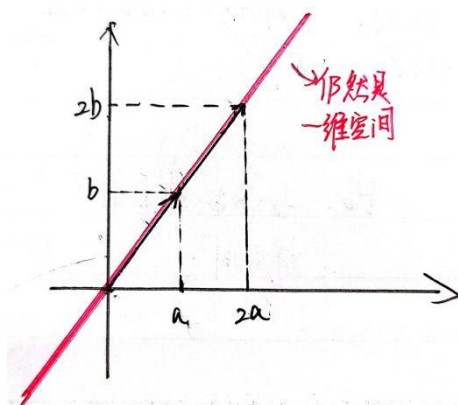
$$A \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

这显然是不合理的，如果要产生单位向量，意味着逆矩阵 A 应当是一个有两行的矩阵。

而由于 $\begin{bmatrix} 1 & 2 \end{bmatrix}$ 只有一行，这也就意味着 A 只能有一列。于是我们可以完善一下上面的表达式，像这样：

$$\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & 2a \\ b & 2b \end{bmatrix}$$

如果我们尝试几何中表示经过这个矩阵变换后的图形，将会是这样：



这仍然是一个一维空间，换句话说，似乎不存在一个矩阵，让一个一维空间的直线展开成为一个二维空间。

这是我们的初步理解，也就是说这个矩阵不存在逆。

我们接着给出进一步深入一些的理解。

首先，我们如何理解矩阵乘法？代表着对“基向量”的改变。比如说我现在有一个共线

的基向量所组成的矩阵，像这样：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

我们是否能找到它的逆呢？

假设存在逆，他将满足这个表达式：

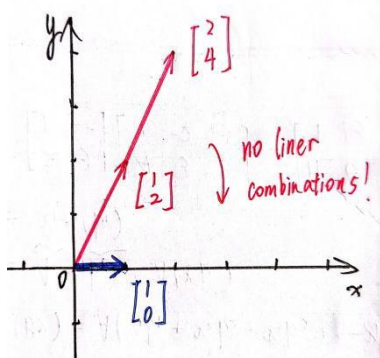
$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

如何看待这个式子？我们回归矩阵乘法的本质。

这意味着先输入一个矩阵 $\begin{bmatrix} a \\ b \end{bmatrix}$ ，接着，用 a 来作为 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 的权，而使用 b 来作为 $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ 的权。

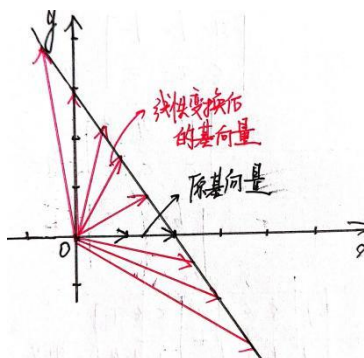
而输出的向量应当是 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ，这当然不可能， $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ 怎么可能线性组合出 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 呢？No way！

在图像中，将会是这样：



下面我们考虑另一种理解，回归什么是“矩阵”？矩阵只不过是操纵线性空间的工具，实现线性空间的线性变换。但是请注意，什么是“线性变换”？

线性变换，指的是输入一个向量，并输出一个向量。那么我们是否存在一个“由线到面”的线性变换呢，这必然是不可能的，因为它将输入一个向量，而将输出无数个向量，如果在几何中，将会是这样：



这也正是第三种理解。

Chap2.6 矩阵的 LU 分解初步

前面，我们研究了矩阵的消元法与矩阵的关联，像这样：

$$\text{对于 } A = \begin{bmatrix} 1 & 2 & 1 & 6 \\ 1 & 1 & 2 & 5 \\ 2 & 1 & 1 & 5 \end{bmatrix}$$

如果我们将它消元，将会变成这样：

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

其中最后化简完成的系数矩阵 $\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -4 \end{bmatrix}$ ，被称为 U 。

我们采用记号 $E_{(2,1)}$ ，代表一个矩阵，使得其作用等效为初等行变换使得第二行，第一列的元素为 0。那么以此类推，刚刚的 3×3 的矩阵的每一个 E ，将会是这样：

$$E_{(2,1)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{(3,1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \quad E_{(3,2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

如果我们简化地写，将会是这样：

$$E_{(3,2)} E_{(3,1)} E_{(2,1)} A = U$$

显然，没有人会这样计算矩阵。但是我们还是来举个例子，我们将 $E_{(3,1)} E_{(2,1)}$ 和相乘，像这样：

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

我们发现，这个矩阵的等效矩阵就是对第三行先减去两倍的第一个方程，再减去三倍的第二个方程。也就是说，是两个矩阵效果的叠加。

为了更有效的计算矩阵，我们给出下面的一个引理，方便“三角形的计算”

对于一个有上三角的矩阵 \times 上三角的矩阵，输出的矩阵仍然是一个上三角，像这样：

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

同理，下三角的矩阵 \times 下三角的矩阵，输出的矩阵仍然是一个下三角，像这样：

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

我们由此继续计算完上面表达式，这下可以一下子写出右上角是一个三角。像这样：

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

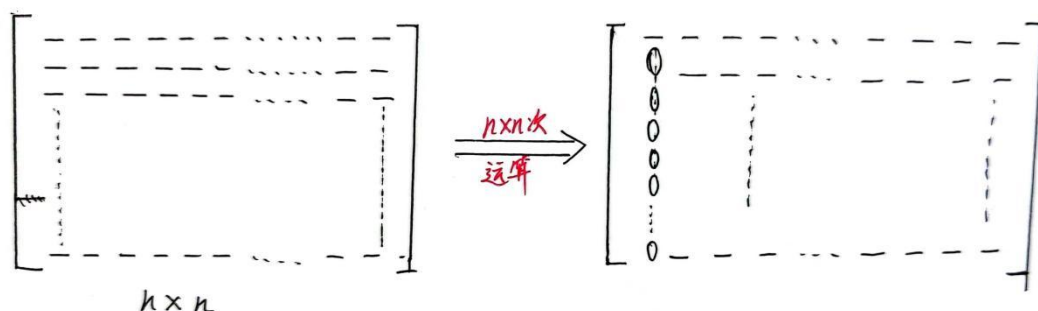
我们继续分析这个矩阵的含义，位于第二行，第一列的-1，表示第二行减去第一行，可以理解。位于第三行第二列的-3，代表第三行减去三倍的第二行，可以理解。但是这“1”是怎么来的？第三行只减去了2倍的第一行呀？

实际上，问题在于第三行总共加上了-2个第一行，并且加上了-3个第二行，但是第二行也曾经加上了-1个第一行。“所以第三行，加上了-3个第二行”等效成了又加上了3个第一行，于是最终等效为加了1个第一行。

但是，这有什么作用呢？

如果我们要让电脑来求解一个线性方程组，那么这也就要求，要寻找一种尽可能快地算法来解除线性方程组。

例如，对于一个n个未知数，n个方程的线性方程组，计算机计算多少次呢？像这样：



如果先对第一列进行消元，那么首先计算机将会根据每一行的不同情况，先对第一行乘一个系数，接着与这一行做加法，从而消去未知量。其中，乘法，由于一共有n个元素，所以乘了n次，而减法，由于有n个元素，减了n次，所以一次操作，如果我们将一次加法，一次乘法看作一次运算的话，那么消除一个元就需要n次运算，而由于一共有n-1列，故一个需要 $n * (n-1)$ 次运算。

同理 $(n-1) * (n-2)$ ， $(n-2) * (n-3)$ ，.....

我们将 $n * (n-1)$ 近似的认为是n的平方，因为这样比较方便估算，于是总的运算次数将会是：

$$\sum_{i=1}^n i^2 = 1 + 2^2 + 3^2 + \dots + n^2 \approx \frac{n^3}{3}$$

这也就意味着，随着线性方程组中未知量的个数增加，运算的次数将会随着未知量的个数的立方增长。

我们的目的就是求解出x，像这样：

$$Ax = b$$

有没有更加优化的算法呢？其实上，如果没有“其他条件”，我们或许没有更优化的算法。但是，如果对于固定的A而有多组b需要求解x，此时高斯消元便略显吃力

比如说，我们现在有 m 个 b 需要求解 x ，那么如果用普通的高斯消元法，对于计算机来说，将会计算 $m \frac{n^3}{3}$ 次，如果 m 和 n 都很大，这相当于一个接近于四次方的增长。

我们发现，许多计算都在进行重复的操作，即对 A 的行变换，这个行变换当然不需要这么多次。我们只需要进行一次高斯消元法，找到“消元”的方法，然后把这个方法告诉 b 即可，让 b 自己消元产生 A 即可。

如何把这个“消元的方法”正确地告诉 b 呢？

前面，我们讨论了矩阵与消元的关系，不妨把式子写成这样：

$$E_{(3,2)}E_{(3,1)}E_{(2,1)}A=U$$

或者，简化一点，写成这样：

$$EA=U$$

而我们的目的就是求解 x ，于是我们只需要去计算这件事即可，这就是我们前面讨论的“逆”运算：

$$EAx=Ux=b \Leftrightarrow Ax=E^{-1}Ux=b$$

但是，如何计算出逆呢？首先对于 $E_{(3,2)}E_{(3,1)}E_{(2,1)}$ ，它的逆应该是什么？这又回到了我们说的“矩阵的本质”，其本质就在于对于线性空间的变换。我们可以将每一个矩阵 $E_{(3,2)}$ ， $E_{(3,1)}$ ， $E_{(2,1)}$ ，看作是对线性空间的一个变换，而他们的逆矩阵，就是将这个线性空间的变换逆过来。

举个例子，我们如果定义矩阵 A 的含义是将空间逆时针旋转 90° ，矩阵 B 的含义是将空间向下压缩。那么 AB 的含义，将会是先向下压缩，再逆时针旋转 90° 。很自然地，如果将它逆回去，应该是先顺时针旋转 90° ，再向上拉伸，这也是 AB 的逆的含义。如果写成数学表达式将会是这样：

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

是的，同理，三个变换也是一个道理，像这样

$$E_{(2,1)}E_{(3,1)}E_{(3,2)}=E_{(3,2)}^{-1}E_{(3,1)}^{-1}E_{(2,1)}^{-1}$$

如果比喻成生活中的道理的话，将会是，先穿上袜子（矩阵 A ），再穿上鞋（矩阵 B ），再系上鞋带（矩阵 C ）。那么将这个逆过来，将会是先解开鞋带（ C 的逆），在脱下鞋（ B 的逆），在脱下袜子（ A 的逆）。只不过这里把穿衣服比喻成了对线性空间的变化罢了。

于是，我们将会下面这个式子：

$$Ax=E_{(3,2)}^{-1}E_{(3,1)}^{-1}E_{(2,1)}^{-1}Ux=b$$

是的，下面，我们将 $E_{(3,2)}^{-1}E_{(3,1)}^{-1}E_{(2,1)}^{-1}$ 记作 L ，于是上面的式子将写作：

$$A=LU \rightarrow Ax=LUx=b$$

这为什么简化了计算？我们可以这么看：

$$Ax=LUx=L(Ux)=Ly=b$$

换句话说，首先通过 $Ly=b$ ，求解出 y ，然后再由 $Ux=y$ ，求解出 x 的值，而这能简化

算法的本质，就是因为 L 和 U 都是上三角和下三角形状的，这意味着有一半的趋于都不用算。

或许这有些抽象，无所谓，我们从上面的例子来看。

$$\text{对于 } A = \begin{bmatrix} 1 & 2 & 1 & 6 \\ 1 & 1 & 2 & 5 \\ 2 & 1 & 1 & 5 \end{bmatrix}$$

如果我们将它消元，将会变成这样：

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

我们采用记号 $E_{(2,1)}$ ，代表一个矩阵，使得其作用等效为初等行变换使得第二行，第一列的元素为 0。那么以此类推，刚刚的 3×3 的矩阵的每一个 E ，将会是这样：

$$E_{(2,1)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{(3,1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \quad E_{(3,2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

首先，我们来求解对应的逆，像这样：

$$E_{(2,1)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_{(3,1)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \quad E_{(3,2)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

然后你是否可以一眼看出来它的逆的乘积呢？

我一眼就看出来了：

$$E_{(2,1)}^{-1} E_{(3,1)}^{-1} E_{(3,2)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

是的，这里为什么没有前面的冲突呢？显然，因为 $E_{(3,2)}$ 意味着先加 3 个第二行， $E_{(3,1)}$ 意味着先加 2 个第一行， $E_{(2,1)}$ 意味着加一个第一行，不存在任何的冲突！这也就是为什么可以直接将高斯消元法的操作直接重叠在这个矩阵之上。

如果你不放心，当然可以再计算一下，像这样：

$$E_{(2,1)}^{-1} E_{(3,1)}^{-1} E_{(3,2)}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

是的，这是正确的。于是，我们将 LU 分解的算法写出来，就是这样：

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

我们把 LU 算法分成了两步：

(1) 通过 $Ly=b$, 求解出 y

观察一下, y 是一个向量, 输入一个矩阵 L , 而这个矩阵 L 的右上都是 0 或者 1 组成的三角, 所以这部分不需要任何的计算。

(2) 通过 $Ux=y$, 求解出 x

观察一下, x 是一个向量, 输入一个矩阵 U , 而这个矩阵 U 的左下都是 0 或者 1 组成的三角, 所以这部分不需要任何的计算。

这就是为什么这个算法可以再 A 固定时简化计算的本质, 因为有接近一半都没有计算。

对于一个 $n \times n$ 的矩阵来说, 像这样:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ * & 1 & 0 & & \\ * & * & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & * & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & * & * & \dots & * \\ 0 & 1 & * & & \\ 0 & 0 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} x = b$$

下面我们具体分析这个算法究竟进行了多少次计算, 像这样:

Step 1

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ * & 1 & 0 & & 0 \\ * & * & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & * & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} \Rightarrow \begin{array}{l} y_1 = b_1 \\ \downarrow \\ \text{回代 (乘法+减法) 1次} \\ y_2 = b_2 - \lambda y_1 \\ \downarrow \\ \text{回代 (乘法+减法) 2次} \\ y_3 = b_3 - \lambda_2 y_2 - \lambda_1 y_1 \end{array}$$

所以总的计算次数将会是:

$$\sum_{i=1}^n i = 1 + 2 + 3 + \dots + n \approx \frac{n^2}{2}$$

是的同理, $Ux=y$ 的计算, 也需要相同的次数, 所以从共需要 n^2 次计算。这之前的高斯消元法相比, 减少了一些复杂度, 从 n 的三次方, 变成 n 的平方, 简化了算法。

Chap2.7 转置的本质

前面，我们的 LU 分解，是建立在不需要“行交换”的基础上进行的。但实际上，很多的矩阵都需要行变换，比如说我们在 Chap2.5 中就遇到了行交换的情况，像这样：

$$\begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

是的，如果我们要实现这样的一个矩阵的变换，我们需要乘上这样一个矩阵：

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

对于二维的话，行的变换就有两种，一种是不变换，一种是第一行和第二行变换，也就意味着只有两个行变换的矩阵，像这样：

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

对于三维的话，行的变换就有六种，相当于就是行的“有顺序的排列”一共有六种，分别是 (1,2,3)，(1,3,2)，(2,1,3)，(2,3,1)，(3,2,1)，(3,1,2)，分别对应的矩阵长成这样：

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

如果是 $n \times n$ 的矩阵，那么应当有 $n!$ 种行变换的矩阵。

下面我们考虑这些矩阵的性质，先从简单的开始，对于矩阵 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，那么它代表着将第一行和第二行调换过来这样一种变换。那么它逆过来的变换将会是什么呢？显然，仍然是将第一行和第二行调换一下，这样就像是好像没有进行任何操作一样。于是我们得到：

$$A^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = A$$

这是否是巧合呢？我们再找一个例子：

这代表着将第二行和第三行交换。自然地它的逆就是再将第二行和第三行交换，就是它本身，于是我们再次得到：

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = A$$

但是，如果我们观察这样一个矩阵：

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

它代表的含义是将第三行放到第一行，将第一行放到第二行，将第二行放到第三行。它的逆显然就不是它本身了，再进行相同的操作会使得第二行被放到第一行，并没有达到效果。

于是，为了引出这背后的规律，我们引入一个概念——矩阵的转置。

顾名思义，之前我们矩阵中的向量都是列向量，现在我们希望它变成行向量，这也就是矩阵的内涵，例如：

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

它的列向量分别是：

$$i = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, j = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, k = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

将这些列向量写成行向量也就是：

$$\begin{aligned} i &= [0 \quad 1 \quad 0] \\ j &= [0 \quad 0 \quad 1] \\ k &= [1 \quad 0 \quad 0] \end{aligned}$$

将它们拼在一起，也就得到了矩阵的转置，我们把它记作 A^T 。

$$A^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

下面我们来研究它的性质。

这真是一个天才的想法，你看到了吗？对于原来的三个基向量， i, j, k 都是相互垂直的，而矩阵乘法是行乘列，这也就意味着，如果我将一个矩阵转置过来，将会使得 i, j, k 与 i, j, k 点乘，而自身与自身点乘由于是单位向量，将会得到 1 ，但是如果是与其他向量点乘，由于相互垂直，将会得到 0 ，这不正是单位矩阵吗？

于是，我们尝试着将 $A^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ 与 $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ ，像这样：

$$AA^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

于是我们得到了这样一件事，对于这种“实现行变换的矩阵”，有这样一个性质，即：

$$A^{-1} = A^T$$

这就是正交矩阵转置的一个十分良好的性质。

还有一类矩阵有关转置有良好的性质，这一类矩阵被称为“对称矩阵”，像这样：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 10 \\ 3 & 10 & 9 \end{bmatrix}$$

如果将它转置，它的行依旧是行，列依旧是列。因为在原来的矩阵当中，第一行和第一列相同，第二行与第二列相同，第三行与第三列相同。所以转置过来道理是一样的，于是，我们有：

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 10 \\ 3 & 10 & 9 \end{bmatrix} = A$$

对称矩阵，可以通过一个向量 x ，以及其转置过来的矩阵 x^T 相乘得到，比如说：

$$x = \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix}$$

那么它与自身的转置相乘，将会得到：

$$x^T x = \begin{bmatrix} 2 & 3 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 14 \\ 6 & 9 & 21 \\ 14 & 21 & 49 \end{bmatrix}$$

下面，我们考虑矩阵的逆在矩阵乘法时的运算法则。对于一个矩阵 AB ，它的逆是什么？

我们先来考虑简单的情况，即 (Ax) 的逆，因为矩阵 B 每一行的本质就是一大堆向量，把每一个向量输入 A ，即为矩阵乘法的本质，所以自然地，我们先研究 Ax 的逆，进而延伸到 AB 。

$$Ax = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix}$$

首先， Ax 的含义，代表着以 x_1, x_2, \dots, x_n 为权，对 A 中的基向量进行线性组合，注意这里是“列向量”的线性组合，最终会产生一个矩阵 $n \times 1$ 。

那么，如果将这个矩阵转置过来，将会是 $1 \times n$ 的一个矩阵。每一个元素仍然是“线性组合”而来的。所以自然地，我们将 A 翻转过来，然后将 x 也翻转过来，将 x 放在前面，从而对 A 的各个行进行同样的线性组合，这和对列的线性组合是一样的，即：

$$(Ax)^T = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} A_1^T \\ A_2^T \\ A_3^T \\ \dots \\ A_n^T \end{bmatrix} = \begin{bmatrix} x_1 A_1^T \\ + \\ x_2 A_2^T \\ + \\ \dots \\ + \\ x_n A_n^T \end{bmatrix}$$

也就是说：

$$(Ax)^T = x^T A^T$$

那么对于 AB 呢？它的转置是什么呢，我们依旧从矩阵乘法的本质来分析。这意味着 B 中含有很多的列向量，这些列向量与 A 中的行向量点乘，进而获得输出矩阵在某一个坐标的位置，这也就意味着，取决于 A 的行 i 和 B 的列 j ，决定了 AB 的 (i, j) 。那么如果转置了矩阵之后， (i, j) 就变换到了 (j, i) ，但是此时我们仍然需要 B 的 j 列与 A 的 i 行进行点乘，并且把结果放在 (j, i) 上，于是我们考虑将 B 转置放在前面， A 转置放在后面，如此，也就得到了 B 的 j 行与 A 的 i 列点乘，从而将结果放到了 (i, j) 上。

于是我们得到了下面这个式子：

$$(AB)^T = B^T A^T$$

如果我们换一种理解，顺应刚刚的推论： $(Ax)^T = x^T A^T$ 。我们可以将 B 看作是许多个向量 x 组成的，像这样：

$$AB = A \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} Ax_1 & Ax_2 & \dots & Ax_n \end{bmatrix}$$

于是， AB 的每一列都将是一个经过 A 线性变换的向量。如果将 AB 转置过来，那么将会是将每一个线性变换后的向量转置过来，也就是：

$$(AB)^T = \begin{bmatrix} (Ax_1)^T \\ (Ax_2)^T \\ \dots \\ (Ax_n)^T \end{bmatrix} = \begin{bmatrix} x_1^T A^T \\ x_2^T A^T \\ \dots \\ x_n^T A^T \end{bmatrix}$$

于是我们观察后右面，于是得到，实际上，就是 B 的转置与 A 的转置相乘罢了。像这样：

$$(AB)^T = \begin{bmatrix} (Ax_1)^T \\ (Ax_2)^T \\ \dots \\ (Ax_n)^T \end{bmatrix} = \begin{bmatrix} x_1^T A^T \\ x_2^T A^T \\ \dots \\ x_n^T A^T \end{bmatrix} = \begin{bmatrix} x_1^T & x_2^T & \dots & x_n^T \end{bmatrix} \begin{bmatrix} A^T \\ A^T \\ \dots \\ A^T \end{bmatrix} = B^T A^T$$

于是我们验证了这个公式的正确性：

$$(AB)^T = B^T A^T$$

那么 ABC 的转置呢？同样的道理，我们现将 BC 看作一个整体，像这样：

$$(ABC)^T = (BC)^T A^T = C^T B^T A^T$$

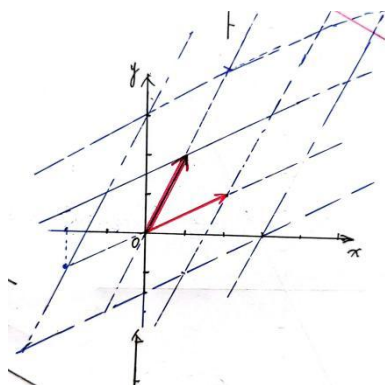
同理，很多矩阵相乘的转置依旧是将这些矩阵的转置倒过来相乘。

Chap2.8 线性相关的本质

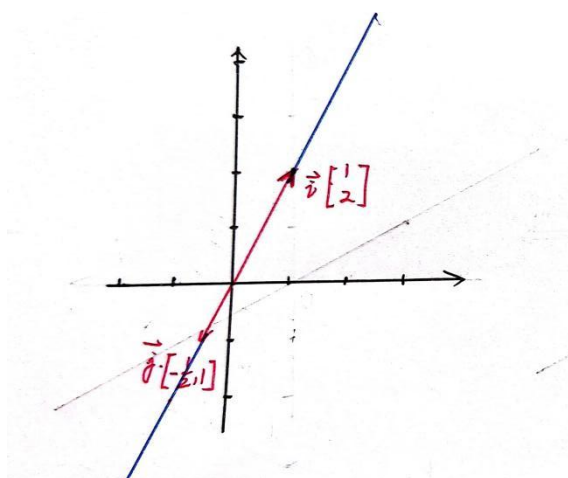
之前我们提到，线性代数的核心，是由基向量的线性组合产生的线性空间的变换。所以我们接下来研究基向量线性组合所产生的线性空间的形态。

例如基向量 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，和基向量 $j = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ，那么这两个向量通过线性组合产生的线性空间

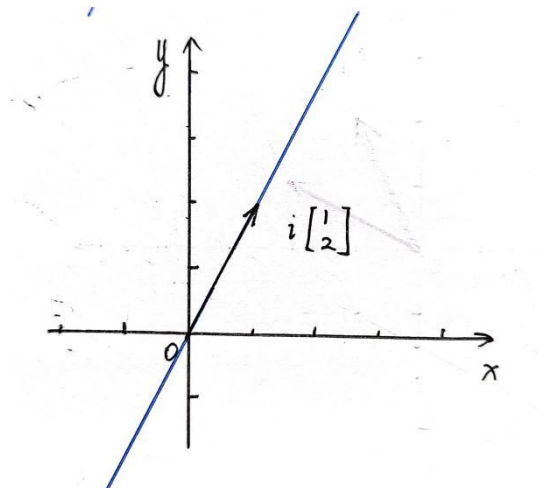
将会覆盖整个二维平面，像这样：



但是如果是两个共线的基向量，例如基向量 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，和基向量 $j = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}$ ，那么这两个基向量线性组合所形成的空间将只会是一维空间，像这样：

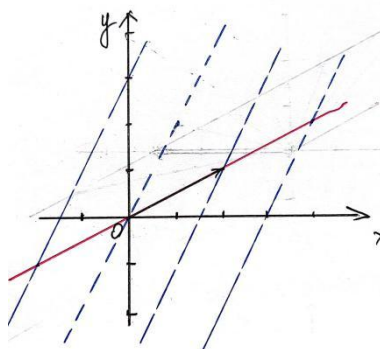


为什么会这样呢？稍加思考我们发现，在第一个例子中，基向量 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 可以提供一个维度，像这样：



而在这个维度的基础上，可以通过另一个基向量的伸缩，即 $j = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}$ 的伸缩，使得

基向量 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 所产生的一个维度（或一个直线），平移起来，从而覆盖整个二维平面，像这样：



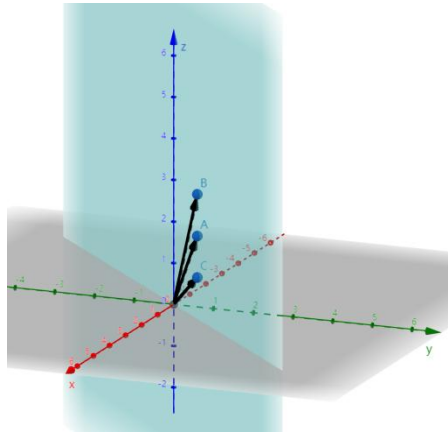
但是，对于基向量 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，和基向量 $j = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}$ ，这两个向量所产生的维度是相同的，同一条直线，同时地落在同一条直线上，其中一个基向量对“维度”没有任何贡献。换句话说， $j = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}$ 可以通过伸缩得到 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ，所以 $i = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 没有存在的意义。我们把可以通过伸缩或者压缩等线性变换可以相互转换的两个向量称为“线性相关”。

换句话说，如果由这件事成立，则称 i 与 j 线性相关：

$$ai = bj$$

此时无法扩展到整个空间当中。

我们再将视野扩展的三维空间，对于许多的基向量，其线性组合都应当是这个三维的立体空间，但是，有一些不是这样的，例如在前几章中提到的这幅图，像这样：



此时向量 A, B, C 并没有通过线性组合产生整个三维空间，而是停留于一个平面。为什么呢，其实从前面的推理就已经可以看出， A 可以由 B 和 C 线性组合而产生，也就意味着它们之间也“有关联”，我们把这件事写出来，也就是：

$$ai = bj + ck$$

我们观察一下这两个式子，便意识到标题“线性相关”的来由。

$$ai = bj \quad ai = bj + ck$$

线性运算包括两个，一个是数量乘法，一个是向量加法，这两者的定义贯穿了向量运算：

$$\vec{a} + \vec{b} = \vec{c}$$

$$k\vec{a} = \vec{ka}$$

但是，这一写法不太美观，于是我们写作：

假若， $ai + bj + ck = 0$ 则称 i, j, k 线性相关

那么我们如何判断向量之间是否线性相关呢？

换句话说我们需要搞清楚，是否存在一个 a, b, c 这样一个都不是 0 的系数，使得 i, j, k 三个基向量通过 a, b, c 这三个系数的线性组合能够得到 0。

那换句话时候，比如说由 i, j, k 三个基向量构成的矩阵是这样的：

$$A = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$$

那么也就要寻找 (x, y, z) ，使得经过 A 矩阵的线性变换后，可以变成 $(0, 0, 0)$ ，像这样：

$$\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

如果写成线性方程组就是这样：

$$ax + dy + gz = 0$$

$$bx + ey + hz = 0$$

$$cx + fy + iz = 0$$

如果写得更简洁一些，将会是这样：

$$\mathbf{Ax} = \mathbf{0} \quad (\mathbf{x} \neq \mathbf{0})$$

这也是最后我们得出的结论，如果向量 \mathbf{i} , \mathbf{j} , \mathbf{k} 所组成的矩阵，可以找到一个向量 $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ （不是零向量）使得这个向量输入矩阵 \mathbf{A} 后，输出零向量，那么就意味着 \mathbf{i} , \mathbf{j} , \mathbf{k} 是线性相关的。

Chap2.9 零空间与通解的本质

之前我们提到了这个方程的求解问题：

$$Ax = 0$$

这个方程的解可能是唯一的，即零向量。也有可能是无穷多个，例如一个矩阵

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 4 & 7 \end{bmatrix}$$

首先，我们是否可以一眼看出来这个矩阵对应的方程 $Ax = 0$ ，一定有无穷多个解呢？是的，这是必定的！因为第一列的向量加上第二列的向量，等于第三列的向量，像这样：

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 7 \end{bmatrix}$$

所以，必定的将会有无穷多个线性组合，使得结果是零向量。

我们来用初等行变换化简这个矩阵，像这样：

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

我们发现只有 2 个主元，下面我们迫切地想求得它的通解，于是，我们不妨继续化简：

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

于是我们将其翻译为线性方程组的形式，首先我们要明白，主元是前两列，而自由变量是第三列，翻译后就是这样：

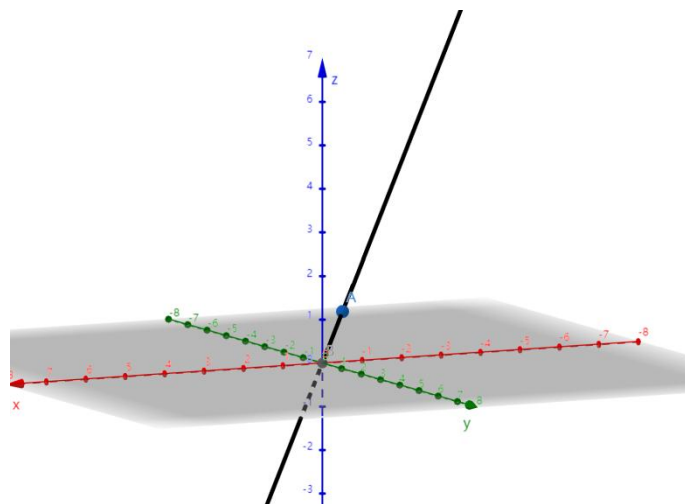
$$x_1 + x_3 = 0$$

$$x_2 + x_3 = 0$$

那么如果将解写作： $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ ，那么将会有下面的结果：

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_3 \\ -x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

如果我们想象一下，这个解在另一个空间当中，那么在直观上就是这样的：



这就是零空间，其内部包含所有满足 $Ax = 0$ 的 x 。

下面我们考虑，如何求解零空间中的全部解，是的，前面我们已经成功求解了，但是并不太优雅。

例如，我们考虑这样的矩阵：

$$A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 6 & 4 & 8 \\ 3 & 8 & 6 & 10 \end{bmatrix}$$

首先，他也必定是由无穷多个解的，因为第三列向量是第一列向量的二倍。我们来进行一下高斯消元法的化简，于是将会得到：

$$A \sim \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 2 & 0 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

我们继续通过之前的方式求解，也就意味着，需要我先将这个矩阵翻译成线性方程组的形式，像这样：

$$\begin{aligned} x_1 + 2x_3 - 2x_4 &= 0 \\ x_2 + 2x_4 &= 0 \end{aligned}$$

这也就意味着，同时存在了两个自由变量。如果我们将解同样地，写作向量的形式 $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ ，

像这样，那么将会得到这样的一个解：

$$\begin{bmatrix} -2x_3 + 2x_4 \\ -2x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ 0 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 2x_4 \\ -2x_4 \\ 0 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

但这很不优雅。

分析一下等式的右侧，也就是两个基向量的线性组合，将会组成一个平面。这和矩阵的

定义是一样的，矩阵的本质就是向量的线性组合不是吗？所以，我们可以把解写成这样的形式，即，

$$\text{solutions} = \begin{bmatrix} -2 & 2 \\ 0 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

下面来观察一下解和简化后的矩阵之间的关系，你是否有一些惊人的发现？

$$\begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

换句话说，我们是否可以一眼看出来这个矩阵的解呢？

首先，最后的一行都是 0，这都不需要考虑，因为这是永远成立的，我们可以将矩阵化简为下面这样：

$$\begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

现在，是否能看出来这其中奇妙的关系了呢？是的，首先对于主元列，其本质就是一个单位矩阵，而后两列是自由变量的系数所构成的矩阵，我们叫它 **F** 矩阵（Free）。于是，上面这个矩阵的求解过程就可以写成这样：

$$[\mathbf{I} \quad \mathbf{F}] \mathbf{x} = \mathbf{0}$$

现在是否能够一眼看出来这个方程的解 \mathbf{x} 究竟是什么呢？显然地， \mathbf{x} 将会是这样：

$$\mathbf{x} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{I} \end{bmatrix}$$

这也这个通解最为优雅的表达形式，但是其本质是与前面的推导没有区别的，只不过很美而已。

Chap2.10 通解与特解的本质

当然，这里谈论的解，指的是线性方程组的解。比如这个线性方程组：

$$\begin{aligned} x + y &= 2 \\ 2x + y &= 3 \end{aligned} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

如果解出来，可以得到 $x=1, y=1$ ，我们用 (x, y) 来表示这个解，是一个向量的形式。

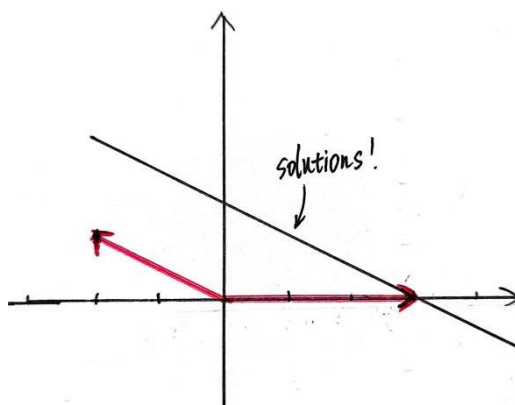
上述情况是一个很幸运的情况，即有唯一解，在前面的讨论中，我们认识到，可能存在无穷解的情况，像这样：

$$x + 2y = 3 \Leftrightarrow \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix}$$

那么主元是 x ，而自由变量是 y ，如此看来，如果将其表示为向量的形式，将会变成这样：

$$x = 3 - 2y \Leftrightarrow \begin{bmatrix} 3 - 2y \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + y \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

那么如果是在几何当中，将会是这样：



前面是通过二维空间的角度，明白解的几何形态，于是我们可以一类简单的线性方程组，像这样：

$$\begin{aligned} x + y &= 0 \\ 2x + y &= 0 \end{aligned} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

这和上面的方程是一样的，我们用同样的角度来解这个方程，由于住院都在系数矩阵，所以这个线性方程组将会有唯一的解：

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

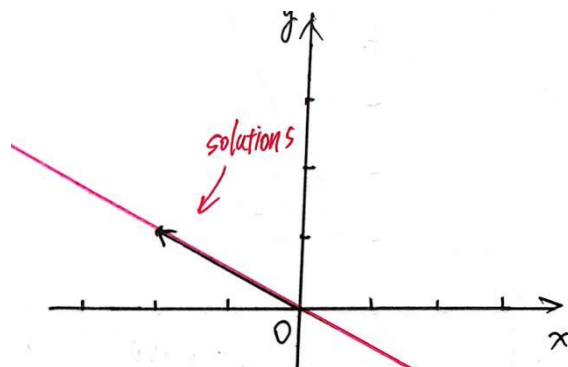
下面，我们考虑这样一个没有唯一解的方程，像这样：

$$x + 2y = 0 \Leftrightarrow \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

和上面一样，我们化简为这样：

$$x = -2y \Leftrightarrow \begin{bmatrix} -2y \\ y \end{bmatrix} = y \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

如果在几何中，将会是这样：

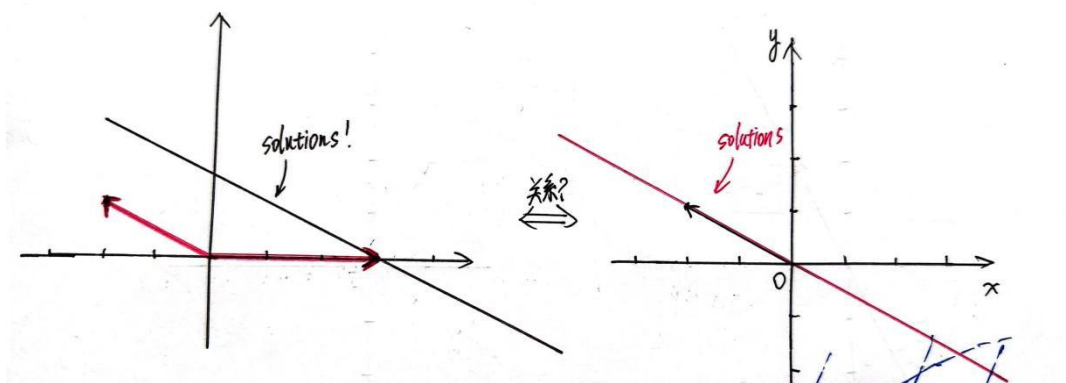


我们来观察前后的两个解，首先从代数上对比：

$$\begin{bmatrix} 3-2y \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + y \begin{bmatrix} -2 \\ 1 \end{bmatrix} \quad \begin{bmatrix} -2y \\ y \end{bmatrix} = y \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

我们可以将左侧的“非齐次的线性方程组”的解抽象为 $\mathbf{x}=\mathbf{p}+t\mathbf{v}$ ，而右侧的“齐次线性方程组”的解，抽象为 $\mathbf{x}=t\mathbf{v}$ 。

可见，似乎这二者之间有一些关系，从几何上看，非齐次线性方程组的解就是齐次线性方程组的解经过了一个平移。



这一结论是否具有普遍性呢？

我们考虑任意的一个矩阵 $\mathbf{A} = \begin{bmatrix} a & b \end{bmatrix}$ ，那么，对于任意一个满足这个矩阵所对应方程组的解 (x, y)

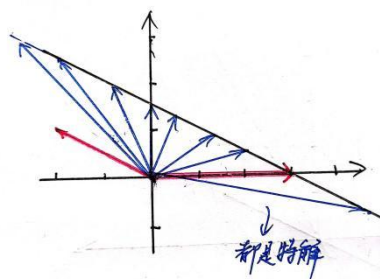
对于齐次方程组，将会有： $ax+by=0$ 也就是 $x=-\frac{b}{a}y$ 。如果写成向量的形式，也就

是 $\begin{bmatrix} -\frac{b}{a}y \\ y \end{bmatrix} = y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix}$ 。也可以看做是 $\mathbf{x}=t\mathbf{v}(t=y, \mathbf{v}=\begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix})$ 的形式。

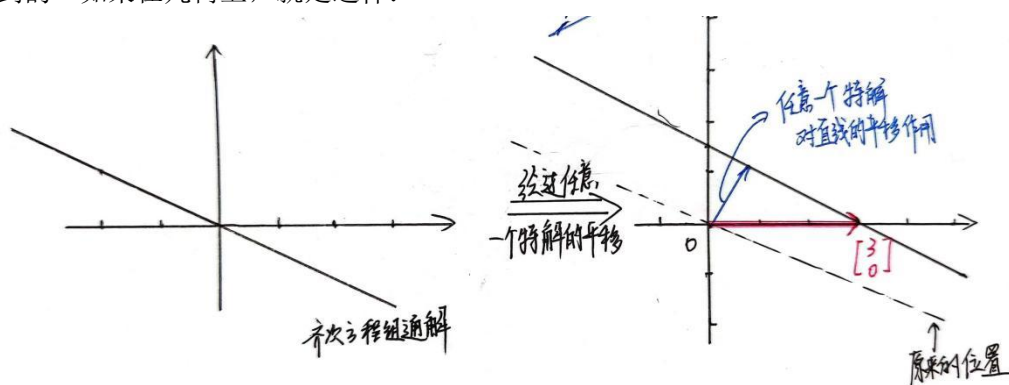
对于非齐次的线性方程组，将会有： $ax+by=c(c \neq 0)$ 也就是 $x=\frac{c}{a}-\frac{b}{a}y$ 。也就是

$$\begin{bmatrix} \frac{c}{a} - \frac{b}{a}y \\ y \end{bmatrix} = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix} + y \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix}。也可以看做是 $x = p + tv (p = \begin{bmatrix} \frac{c}{a} \\ 0 \end{bmatrix}, t = y, v = \begin{bmatrix} -\frac{b}{a} \\ 1 \end{bmatrix})$ 的形式。$$

所以我们发现，非齐次线性方程组的解，就是齐次线性方程组的通解，加上一个非齐次线性方程组的特解。这个特解可以是这个方程组的任意一个解，如果在几何上就表示成这样：



而非齐次线性方程组的全部解，就是由齐次线性方程组的解通过任意一个特解的平移所得到的！如果在几何上，就是这样：



所以，请记住：

非齐次线性方程组的通解=非齐次线性方程组的特解+齐次线性方程组的通解

Chap3.1 秩的本质

是的，这是新的一个章节——线性空间。

首先，我更喜欢“秩”的英文，叫做“Rank”，意味等级。其实在中文中，“秩”代表“秩序”，也是一种顺序，也是一种等级。什么的“等级”，秩一般是矩阵中的概念，也就是矩阵的等级。

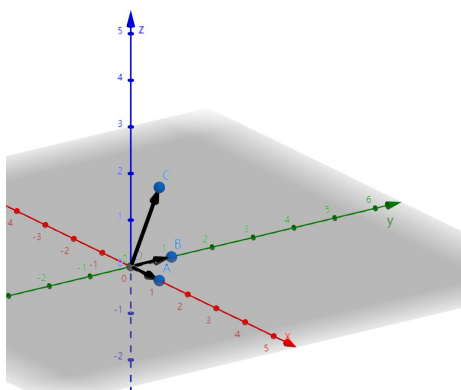
我们先从几何角度来直观说明秩究竟是什么？

前面，我们一直在讨论的核心内容在于，“线性组合”，即对于一个矩阵，我们把他的列看作一个个向量，从而进行线性组合，产生空间中的图形，例如直线，平面或者三维空间等。

例如：

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

对于这个矩阵而言，我们如果检验一下它的每一列，会发现每一列之间是线性无关的。当然这可以一眼看出来，因为前两个向量在 XOY 平面上，而第三个向量并不在 XOY 平面上，像这样：



是的，这三个向量如果进行线性组合的话，我们很自然的可以发现，它将覆盖整个三维空间。请暂时留个印象：这个矩阵线性组合后是三维空间。

但是，总是有些热衷于代数的教授，喜欢运用消元法进行验证，这是一样的，像这样：

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

那么，让我们稍微改一改这个矩阵，像这样：

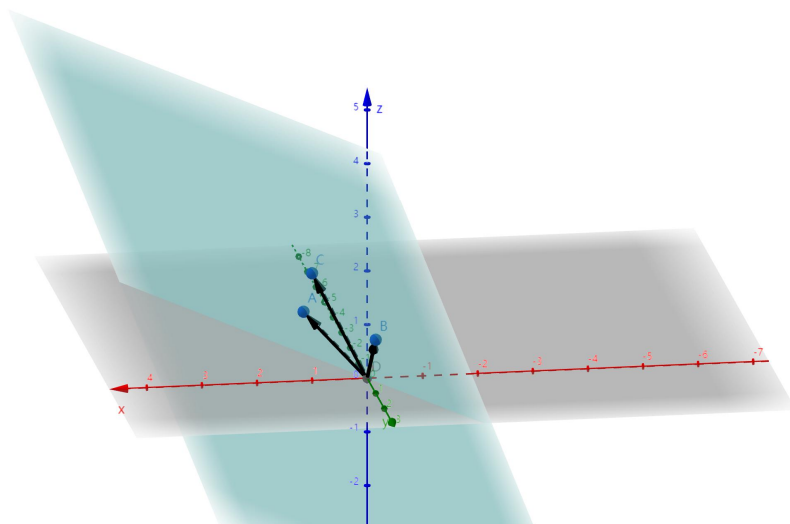
$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix}$$

那么，显然我是精心设计了这样的一个矩阵。因为矩阵前两列构成的向量加起来就等于后一列。对于热衷于代数方法的同学，如果我们将它做消元法的话，也就是观察 $Ax=0$ 是否有非零解的话，将会是这样：

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

显然，对于 $Ax=0$ ，他将会有无穷多个解。

但是，我更热衷于几何直观，如果在图像中，将会是这样：



是的，这个矩阵的基向量将会形成一个平面。

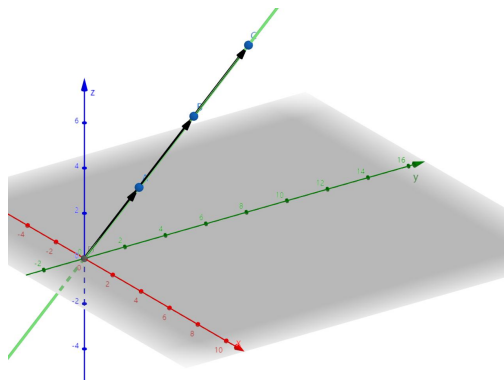
下面，继续更加极端一些，如果这个矩阵长成这样呢？

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

是的，这一眼就看一看出是非常线性相关了，因为每一个列向量互相已经成倍数关系了，那么我们再次满足那些喜欢代数方法的同学，对它进行高斯消元，于是将会得到：

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

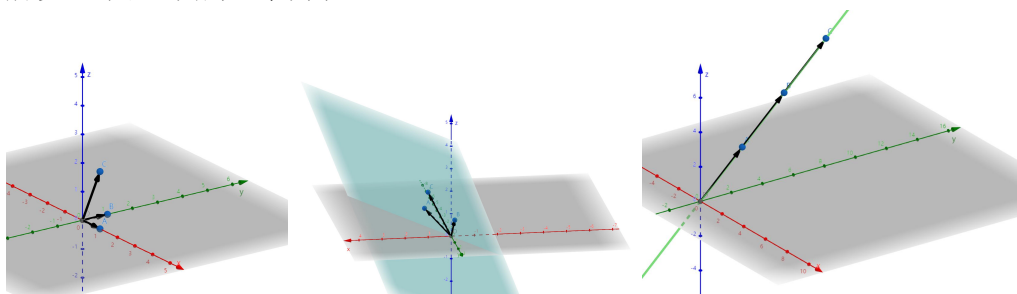
但是，从几何上，这更加直观，这三个向量的线性组合将会是空间中的一条直线，也就是一个一维空间。



所以，从上面这三个例子，我们就可以引出“秩”（rank，记作 r ）的本质了，即矩阵各个列向量形成的列空间的维度。我们把其记作 $\text{Dim}(C(A))$ ，其中 Dim 代表维度，即 dimension，而 C 代表 column，代表矩阵 A 中列向量形成的空间。于是我们有以下定义：

$$r(A) = \text{Dim}(C(A))$$

所以，对于上面的三个例子：



秩将分别对应 3, 2, 1。因为它们的列生成的空间将会是三维，二维，一维。这正是“秩”的本质。

事实上，热衷于代数的人将会不乐意，如何给“秩”一个代数的定义呢？

我们再继续观察前面的三个例子的高斯消元法的结果，将会是这样：

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

是的，它们的主元数量分别是 3, 2, 1。所以，我们也可以对应着说，“秩”代表着一个矩阵的“主元”数量，而“主元”数量和列空间的维度是相等的。也就是说，我们有以下的表达式：

$$r(A) = \text{Dim}(C(A)) = A \text{ 主元数量}$$

于是，很自然的问题是，列空间的维度为什么和主元的数量是相等的？

在前面的章节中，我们认识到，消元法可以用矩阵来描述，像这样：

$$E_{(3,2)}E_{(3,1)}E_{(2,1)}A = U$$

或者，简化一点，写成这样：

$$EA = U$$

而矩阵的本质是线性变换，所以对于一个矩阵，如果消元法之后是这样的：

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

那么，显然，这个化简后的矩阵将会生成一个三维空间，所以如果将 E 逆过来，即我们如果想从 U 回到 A ，像这样：

$$A = E^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

是的一个三维的立体空间，经过 3×3 矩阵的消元矩阵的变换，必然仍然是一个三维空

间，因为每一个消元矩阵的每一列都是线性无关的，所有的消元矩阵相乘，自然也是线性无关的，所以消元矩阵的逆，必然依旧线性无关。这也就意味着 A 所形成的空间，经过三维空间内的线性变换，依旧是一个三维空间。

同理，对于两个主元的情况，我们可以用相同的思路：

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

显然，这将会生成一个平面，因为这个矩阵的第三个分量是 0 ，也就要求生成的所有向量的 z 坐标为 0 ，也就意味着将生成 XOY 平面。

同样地我们通过消元矩阵的逆矩阵将整个过程倒回去，也将会得到仍然是一个平面。

这样我们就再次验证了这个理论的正确性：

$$r(A) = \text{Dim}(C(A)) = A \text{ 主元数量}$$

Chap3.2 行空间的本质

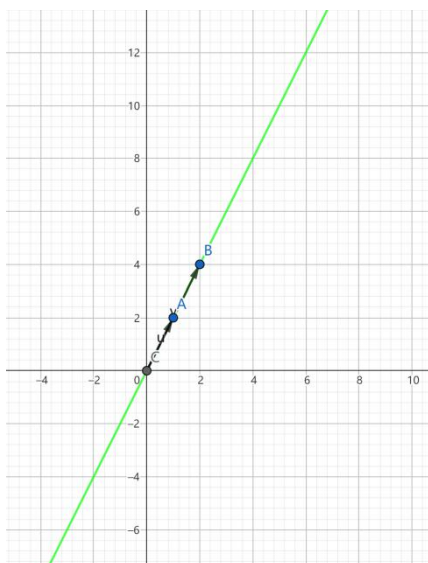
首先，在前面我们已经认识到了列空间。例如，我们给出这样的一个矩阵：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

如果从列的角度，那么它的两个基向量是这样的：

$$i_c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, j_c = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

那么，如果在几何当中，这两个基向量所形成的空间将会是这样的：



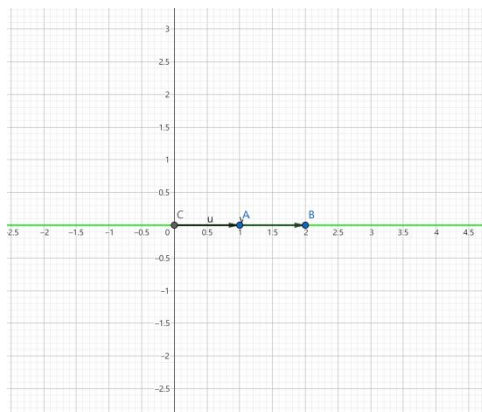
我们尝试着这个矩阵进行初等行变换，于是得到：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

那么新产生的两个基向量将会是这样：

$$i'_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, j'_c = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

如果在空间中，那么列空间将会是这样：



显然，列空间改变了，换句话说， $(1,0)$ 并未位于原来的空间。但是在这个过程中，什么是不变的？本节的题目叫做“行空间”，这也就意味着我们可以将行看作一个向量，像这样：

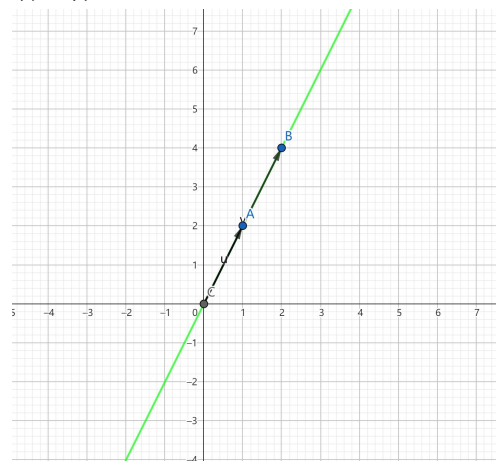
对于矩阵：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

它的行向量是这样的：

$$i_r = [1 \ 2], j_r = [2 \ 4]$$

如果画在图中，代表着这样：



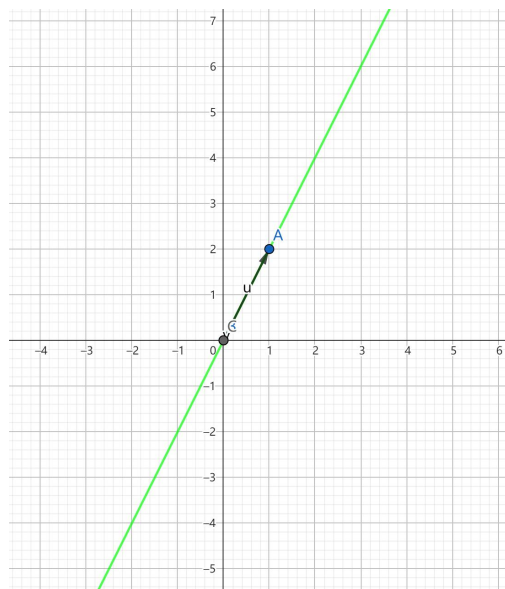
是的，如果我们再从化简后的矩阵来看：

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

它的两个行向量是这样的：

$$i_r = [1 \ 2], j_r = [0 \ 0]$$

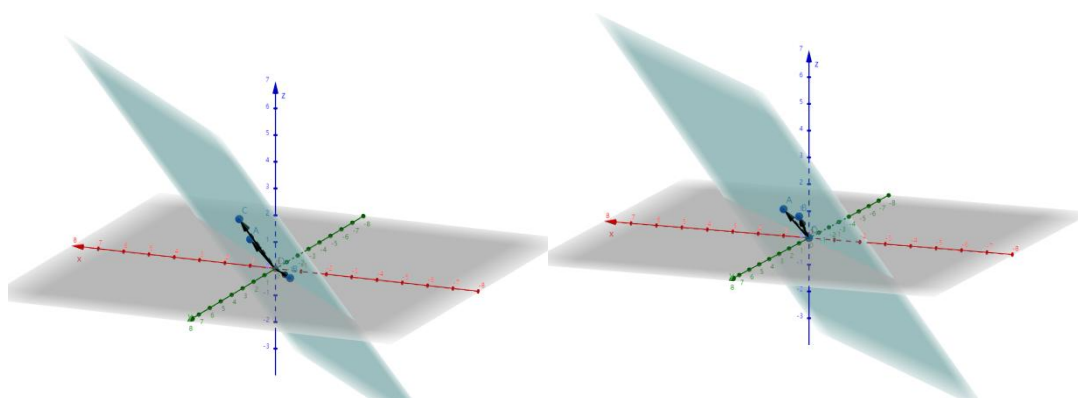
如果画在图中，将会是这样的：



是的，行空间并没有改变。这是否只是一个偶然呢？我们再举几个例子，像这样：

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

可见，它们的行空间也没有改变。



可见，在初等行变换（就是消元法），列空间将会变化，但是行空间却不变化。这是为什么？首先我们先分析为什么列空间会变化——因为列向量变化了。但是行向量也变化了呀。但是行向量是已有向量的线性组合，所以在已有空间内，无论怎样去线性组合，都一定仍然位于这个线性空间当中。这就是为什么行空间是不变的。

于是我们得出下面的结论：

初等行变换不改变行空间，改变了列空间

下面我们考虑行空间和列空间的关系。

首先，由于之前我们一直在考虑“列向量”，或列向量，即“ $C(A)$ ”。首先，行空间意味着对 A 的每一行进行线性组合，那么如果我们将 A 转置过来，就将会称为是对 A 的列进行线性组合，像这样：

$$\text{combs of rows of } A = C(A^T)$$

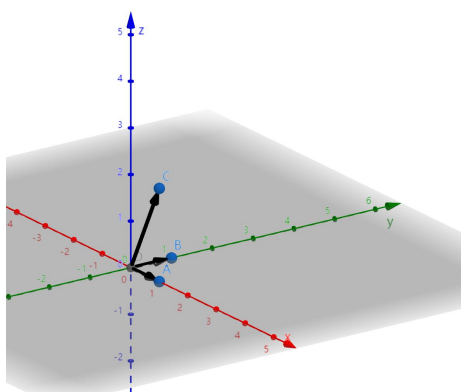
于是我们下面来讨论着二者的关系—— $C(A)$ 和 $C(A^T)$ 。首先让我们观察一下前面的三个例子。

(1) 像这样的一个矩阵：

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

它的每一列向量都是线性无关的，所以列空间将是整个三维空间。它的三个列向量分别是：

$$i_c = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, j_c = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, k_c = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$



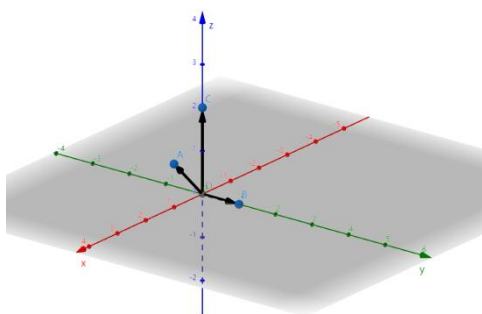
那么如果是行空间的话，我们还是考虑 A^T 吧，首先将这个矩阵转置过来，将会有：

$$A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

那么它的每一个列向量，但实际上是 A 的行向量，将会是这样：

$$i_r = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, j_r = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, k_r = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

如果是画在图里，那么将会是这样：



可见，这个矩阵所形成的空间仍然是三维的，也就意味着列空间和行空间的秩都是 **3**。

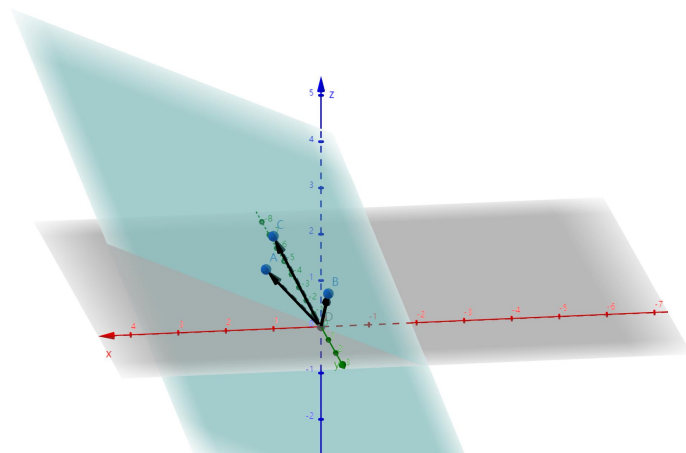
(2) 像这样一个矩阵：

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

它的列向量是这样的：

$$i_c = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, j_c = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, k_c = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

如果在图像中，将会是这样：



是的，这个矩阵的基向量将会形成一个平面，也就意味着这个矩阵所形成的列空间的秩是 2。

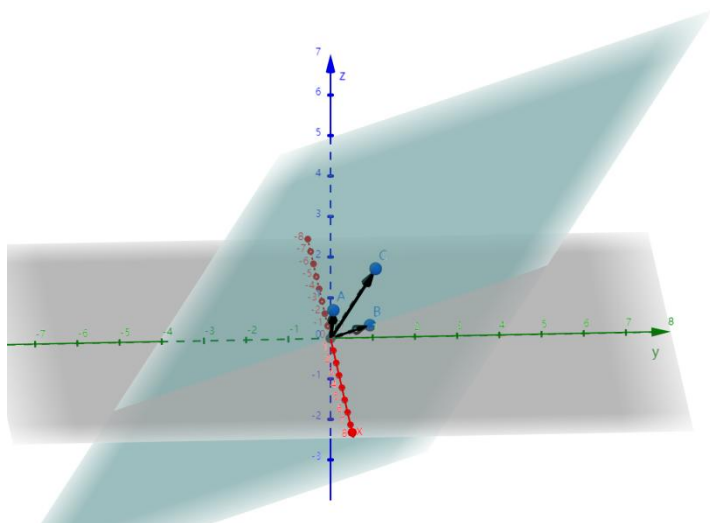
下面我们继续将这个矩阵进行转置，像这样：

$$A^T = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$

那么，它的列向量，也就是 A 的行向量，将会是这样：

$$i_r = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, j_r = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, k_r = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

如果在图像中，将会是这样：



这也就意味着行空间和列空间的秩都是 2。

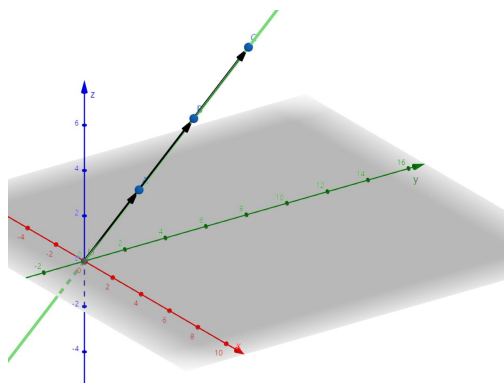
(3) 像这样的一个矩阵：

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{bmatrix}$$

那么这个矩阵的三个基向量将会是：

$$i_c = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, j_c = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, k_c = \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix}$$

从几何上，这三个向量的线性组合将会是空间中的一条直线，也就是一个一维空间。



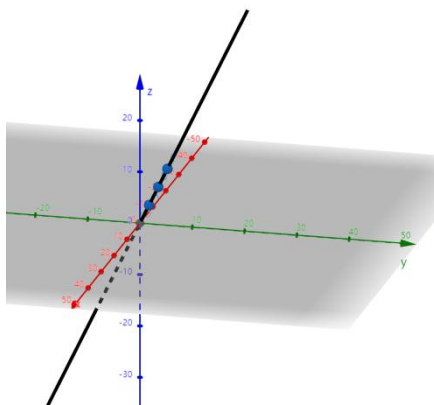
那么，如果我们将它转置过来，将会得到：

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 4 & 8 & 12 \end{bmatrix}$$

那么它的行向量，也就是 A 的列向量，将会是：

$$i_r = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, j_r = \begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix}, k_r = \begin{bmatrix} 3 \\ 6 \\ 12 \end{bmatrix}$$

所以，从集合上，它的图像是这样的：



这也就意味着行空间和列空间的秩都是 1。

或许你会觉得我有些啰嗦。但不是这样的，你是否发现了一些规律？没错，行空间和列空间的秩永远是相同的，也就是说，下面这个结论是永远成立的：

$$\text{rank}(C(A)) = \text{rank}(C(A^T))$$

这看起来是正确的，因为前面，我们的三个例子当中，都谈到了这一点，也验证了它的正确性。

如何去证明呢？我们再次回归上一节所讨论的核心——秩的本质：

$$r(A) = \text{Dim}(C(A)) = A \text{ 主元数量}$$

所以，行空间和列空间的主元数是否相等呢？

例如这两个矩阵：

$$A = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$

显然，如果从这些矩阵来看的话，似乎看不出来，因为这太复杂了。下面我们思考，如何将“行空间”与“列空间”构建关联。

首先我们回顾一下之前的概念“零空间”，这也正是它发挥大用场的时刻。

零空间，也可以理解为 $Ax=0$ 的所有的解。

例如：

$$Ax = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

那么也就意味着寻找 A 的行的线性组合，使得行的线性组合为零向量。那么，

当有 2 个主元的时候，将有 1 个自由变量，此时零空间将是一维的。

当有 1 个主元的时候，将有 2 个自由变量，此时零空间将是二维的。

当有 0 个主元的时候，这也就意味着 A 全部是 0，此时任意的 x 都可以满足 $Ax=0$ ，此时零空间是三维的。

那么如果我们设主元的数量是 rank ，列空间的维度是 n ，那么将会有下面的表达式成立：

$$\text{rank}(N(A)) = n - \text{rank}(C(A))$$

这是否是绝对正确的呢？

我们先考虑一个三维空间的情况，像这样：

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

这个矩阵经过高斯消元后将会变成这样：

$$A \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

其化简之后，有两个主元，也就意味着 $\text{rank}(C(A)) = 2$ ，我们简单记作 $r=2$ 。于是我们可以将这个表达式写的更加优雅一些。如果我们将右侧的自由变量所对应的列都记作 F ，那么 A 可以写作：

$$A \sim \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

其中 I 是 $r \times r$ 的，而 F 是 $r \times (n-r)$ 的。

这也就意味着，关于 $Ax=0$ 的解，将会是这样（之前的章节提到过）：

$$x = \begin{bmatrix} -F \\ I \end{bmatrix}$$

这个解应当一共有 $n-r$ 列，并且由于下面是单位矩阵，所以必然地，每一列将会是线性无关的。所以这个解所组成的空间，也就是零空间的秩，将会是 $n-r$ 。

也就是说，我们得到了以下的结论：

$$\text{rank}(N(A)) = n - \text{rank}(C(A))$$

我们考虑下面的 $m \times n$ 矩阵，像这样：

那么这个矩阵 A 的转置将会是这样：

同样地，我们可以得到一下结论：

$$\text{rank}(N(A^T)) = m - \text{rank}(C(A^T))$$

然后我们需要去寻找列空间和行空间的关系，一切的关系将在零空间时展现出来，对于一下方程：

$$Ax = 0$$

我们将 x 看作是一个一个向量。 A 看作一个一个向量的集合，像这样：

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \dots \\ A_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} \vec{A_1 x} \\ \vec{A_2 x} \\ \dots \\ \vec{A_n x} \end{bmatrix} = 0$$

于是，每一个点乘，或者更确切地说，投影，都应该是 0 ，这和我们 Chap1.6 中谈的点积的本质是一样的。这里每一个 Ax ，都等于 0 ，代表着 x 与 A 的行空间垂直，因为正是这些

行向量组成了行空间。而 \mathbf{x} 有可以代表是 \mathbf{A} 的零空间，所以我们有以下两个结论成立：

$$\text{rank}(N(\mathbf{A})) = n - \text{rank}(\mathbf{C}(\mathbf{A}))$$

$$N(\mathbf{A}) \perp \mathbf{C}(\mathbf{A}^T)$$

是的，这样我们就似乎能找到二者之间的关系了。因为行空间和列空间通过零空间联系在了一起。

下面，我们一步一步来思考如何进一步翻译这个条件。

如果在二维空间中。

例如 $\mathbf{C}(\mathbf{A}^T)$ ，恰好是一条直线，那么 $N(\mathbf{A})$ ，由于垂直于一条直线，那么它将会也是一条直线。

例如 $\mathbf{C}(\mathbf{A}^T)$ ，是整个二维平面，那么 $N(\mathbf{A})$ ，由于垂直于一个平面，那么它将会只是一个点。

如果在三维空间中，

例如 $\mathbf{C}(\mathbf{A}^T)$ ，恰好是一条直线，那么 $N(\mathbf{A})$ ，由于垂直于一条直线，那么它将会也是一条平面。

例如 $\mathbf{C}(\mathbf{A}^T)$ ，是整个二维平面，那么 $N(\mathbf{A})$ ，由于垂直于一个平面，那么它将会只是一条直线。

你发现了关系吗？也就是这个关系：

$$\text{rank}(N(\mathbf{A})) = m - \text{rank}(\mathbf{C}(\mathbf{A}^T))$$

是的，如果我们将这个式子与上面的式子联系在一起。

$$\text{rank}(N(\mathbf{A})) = n - \text{rank}(\mathbf{C}(\mathbf{A}))$$

$$\text{rank}(N(\mathbf{A})) = n - \text{rank}(\mathbf{C}(\mathbf{A}^T))$$

是的我们化简可以得到：

$$\text{rank}(\mathbf{C}(\mathbf{A})) = \text{rank}(\mathbf{C}(\mathbf{A}^T))$$

所以行空间的秩和列空间的秩是相等的。

在前面，我们证明了矩阵的行空间和列空间的秩是相同的。我们通过的是寻找零空间与行空间的关联进而找到了秩是相同的。下面我们继续完善这个理论，并且试图从几何的角度直观地展现出这个结论：

举个例子：

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

那么它的行空间将会是：

$$\mathbf{C}(\mathbf{A}^T) = c \begin{bmatrix} 1 \\ 2 \end{bmatrix} (c \in \mathbb{R})$$

而其零空间，由于这个矩阵一眼就看得出来秩是 1，所以零空间的维度应当也是 1，即一条直线。

为了求出零空间，我们对 A 进行初等行变换。

于是，我们得到：

$$A \sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

是的，如果你还记的那个优美的公式，将会一眼看出来零空间是：

$$I = c \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

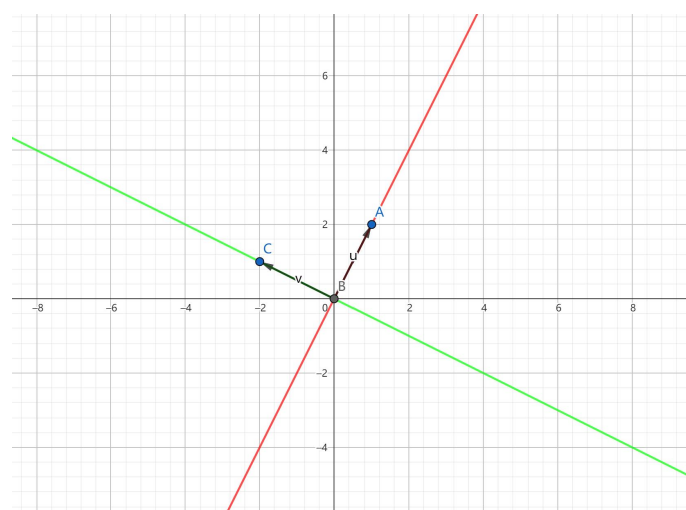
如果没有看出来，我们同样地将 A 转化为线性方程组，于是：

$$x_1 + 2x_2 = 0$$

于是，得到同样的答案：

$$I = \begin{bmatrix} -2x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

我们将这两个空间在图中表示出来：



你发现了什么？

是的，这正是我们上节课的结论，行空间和零空间是垂直的，也就意味着我们有以下的结论：

$$N(A) \perp C(A^T)$$

下面，我们考虑 $N(A^T)$ 和 $C(A)$ 的关系，依旧是刚刚的这个矩阵：

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

那么，它的行空间就是列空间，而转置后也仍然是一样的，因为它是一个对称矩阵。于是我们换一例子吧。可恶。例如，这样一个矩阵：

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

那么它的列空间将会是：

$$C(A) = c \begin{bmatrix} 1 \\ 2 \end{bmatrix} (c \in R)$$

然后，如果将 **A** 转置过来，将会是：

$$A^T = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

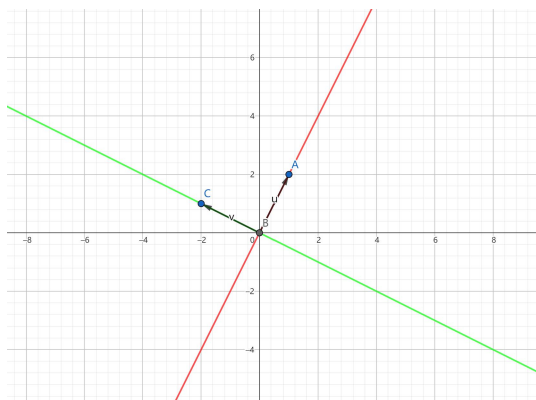
那么我们同理对 A^T 进行化简，像这样：

$$A^T = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

于是 $N(A^T)$ 将会是：

$$N(A^T) = c \begin{bmatrix} -2 \\ 1 \end{bmatrix} (c \in R)$$

我们依旧在图形中画出来，将会是这样：



于是我们可以大致地得出这样的一个结论：

$$N(A^T) \perp C(A)$$

或许你觉得这太草率了，不是吗？我们现在进军三维空间，例如这样一个矩阵：

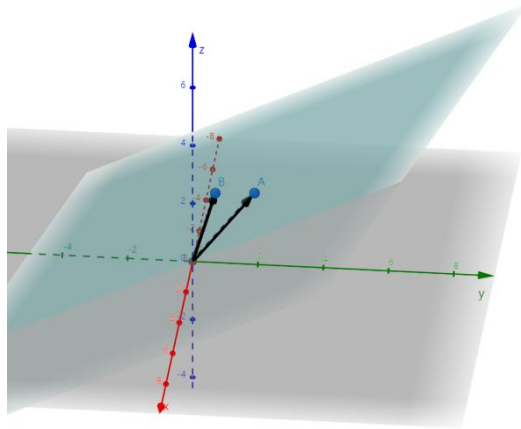
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 1 & 4 \end{bmatrix}$$

首先一眼可以看出这个矩阵的秩是 2，因为第三列是第一列和第二列的和，所以列空间的秩是 2，行空间的秩也是 2。这也就意味着列空间和行空间都应当是 2 维的。

同样地，我们先研究行空间，也就意味着是三个行向量的线性组合：

$$C(A^T) = a \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + b \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \quad (a, b \in \mathbb{R})$$

那么如果在图像中，将会是这样：



下面我们考虑零空间：

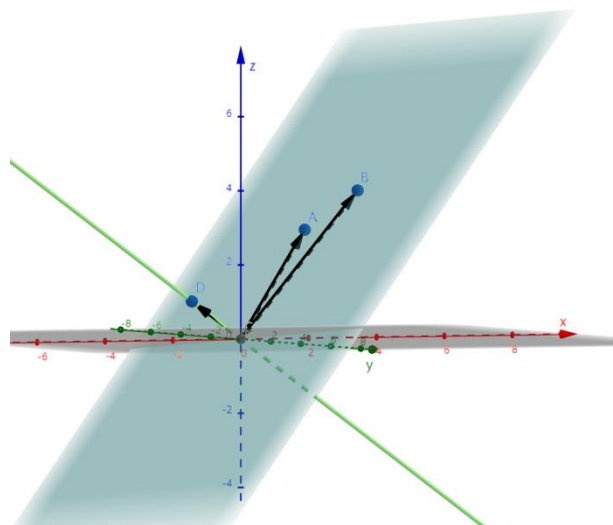
同样地，依旧对 A 进行消元法：

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

于是乎，我们可以看出来，零空间将会是：

$$N(A) = c \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \quad (c \in \mathbb{R})$$

于是这个直线画在刚刚的图像中，将会是这样：



所以，我们再次验证了这个公式的正确性：

$$N(A) \perp C(A^T)$$

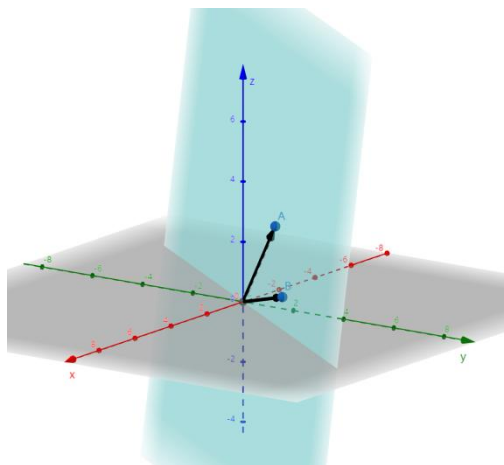
下面，我们考虑 $N(A^T)$ 和 $C(A)$ 的关系，依旧是刚刚的这个矩阵：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 1 & 4 \end{bmatrix}$$

首先，由于秩是 2，那我们只需要选择任意的两列列向量即可，比如前两列即可：

$$C(A) = a \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + b \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \quad (a, b \in R)$$

如果画在图中将会是这样：



下面我们转置这个矩阵，像这样：

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 5 & 4 \end{bmatrix}$$

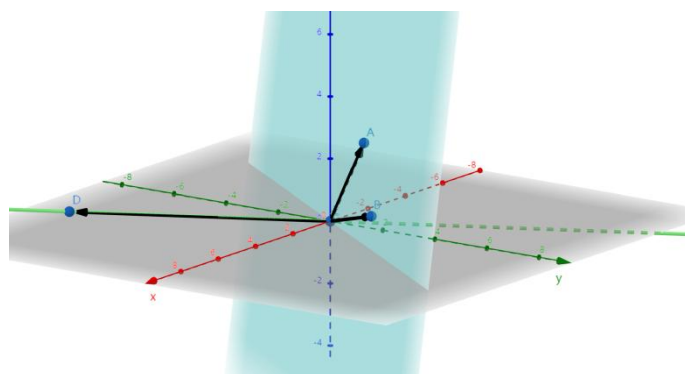
那么，进行消元法化简，于是得到：

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 5 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & -1 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

那么，可以得到 $N(A^T)$ 将会是：

$$N(A^T) = c \begin{bmatrix} 7 \\ -5 \\ 1 \end{bmatrix} \quad (c \in R)$$

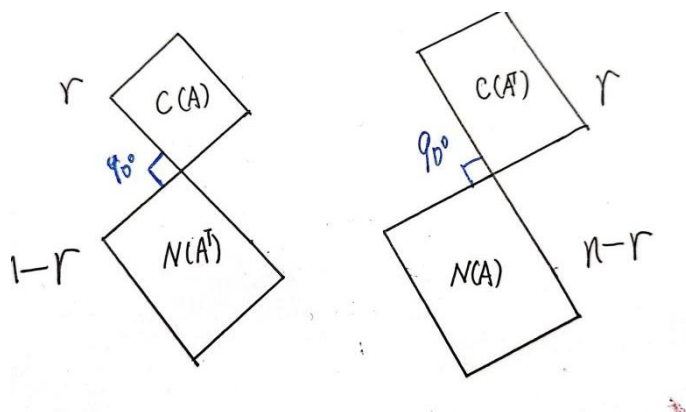
如果画在图中，将会是这样：



这也就意味着我们再次验证了这个结论：

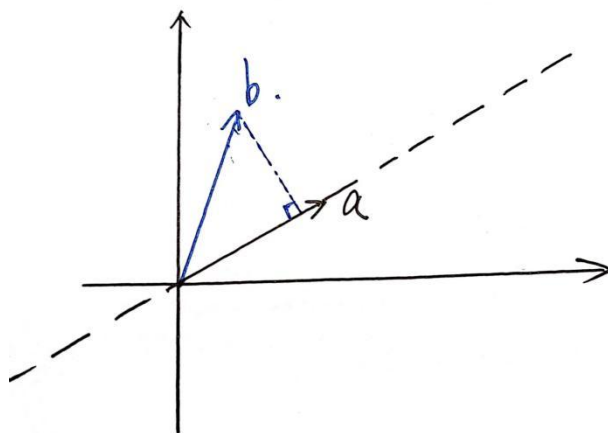
$$N(A^T) \perp C(A)$$

这些结论在上一章已经得到了证明，故这里知识提供一个直观理解，而不再继续证明。
那么我们把刚刚得到的结论用直观地几何语言描述出来，将会是这样：



Chap3.3 投影矩阵的本质

下面，我们考虑投影矩阵，在 Chap1.6 中，我们已经简单地引入了“投影”这一变换。那么下面我们继续讨论一维的情况，像这样：

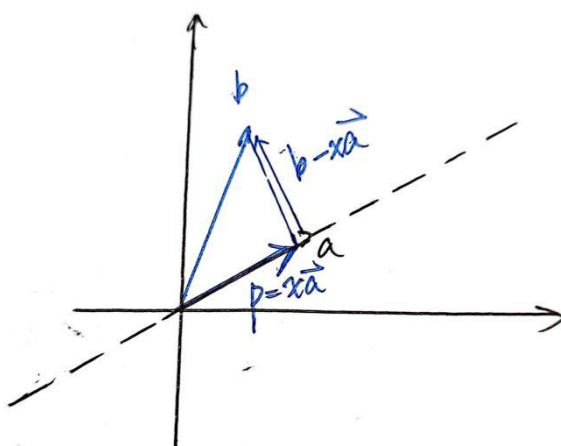


那么显然地，例如我们想要获得向量 b 在向量 a 所在直线上的投影，我们只需要进行一下操作：

$$b_{\text{投}} = \frac{1}{\sqrt{a^T a}} a^T b$$

首先后面的 $a^T b$ ，代表着 a 与 b 的点乘，将 b 投影到了 a 上，而下面的 $a^T a$ ，代表着 a 向量的长度，这样我们就获得了 a 方向的单位向量，如此就可以得到投影向量。

但是这一想法如果拓展到一条直线投影到一个平面之上就变得不那么显然，所以我们给出一个新的思路，像这样：



那么，我们设在垂直方向上的向量为：

$$e = b - xa$$

其中 xa 代表 b 的投影向量， x 表示 b 的投影向量的长度与 a 的长度的比例。那么应当有：

$$e \cdot a = 0$$

也就是，我们得到了：

$$a^T(b - xa) = 0$$

继续化简，我们有：

$$a^T b - xa^T a = 0$$

$$a^T b = xa^T a$$

所以我们就得到了 x ：

$$x = \frac{a^T b}{a^T a}$$

接着我们把 x 乘到 a 中，就得到了投影向量：

$$p = ax = a \frac{a^T b}{a^T a}$$

我们如果不这样写，而将以下这个式子记作投影矩阵：

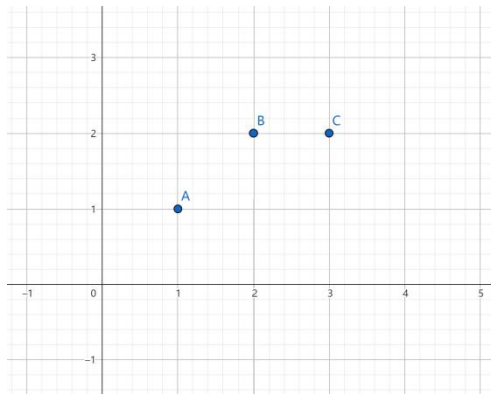
$$P = \frac{aa^T}{a^T a}$$

那么 b 在 a 所在直线上的投影将会是：

$$p = Pb$$

于是我们就获得了投影。但是这实际上除了解决一些力学点乘问题之外，也没有其他的用处。我们再来看一看三维空间内的情况。

在引入三维空间之前，我们先引入这样一个问题。例如我们在做一些科学研究，那么所得到的数据不一定是完美地符合真理，因为必然地会存在误差。比如说，我们现在有三个点： $(1,1)$ ， $(2,2)$ ， $(3,2)$ 。我需要找到一条直线，使得这条直线能够展示出这三个点所窥探出的趋势。



我们假设一条直线 $y = Cx + D$ 能够做到这件事，那么应该满足一下方程，像这样：

$$C + D = 1$$

$$2C + D = 2$$

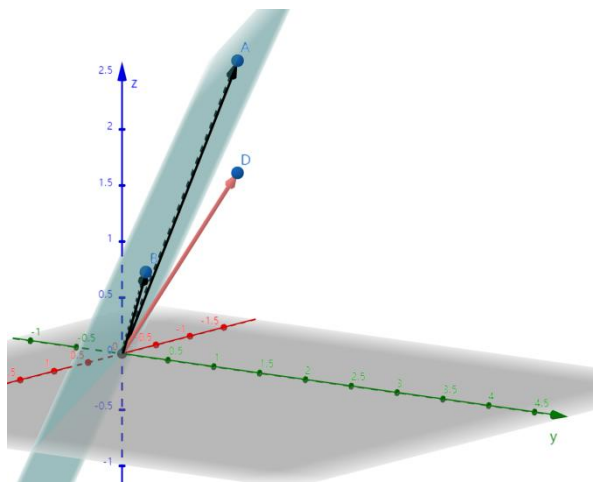
$$3C + D = 2$$

下面，如果我们把这个线性方程组写成矩阵的形式，将会是这样：

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

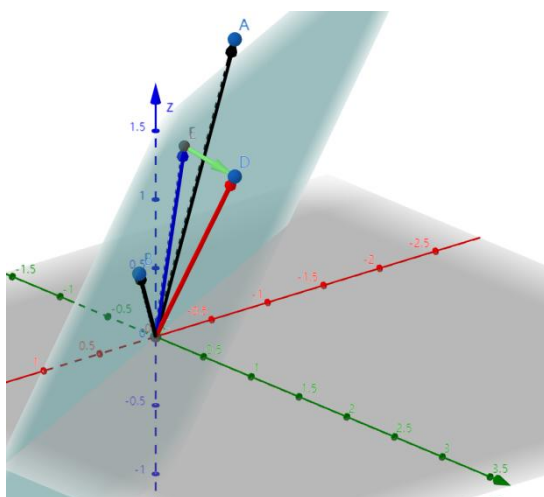
显然，这是不存在解的，因为如果我们从列空间来看，那么如果有解，应当要求向量 $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

位于 $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 所组成的列空间当中，如果我们把在三维空间中的这个列空间给画出来，那么将会是这样：



其中红色的向量，代表向量 $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ ，那么这个向量显然不属于这个列空间，自然地，也就不可能存在一个解。但是我们又必须要求有一个解，该怎么办呢？

于是，我们考虑将这个 $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ 投影到列空间所在的平面上，以寻找一个最相近的点，进而找到一个合理的解，像这样：



如果我们能够找到这个投影就可以了。下面的问题转化为，如何找到这个投影向量我们考虑下面的向量，把 OD（红色）设为 \mathbf{b} ，即我们找不到解的 $\mathbf{Ax}=\mathbf{b}$ 。把 ED，即绿色的向量，这个向量垂直于整个平面，我们把它设为 \mathbf{e} ，还有一个向量 OE（蓝色），是投影向量，我们把它设做 \mathbf{p} 。于是将会有一下表达式成立：

$$\mathbf{e} = \mathbf{b} - \mathbf{p}$$

下面，我们考虑“ \mathbf{p} ”是什么？ \mathbf{p} 位于列空间所生成的平面当中，所以这也就代表着，它将会是列空间的两个基所线性组合产生的一个向量。自然地，我们设这个线性组合的“权”为 $\hat{\mathbf{x}}$ ，那么 \mathbf{p} 将可以表达成这样：

$$\mathbf{p} = \mathbf{A}\hat{\mathbf{x}}$$

那么 \mathbf{e} 的表达式，将会是：

$$\mathbf{e} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$$

下面，由于 \mathbf{e} 垂直于整个平面，我们之前说过，研究平面的所有向量，只需要研究这个平面的基向量就可以了，于是我们让 \mathbf{e} 垂直于这个平面的两个基向量，比如我们设这两个基向量分别为 \mathbf{a}_1 ， \mathbf{a}_2 。于是，我们将有以下表达式成立：

$$\mathbf{a}_1 \cdot \mathbf{e} = 0$$

$$\mathbf{a}_2 \cdot \mathbf{e} = 0$$

于是：

$$\mathbf{a}_1(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0$$

$$\mathbf{a}_2(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0$$

但是，这么写真是太不优雅了，于是我们写作：

$$\mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0$$

于是，我们得到 $\hat{\mathbf{x}}$ ，将会是这样：

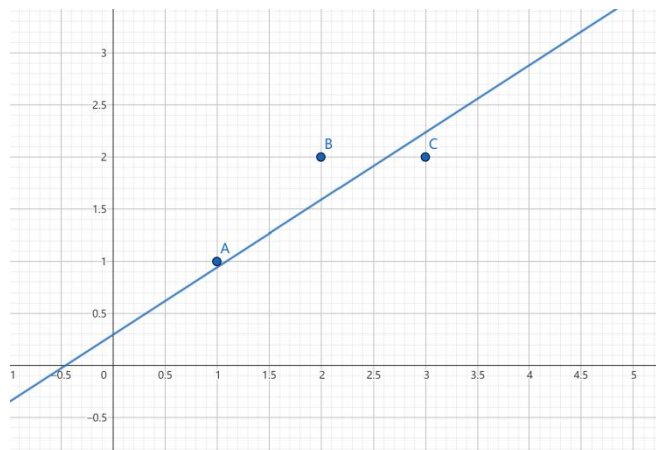
$$\mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}}$$

$$\hat{\mathbf{x}} = \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}}$$

是的，这正是我们的解，如果我们计算以下的话，将会得到解是这样的：

$$\hat{x} = \frac{\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}{\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}} = \frac{\begin{bmatrix} 11 \\ 5 \end{bmatrix}}{17} = \begin{bmatrix} \frac{11}{17} \\ \frac{5}{17} \end{bmatrix}$$

所以这正是 C 和 D 的近似解，如果我们画在图中，将会是这样：



你是否发现我刚刚计算时的问题？

首先这里，这里是两个矩阵相乘，它并得不到一个数！而是一个矩阵！

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix}$$

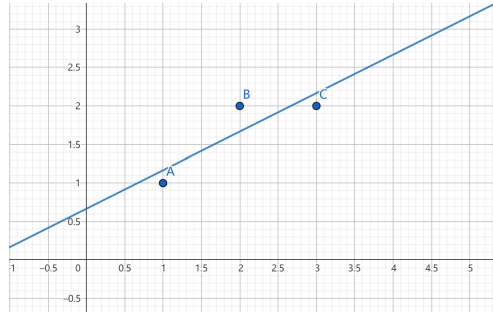
哈哈哈哈哈，所以刚才我们的计算方法是错误的，我们现在来修正。不要忘了！这里是矩阵，矩阵是没有除法的！我们只能写成逆的形式。

$$A^T b = A^T A \hat{x} \Leftrightarrow \hat{x} = (A^T A)^{-1} A^T b$$

下面，我们再进行正确一些的计算：

$$\begin{aligned} \hat{x} &= \left(\begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & -1 \\ -1 & \frac{7}{3} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{4}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} + \frac{1}{2} * 2 \\ \frac{4}{3} + \frac{1}{3} * 2 - \frac{2}{3} * 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{2}{3} \end{bmatrix} \end{aligned}$$

下面我们考虑投影矩阵在多维空间内的表示，前面，我们在二维空间当中讨论时，得到的投影向量是这样的：



是的，我们下面考虑投影向量的表达式：

$$\hat{x} = (A^T A)^{-1} A^T b$$

我们有了投影向量的权，下一步，我们考虑投影矩阵是，前面我们有表达式：

$$p = A\hat{x}$$

这个表达式代表着什么？这代表着投影向量 p 是 A 的各个列向量的线性组合。自然地，我们将 \hat{x} 代入，得到：

$$p = A(A^T A)^{-1} A^T b$$

是的，我们可以再像刚刚一样，找到投影矩阵，像这样：

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

我们将一维与多维进行对比，道理是一样的：

$$P = \frac{aa^T}{a^T a} \Leftrightarrow P = A(A^T A)^{-1} A^T$$

下面，我们讨论投影矩阵的一些美好的性质。

首先，我们考虑 P^n ，也就代表着对矩阵 P 做多次乘法运算。如果你拥有几何直观，那么你可以一眼看出这个表达式：

$$P^n = P$$

这是因为，经过一次投影后，向量将落在我们所希望的向量空间里，那么继续投影，它必然还在那里，再投影，必然还会在那里，所以我们通过肉眼就看出来了这件事情的原因。

下面我们从代数的角度进行考虑：

为了简单起见，我们考虑 P^2 ，像这样：

$$\begin{aligned} P^2 &= A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T \\ &= A(A^T A)^{-1} (A^T A(A^T A)^{-1}) A^T \\ &= A(A^T A)^{-1} I A^T \\ &= A(A^T A)^{-1} A^T \end{aligned}$$

是不是很美妙呢，乘 n 次也是同样的道理。

接下来这个性质似乎并不那么显然，我们来观察一下这个 P ，如果将它转置呢？

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

显然，由于 $\mathbf{A}^T \mathbf{A}$ ，是对称矩阵，所以它的逆也自然是对称矩阵，转置后依旧是对称矩阵。所以我们可以得到这样的一个美妙的性质：

$$\mathbf{P}^T = \mathbf{A}^{TT} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{P}$$

这就是投影矩阵的第二个性质，也就是是转置后不变，或者换句话说，我们证明了这个投影矩阵是对称矩阵。

Chap3.4 施密特正交化的本质

前面我们讨论了投影矩阵。下面我们讨论正交矩阵。

首先，正交矩阵意味着矩阵的所有列向量都是垂直的，这些矩阵的每一个列向量之间都是相互垂直的，这也就意味着矩阵的任意一列之间相互点乘都是 0。例如以下的矩阵：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

正交化矩阵还有一类，称为单位正交化矩阵，意味着每一个正交矩阵的长度都是 1 像这样：

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

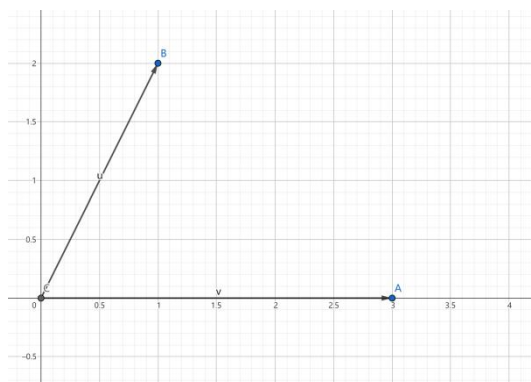
那么首先，我们先来讨论施密特正交化的性质，首先，正交矩阵的逆和转置之间存在关联，例如我们设一个正交矩阵叫做 Q ，很自然地，我们有：

$$Q^T Q = I$$

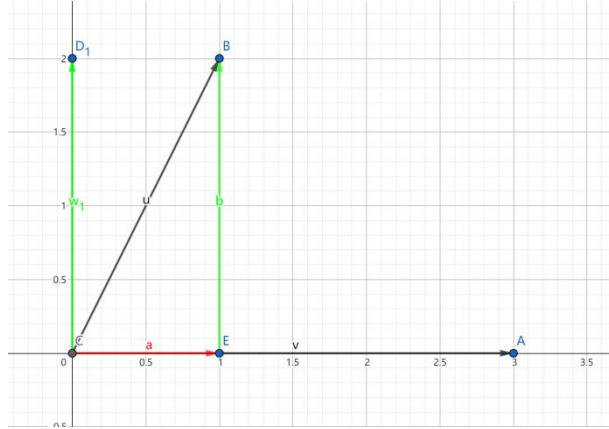
这个公式的含义在于计算正交矩阵的每一行和每一列相乘，很自然地，只有自己与自己相乘会得到 1，而与其他点乘都将得到 0。这也就是说：

$$Q^{-1} = Q^T$$

接下来我们回归正题，正交化，意味着原来并不是正交的，然后通过一种方法，让两个向量变得正交了，例如这两个向量：



我们如何才能让这两个向量正交呢？很简单，我们只需要 B 的垂线就可以了，也就像这样：



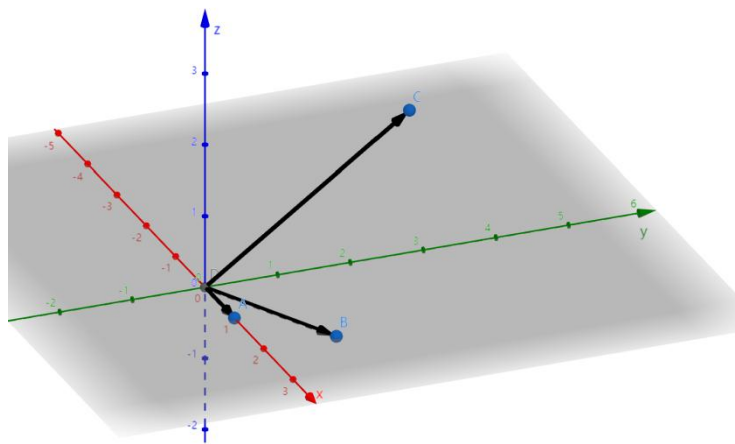
这就意味着我们只需要去寻找到绿色的向量即可。首先我们可以轻松的得到红色的向量，因为前面我们已经得到向量 b 在 a 的投影矩阵，像这样：

$$p = \frac{aa^T}{a^T a} b$$

那么垂直地矩阵将会是：

$$e = b - p = b - \frac{aa^T}{a^T a} b$$

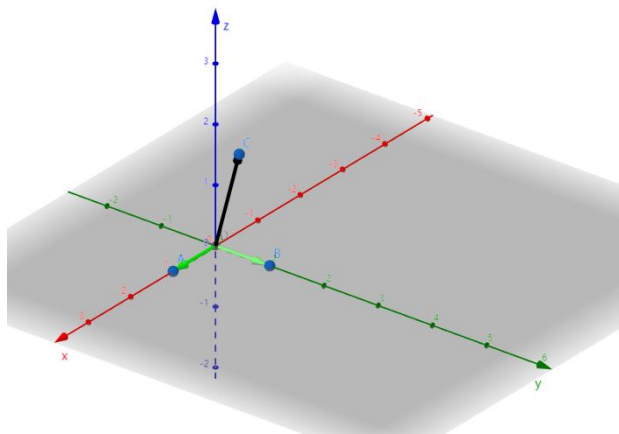
下面，我们考虑三个向量的情况，像这样：



那么如果我们有三个向量， a ， b ， c ，我们希望将这三个向量也正交化，首先我们先将 a ， b 正交化，像这样，用我们刚刚的公式，于是：

$$b_{\text{正}} = b - p = b - \frac{aa^T}{a^T a} b$$

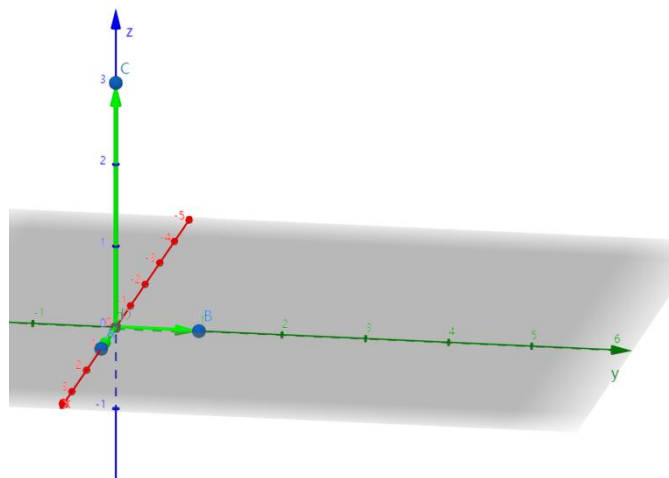
如果在图中，那么 a ， b 将会是这样：



这样吧，我们继续将 c 也正交化，自然地，我们希望它垂直于 a 和 b ，自然地，我们只需要垂直于 a, b 的向量，而不需要 a, b 方向的向量，于是我们考虑减去向量 c 在向量 a, b 方向的分量，从而剩下的所有分量都应当是垂直于 a 和 b 的，像这样：

$$c_{\text{正}} = c - p_a - p_b = c - \frac{aa^T}{a^T a} c - \frac{bb^T}{b^T b} c$$

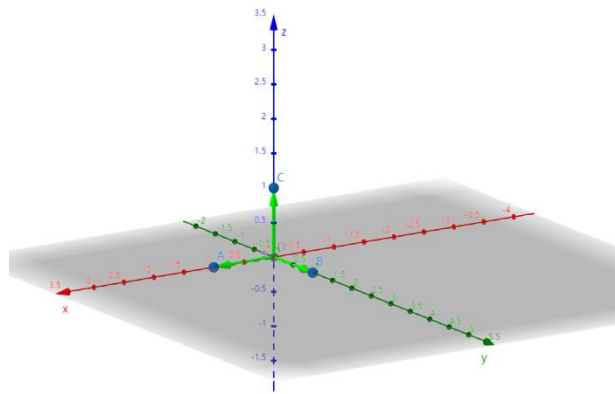
如果在几何中，将会是这样：



下面，为了让它们三个基向量变成标准的正交矩阵，那么将会是这样，我们将每一个列向量除以各自的长度，像这样：

$$a = \frac{a}{\sqrt{a^T a}}, b = \frac{b}{\sqrt{b^T b}}, c = \frac{c}{\sqrt{c^T c}}$$

于是，我们再进行一步变换，将会是这样：



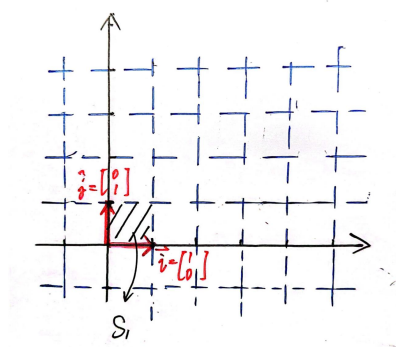
请记住：正交化的本质在于换一组空间的基，这样就更方便地表示空间了。

Chap3.5 行列式的本质

前面我们讨论了矩阵，每一个矩阵都有一个美妙的行列式。但我们先不给出行列式的表达式。而首先给出其诸多性质，进而推理出其表达式。

但事实上，我不想给出性质，因为这些性质也是“凭空产生”的，并不能达到最自然的程度。我们首先从几何角度，给出行列式的定义。

还记得我们很久很久以前讲的线性空间吗？例如，我们先给出一个最简单的线性空间，像是这样：



这是一个非常简单的线性空间，其对应的基向量分别是：

$$i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

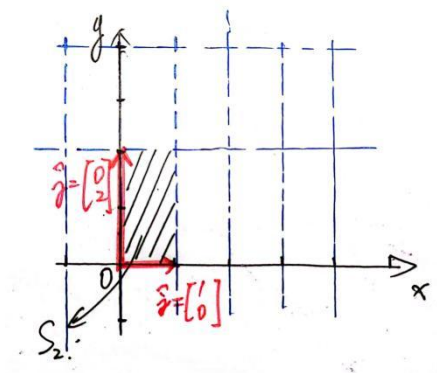
于是，这两个基向量所对应的矩阵就是单位矩阵，像这样：

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

我们观察到图中的面积 S_1 ，这就是行列式的值，即两个基向量所围成的平行四边形的面积。我们将它记作：

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

下面我们稍微复杂一点的线性空间，比如，我们可以向上拉伸空间。



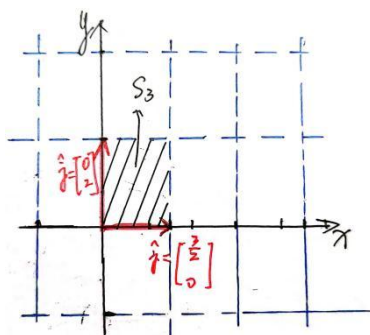
从代数上，这个线性变换所对应的矩阵将会是这样：

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

那么这个矩阵对应的行列式的值将会是图中两个基向量所围成的面积，像这样：

$$\begin{vmatrix} 1 & 0 \\ 0 & 2 \end{vmatrix} = 2$$

那么下面我们在尽显拉伸，像这样，沿着水平方向拉伸整个线性空间，那么空间将会变成这样：



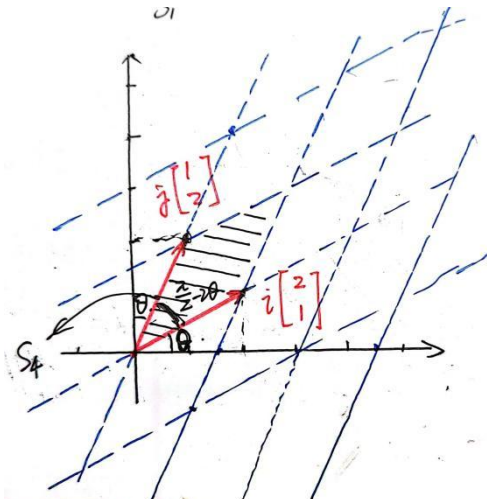
在代数上，意味着由 $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ，进行了一下矩阵的变化：

$$\begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

那么自然地，这个矩阵的行列式将会是：

$$\begin{vmatrix} \frac{3}{2} & 0 \\ 0 & 2 \end{vmatrix} = 3$$

于是我们接着考虑复杂一些的内容，如果经过了这样的一个变换：



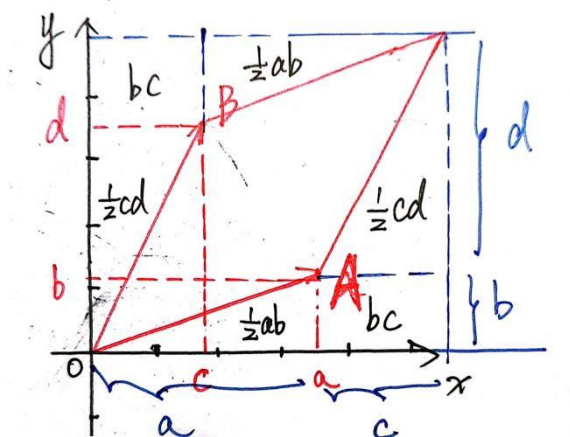
那么这个变换经历的矩阵就是：

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

如果我们要计算这个线性空间的行列式，首先可以采用三角函数，这很自然，但是也很愚蠢：

$$\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = \sqrt{5}\sqrt{5} \cos\left(\frac{\pi}{2} - 2\theta\right) = 5 \cdot \frac{3}{5} = 3$$

下面给出线性代数的做法：



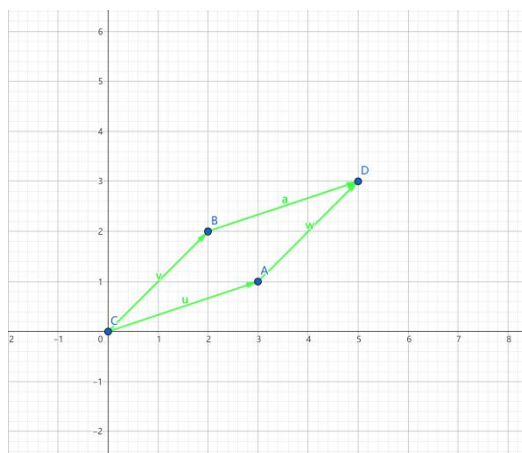
其计算的基本思想是割补：

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = (a+c)(b+d) - 2bc - ab - cd = ad - bc$$

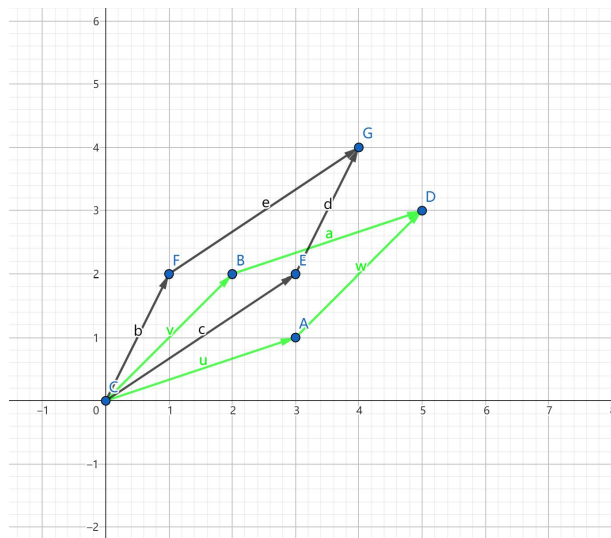
前面，我们一直都在讨论列空间，那么行空间呢？

这二者的行列式是否相等呢？

例如矩阵 $\begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$ ，其列向量形成的线性空间将会是这样：



那么它的行向量呢？如果用红笔在图中画出来将会是这样：

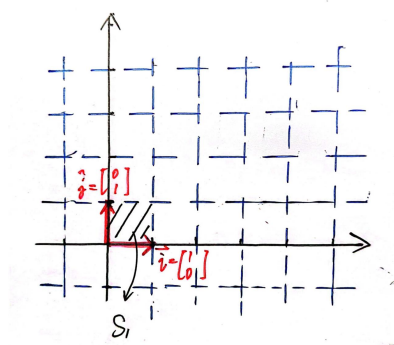


好吧，诚实的说，似乎看不出任何东西。

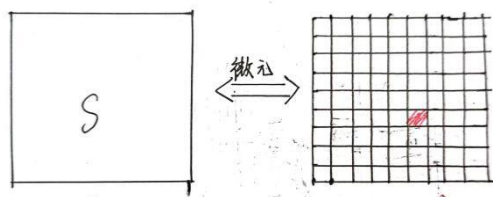
当没有思路的时候，我们就需要回到线性代数的本质——线性。

我们再次思考行列式的意义，它代表着基向量组成平行四边形的面积。这似乎也是肤浅的理解，如果我们这样做呢？矩阵的本质是对线性空间进行变换，那么我们来观察一下再变化过程中的每一个“小方块”的变化，这是微积分的思想，我称这些小正方形为“面积元”。

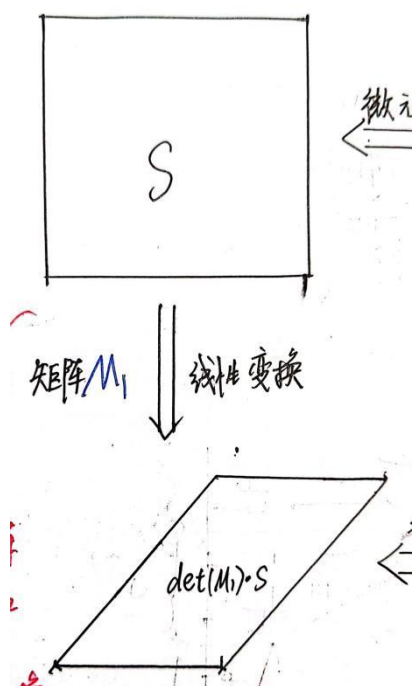
对于一个没有变换的单位矩阵所对应的基向量所张成的线性空间，将会是这样的：



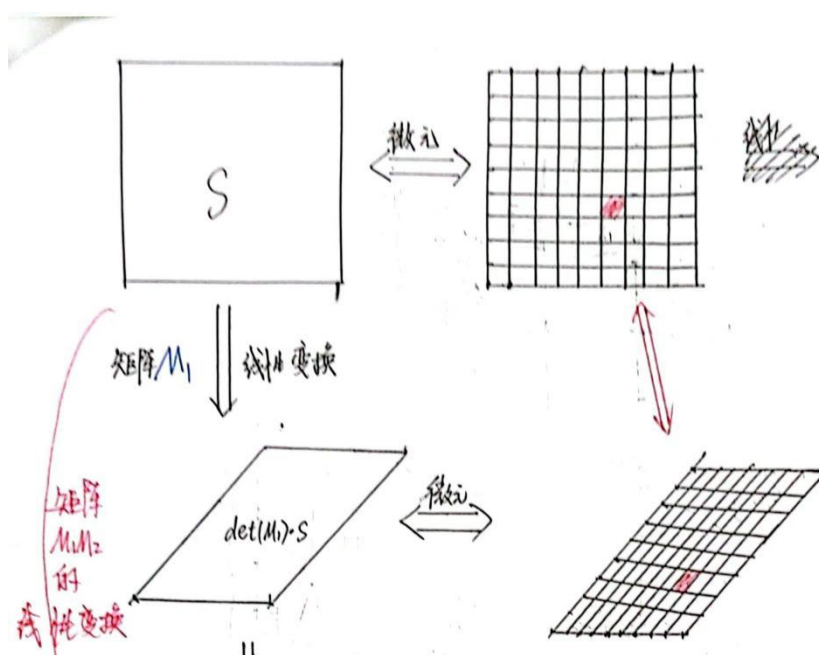
我们考虑对他进行某一个矩阵所对应的线性变换，我们从微观的角度进行观察，进而从几何上认识这个美妙的结论。看下面的例子，我们将这个原始的面积进行微元：



假若我们对这个空间进行一个线性变换，即对应了一个矩阵，像这样：



那么，如果我们将视角放在微观，那么对于每一个面积来说，将会是下面这样的变换：

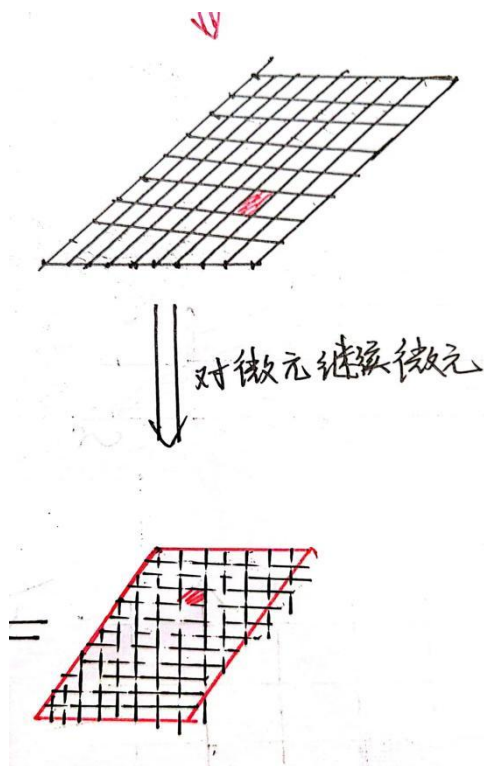


每一个小正方形线性变换后，也将会进行同样的线性变换，也就是说，每一个小的面积元都会进行相同线性变换。

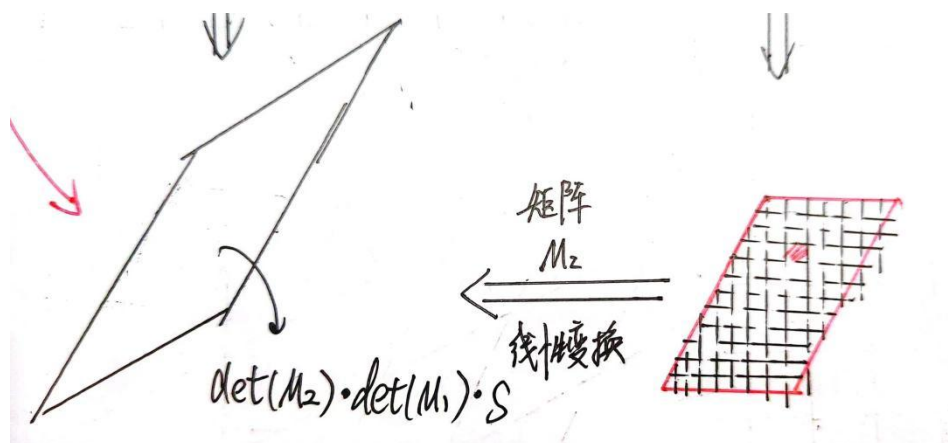
$$S_{M1} = \sum dS_{M1} = \sum \det(M_1) dS = \det(M_1) S$$

也就代表着，行列式代表着将面积进行变化，对于空间中的每一个面积都将变化一个相同的比例，这个比例就是这个矩阵所代表的行列式。

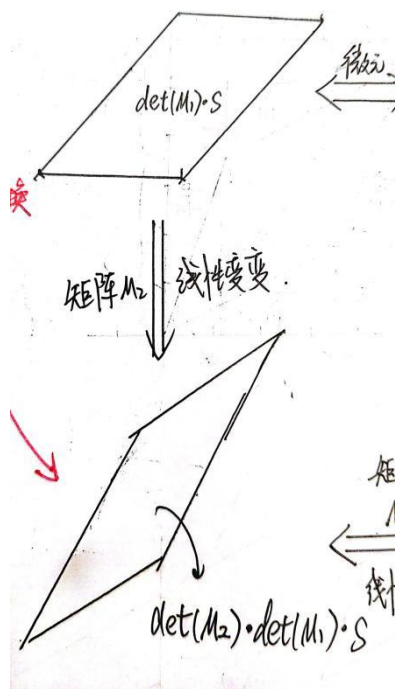
接下来，我们继续对一个平行四边形的面积进行微元，同样地，进行线性变换。



下面我们考虑将这个已经进行过线性变换的空间再次进行一次线性变换，对应的矩阵是 M_2 ，也就意味着将这个 S_{M_1} ，进行相同的线性变换，也就意味着将这个 S_{M_1} 空间进行切分成一个一个的小正方形，然后同样地进行 M_2 变换。



这等效为将上面的整个大的平行四边形进行 M_2 的线性变换，像这样：

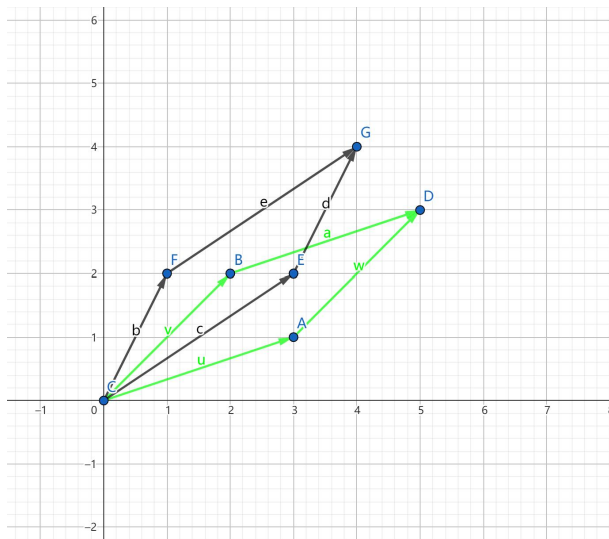


换句话说，我们将有以下这件事情成立：

$$\det(M_1 M_2) = \det(M_1) \cdot \det(M_2)$$

哦哦哦，别让我们沉溺在微积分思想的震撼中了，回到刚刚哪个问题——行空间的行列式有什么关系呢？

矩阵 $\begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$ ，其列向量形成的线性空间将会是这样：

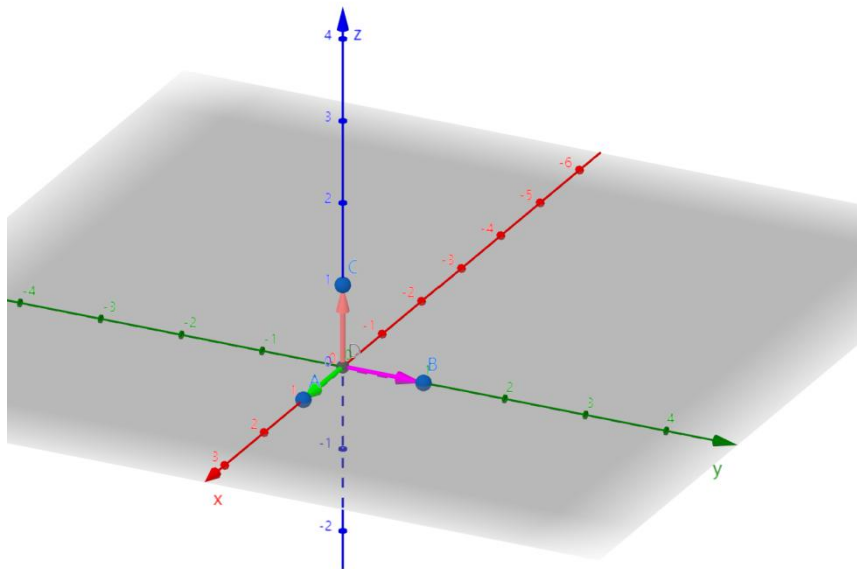


我们如何理解“行空间”，其实就是转置的列空间而已，我们下面寻找转置矩阵：

$$\text{row space} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

而这里，由于转置矩阵的行列式显然是 1，所以行空间和列空间的行列式是相等的（这里有些不严谨，因为事实上是-1。）

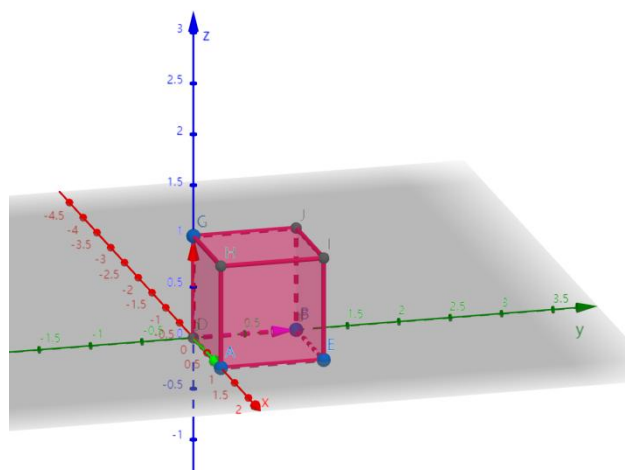
下面我们将思路拓展到三维空间，行列式将会是什么样子呢？应当也是基向量所组成的某个量，在二维空间中是面积，自然地，在三维空间中，将会是体积，例如，最简单的，像这样：



这三个基向量对应的矩阵将会是这样：

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

那么它的行列式，将会是空间当中的这个立方体的体积：



于是我们可以一眼看出来，这个行列式的值将会是 1，像这样：

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

我们接着，从行列式的几何性质开始，总结出行列式所对应的代数性质。这样我们从性质入手，推导出行列式的定义，似乎就可以通过这种手段，窥探出四维空间中的“某种体积”。

首先，**第一条性质**：

单位矩阵的行列式为 1

$$|I| = 1$$

这很自然，因为所有的基向量是单位矩阵所决定的体积或者面积都是 1。

其次，给出**第二条性质**：

如果有两列相同，那么行列式为 0

这也很自然，因为假设有两点是共线的，那么不可能存在面积或者体积之类的。

接着，我们该处**第三条性质**：

如果一行都乘以一个 t，则 t 可以提出来

$$\begin{vmatrix} ta & tc \\ b & d \end{vmatrix} = t \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

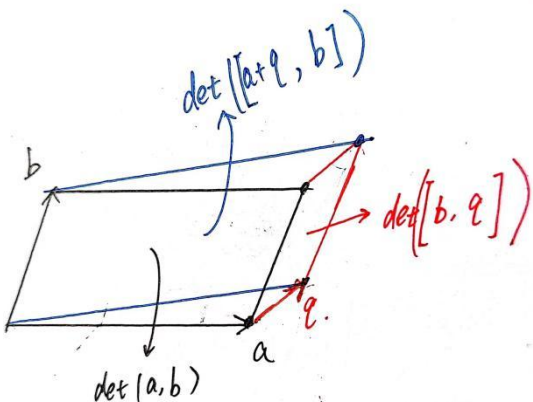
这也很自然，因为我们可以用行空间中的行向量来思考。把一个行向量增长 t 倍，自然地面积也会扩大 t 倍，体积也会扩大 t 倍。

并且，我们从行空间的角度还可以给出**第四条性质**：

行向量的线性可以表现在行列式上

$$\begin{vmatrix} a+a' & c+c' \\ b & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} + \begin{vmatrix} a' & c' \\ b & d \end{vmatrix}$$

这一性质可以从几何中看出来：



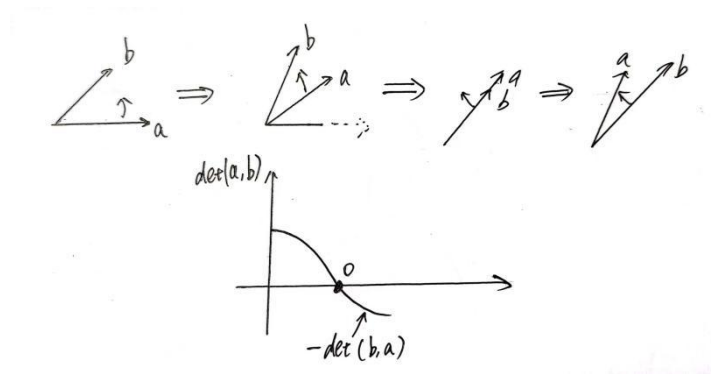
在这幅图中，我们依旧考虑行向量，其中向量 q 为 $\begin{bmatrix} a' \\ b' \end{bmatrix}$ 。

接下来，我们考虑最后一条性质，或许你会觉得这个性质有些不太自然，但实际上，这是很自然的，**第五条性质**：

列向量对调，行列式反号

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = - \begin{vmatrix} c & a \\ d & b \end{vmatrix}$$

我们依旧从几何看出来：



下面我们对上面的几个性质进行总结，并且，我们可以用同样地思路，将结论拓展到行向量和列向量都满足的情况！因为行列式对行和列的向量都满足这些关系！

第一条性质：

单位矩阵的行列式为 1

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

第二条性质：

如果有两列相同，那么行列式为 0

$$\begin{vmatrix} a & a \\ b & b \end{vmatrix} = 0$$

如果有两行相同，那么行列式为 0

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0$$

第三条性质：

如果一行都乘以一个 t，则 t 可以提出来

$$\begin{vmatrix} ta & tc \\ b & d \end{vmatrix} = t \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

如果每一列都乘一个 t，则 t 也可以提出来

$$\begin{vmatrix} ta & c \\ tb & d \end{vmatrix} = t \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

第四条性质：

行向量的线性可以表现在行列式上

$$\begin{vmatrix} a+a' & c+c' \\ b & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} + \begin{vmatrix} a' & c' \\ b & d \end{vmatrix}$$

列向量的线性可以表现在行列式上

$$\begin{vmatrix} a+a' & c \\ b+b' & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} + \begin{vmatrix} a' & c \\ b' & d \end{vmatrix}$$

第五条性质：

列向量对调，行列式反号

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = - \begin{vmatrix} c & a \\ d & b \end{vmatrix}$$

行向量对调，行列式反号

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = - \begin{vmatrix} b & d \\ a & c \end{vmatrix}$$

下面，我们考虑用这五条性质，推出来一些性质。

我们前面发现，对于行向量而言，有加法，有数量乘法的性质，那么它满不满足“初等行变换”呢？

我们来尝试一下，如果我们让第二行减去 l 倍的第一行，像这样：

$$\begin{vmatrix} a & c \\ b-la & d-lc \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} - l \begin{vmatrix} a & c \\ a & c \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

那么我们就得到了第六条性质：

初等行变换中的（行之间的加减）不改变行列式

$$\begin{vmatrix} a & c \\ b-la & d-lc \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

接下来，我们考虑拓展性质六，如果一个行列式，经过初等行变换后，他将会得到对角线上全部是主元。注意，行交换会改变行列式的符号；行也不能自己进行除法操作，只有行之间的加减是保持行列式保持不变的。

那么，最终行列式进行行之间的加减后，将会变成这样：

$$\begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix} = abc \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = abc$$

于是，我们获得了第七条性质：

行列式的值，等于主元相乘

$$|A| \sim \begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix} = abc$$

下面我们继续推导。引出**第八条性质**：

如果有一行全部是 0，那么行列式为 0

$$|A| \sim \begin{vmatrix} a & c & e \\ b & d & f \\ 0 & 0 & 0 \end{vmatrix} = 0$$

这是很自然的，因为这也就意味着这三个列向量都将处于一个平面之内。如果你不满足这草率的几何描述，我们从代数依旧可以解答你的疑惑——我们注意到性质三，提到，如果一行全部是 t 的倍数，那么 t 就可以提出来，所以自然地，我们有：

$$|A| = \begin{vmatrix} a & c & e \\ b & d & f \\ 0g & 0h & 0I \end{vmatrix} = 0 \quad \begin{vmatrix} a & c & e \\ b & d & f \\ g & h & I \end{vmatrix} = 0$$

接下来，我们综合以上的八条性质，得到一条新的性质，**第九条性质：**

如果行列式有奇异矩阵，则行列式为 0，否则不为 0

最终我们通过以上的九条性质，已经可以得出行列式的表达式了，我们下一章再继续讨论。

前面我们讨论了行列式的诸多性质，接下来我们通过性质来讨论行列式的计算公式。

先从二阶行列式进行推导，例如任意的一个二阶行列式，那么：

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{vmatrix} = ad-bc$$

同理，我们可以考虑三阶行列式，那么，如果我们再去消元是会吐血的，因为你会迷失在数字森林当中，而丧失对行列式性质的真正理解。

下面，我们给出

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{vmatrix} = ad-bc$$

下面我们给出另一种思路，也就是这样，我们利用行列式的性质来求解：

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = ad-bc$$

是的我们可以考虑将一个行列式进行拆分，进而找到它的值。接下来我们考虑三维的情况，由于用电脑打字打矩阵我会吐血，所以我打算采用手写版：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} \\ + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\ = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13}$$

我们总结一下：

首先，我们有了二阶行列式的计算公式，像这样：

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = ad-bc$$

并且，我们也有了三阶行列式的计算公式，像这样：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} \\ + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\ = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

可见，当行列式的 n 增大时，行列式的复杂程度也增大了许多。二阶行列式有 2 项，三阶行列式有 6 项，如果大胆猜想的话， n 阶行列式将会有 $n!$ 项。事实的确如此，因为如果我们要将一个行列式像上面分解的话。第一行有 n 个选择，第二行有 $n-1$ 个选择，.....最后一行只剩下一个选择，这样才能让分解出来的行列式都不是 0，并且还可以组合成原来的行列式。

下面，我们先来窥探一下 n 阶行列式的样子，首先和上面一样，它应当是从第一行选择一个，从第二行选择一个.....从第 n 行选择一个，并且有可能是正的，有可能是负的，所以我们先姑且认为，其表达式长成了这样：

$$\det(n \times n) = \sum \pm a_{1\alpha} a_{2\beta} \dots a_{n\eta}$$

那么 \pm 代表着这个式子可能是正的，也可能是负的。后面的 α, β, \dots 代表着从 1 到 n 的一个排列。接下来先考虑正负号的问题。

首先，究竟是什么决定了正负号呢？我们用四阶行列式来举个例子，也就意味着这四项的后面的数字将会是 1,2,3,4 的一个排列。如果我们排列时 1,2,3,4，也就是这一项是 $a_{11}a_{22}a_{33}a_{44}$ ，那么这一项必然是正的，因为这个拆解出来的行列式不需要任何的行变换，

所以也不存在正负号的改变。再比如如果是 (4,3,2,1)，也就意味着这一项是 $a_{14}a_{23}a_{32}a_{41}$ 。

那么也必然地是正的，因为我只需要将“4”和“1”交换，“2”和“3”交换，交换两次，就可以得到正确的对角线的排列，即 (1,2,3,4)，相当于这个过程：

$$\begin{vmatrix} 0 & 0 & 0 & a_{14} \\ 0 & 0 & a_{23} & 0 \\ 0 & a_{32} & 0 & 0 \\ a_{41} & 0 & 0 & 0 \end{vmatrix} \rightarrow \begin{vmatrix} a_{41} & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 \\ 0 & 0 & a_{23} & 0 \\ 0 & 0 & 0 & a_{14} \end{vmatrix}$$

下面，我们从三阶行列式出发，引入一个概念——代数余子式。

首先三阶行列式的展开在这里，我们观察这一项：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} \\
 + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\
 = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

如果我们对它进行代数化简，将会是这样：

$$\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & 0 & a_{32} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{23} & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} \\
 = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) \\
 = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$

同理，我们发现其他的项也可以从三阶行列式变成二阶行列式，像这样：

$$\begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} = \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{31} \end{vmatrix} = -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{31} \end{vmatrix} \\
 \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} = a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

下面，我们总结一下，一个三阶行列式可以拆成下面这样：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12} \left(- \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \right) + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

其中，这些二阶行列式就是“代数余子式”，有正，有负，这是为啥？

如果我们从更高的眼光去看待这个问题，也就是回到行列式的性质：

如果我们已经认为这个式子，将会得到正的余子式，那么进行一次替换，将会变成负的，像这样：

$$\begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{12} & 0 & 0 \\ 0 & a_{21} & a_{23} \\ 0 & a_{31} & a_{33} \end{vmatrix} = a_{12} \left(- \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \right)$$

同理，对于其他的也是一样的，所以在三维空间内，代数余子式对应的正负号是这样的：

$$\begin{vmatrix} + & - & + \\ * & * & * \\ * & * & * \end{vmatrix}$$

如果我们不从第一行进行代数余子式的展开，而是从第二行开始进行代数余子式的展开，那么也就意味着进行了一次交换，进行交换就意味着产生了一个负号。

$$\begin{vmatrix} + & - & + \\ - & + & - \\ * & * & * \end{vmatrix}$$

同理，如果我们是四节行列式，我们把整个的行列式都写出来将会是这样：

$$\begin{vmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{vmatrix}$$

我们可以用代数余子式就可以构造出来递推，进而求解 n 阶行列式。

Chap3.6 克拉默法则

前面，我们讨论了行列式的几何性质和代数性质，其本质基于列空间和行空间。这一节，我们考虑行列式在线性代数中的具体应用。

(1) 行列式与逆矩阵的关系

我们之前求解逆矩阵的方法是高斯-若尔当方法。下面给出行列式的计算方法：

如果是二阶行列式，那么它的逆矩阵将会是这样：

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}^{-1} = \frac{1}{\begin{vmatrix} a & c \\ b & d \end{vmatrix}} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

我们先从二阶行列式窥探出一些规律。

首先，我们如何去看待右侧的这个奇怪的矩阵？我们可以将 d 看作是 a 所对应的代数余子式； a 看作是 d 对应的代数余子式；但是如果这样看的话，“ $-c$ ”似乎不满足这一规律，没关系，我们可以对这个矩阵进行一次转置，像这样：

$$\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}^T = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

于是那么我们就可以将“ $-b$ ”理解成 c 对应的代数余子式；“ $-c$ ”看作是 b 对应的代数余子式。如果这样看的话，这一结论是必然成立的。因为如果我们将这个逆矩阵和原来的矩阵相乘，将会得到：

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} = \det(A) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

那么，如果是 n 阶是否还成立呢？我们用三阶行列式来说明这件事情是绝对成立的。像这样：

$$\begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{12} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} =$$

首先，我们可以很轻易地看出对角线上的元素都是 $\det(A)$ ，像这样：

$$\begin{bmatrix} \det(A) & 0 & 0 \\ 0 & \det(A) & 0 \\ 0 & 0 & \det(A) \end{bmatrix}$$

那么其他的元素是否都是 0 呢？当然是的，因为我们举这样一个例子，将这一行和这一列相乘，是否能得到 0 呢？

$$\begin{vmatrix}
 |a_{22} & a_{23}| & -|a_{12} & a_{13}| & |a_{12} & a_{13}| \\
 |a_{32} & a_{33}| & -|a_{32} & a_{33}| & -|a_{22} & a_{23}| \\
 -|a_{22} & a_{23}| & |a_{11} & a_{13}| & -|a_{11} & a_{13}| \\
 -|a_{32} & a_{33}| & -|a_{31} & a_{33}| & -|a_{21} & a_{23}| \\
 |a_{21} & a_{22}| & -|a_{11} & a_{12}| & |a_{11} & a_{12}| \\
 |a_{31} & a_{32}| & -|a_{31} & a_{32}| & -|a_{21} & a_{22}|
 \end{vmatrix}
 \begin{bmatrix}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
 \end{bmatrix} =$$

显然，是的，因为这些代数余子式可以再组成一个行列式：

$$\begin{aligned}
 & a_{12} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{22} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{32} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\
 & = \begin{vmatrix} a_{12} & a_{12} & a_{13} \\ a_{22} & a_{22} & a_{23} \\ a_{32} & a_{32} & a_{33} \end{vmatrix} = 0
 \end{aligned}$$

可以证明除了对角线以外的元素，其相乘可以对应一个行列式，而这个行列式当中有两个行向量是共线的，这也就表明这个行列式将会得到 0。

于是，我们便证明了这个求解逆矩阵的公式，我们设矩阵每一个元素对应的代数余子式所产生的矩阵称为伴随矩阵 C ，那么逆矩阵可以写成这样：

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

下面我们讨论行列式的另一个重要作用——克拉默法则

(2) 克拉默法则

首先，我们从代数角度先认识克拉默法则。

对于一个线性方程组，我们总可以将它表示成下面的样子：

$$Ax = b$$

这是一个矩阵形式，其中 x 是向量，包含了所有的解。如果我们要把 x 求出来，只需要求解 A 的逆矩阵即可，像这样：

$$x = A^{-1}b$$

如果和上面的那个求解逆矩阵的公式联系起来，将会是这样：

$$x = \frac{1}{\det(A)} C^T b$$

那么分子上是 A 的行列式，很漂亮且美观。但是分子上却显得有些吃力，我们依旧举特殊的例子来帮助我们窥探答案。

如果是二阶，那么伴随矩阵将会是这样：

$$\begin{matrix} A & & C \\ \left| \begin{array}{cc} a & b \\ c & d \end{array} \right| & \longrightarrow & \left| \begin{array}{cc} d & -b \\ -c & +a \end{array} \right| \longrightarrow \end{matrix}$$

如果将伴随矩阵转置，像这样：

$$\begin{array}{c} C \\ \left| \begin{array}{cc} d & -c \\ -b & +a \end{array} \right| \end{array} \longrightarrow \begin{array}{c} C^T \\ \left| \begin{array}{cc} d & -b \\ -c & a \end{array} \right| \end{array}$$

并且得到线性方程组的解，将会是这样：

$$x_1 = \frac{\begin{vmatrix} b_1 & b \\ b_2 & d \end{vmatrix}}{\det(A)}, x_2 = \frac{\begin{vmatrix} a & b_1 \\ c & b_2 \end{vmatrix}}{\det(A)}$$

如果是三阶，那么伴随矩阵将会是这样：

$$\begin{array}{c} A \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \end{array} \longrightarrow \begin{array}{c} C \\ \left[\begin{array}{ccc} \begin{vmatrix} a_{32} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ - \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{array} \right] \end{array}$$

如果将伴随矩阵翻转过来，那么将会是这样：

$$C^T \left[\begin{array}{ccc} \begin{vmatrix} a_{32} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ - \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{array} \right] \xrightarrow{\vec{b}}$$

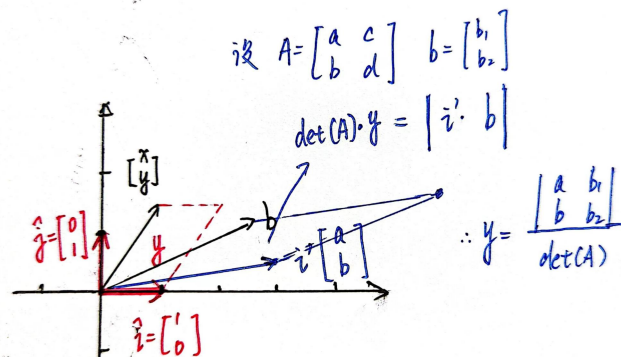
如果我们将 \vec{b} 乘进去，将会是这样：

$$= \begin{bmatrix} \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix} \\ \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix} \\ \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix} \end{bmatrix}$$

如果我们综合以上的讨论，可以求得求解线性方程组的方程，像这样：

$$x_i = \frac{\det(A'_i)}{\det(A)} \quad (A'_i \text{ 表示第 } i \text{ 列被 } b \text{ 替换})$$

下面，我们考虑克拉默法则的几何含义。

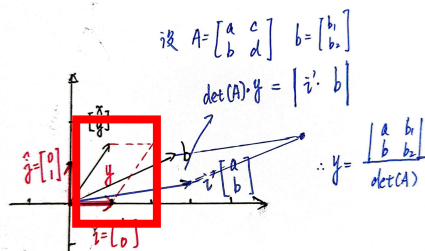


首先，什么是“解方程”？例如下面的这个方程组，我们一般写作 $Ax=b$ 。

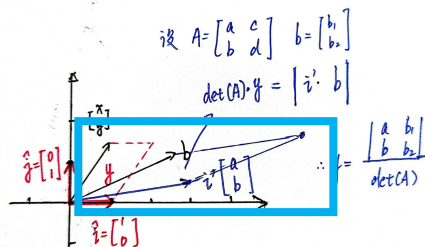
$$Ax = b$$

这代表着向量 x ，经过 A 这个矩阵的线性变换后与 b 重合。我们如果有了这个视角，就能一步一步推理出克拉默法则的几何含义了。

几何含义其实很直观，首先我们关注红色部分的 y ，这里：



这代表着还没有经历线性变换的 x 与最初的基向量 $(1,0)$ 所围成的平行四边形的面积，即“ y ”。下面我们考虑进行矩阵 A 的线性变换，将 x 变换到与 b 重合，那么也就变成了下面这样（蓝圈圈住的部分）：



那么也就意味着 y 这个面积变成了右侧的这个平行四边形的面积，我们设其为 $area$ 。

那么由于矩阵的线性变换将会把空间中的面积扩大行列式倍，也就意味着有以下表达式成立：

$$area = \det(A)y$$

另外的，假设基向量有 $(1,0)$ ，变换到了 (a, b) ，那么我们可以通过另一个行列式再次求解 area 的面积，像这样：

$$\text{area} = \begin{vmatrix} a & b_1 \\ b & b_2 \end{vmatrix}$$

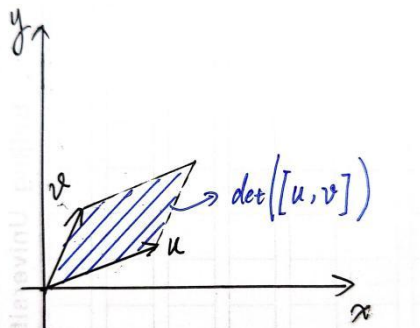
于是，我们就有了以下表达式成立：

$$y = \frac{\begin{vmatrix} a & b_1 \\ b & b_2 \end{vmatrix}}{\det(A)}$$

这就是克拉默法则的几何解释，仍旧回归于线性代数的核心思想——线性变换。

Chap3.7 叉乘的本质

前面，我们讨论了行列式，行列式有几何含义，对于二维空间而言，叉乘显而易见地应该定义成这样，从行列式的几何意义我们可以清楚地看出，这代表着两个向量的平行四边形所围成的面积，像这样：



于是，在二维空间当中，我们把向量的叉乘记作：

$$u \times v$$

下面，我们考虑在三维空间内的情况。三维空间内，两个向量的叉乘我们依旧认为是这两个向量所围成的面积，那么应该如何去计算呢？

首先我们无法计算这种矩阵的行列式

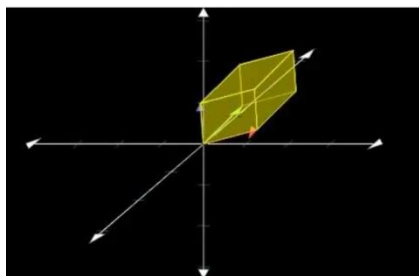
$$\begin{bmatrix} u_1 & v_1 \\ u_1 & v_2 \\ u_1 & v_3 \end{bmatrix}$$

因为行列式在三维空间在三维空间内只能表示体积，两个向量不可能构成一个体积。难道我们要将这两个向量进行“解析几何”的某种暴力求解方法，然后给出一个公式，之后背下来吗？怎么可能？你阳哥还没有蠢到这个地步。

我们想要利用行列式的性质，三维空间内，行列式的性质很简单，就是体积，所以我们构造这样的一个函数，其与前面的两个向量，组合成一个平行六面体的体积，像这样：

$$f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{vmatrix} x & u_1 & v_1 \\ y & u_1 & v_2 \\ z & u_1 & v_3 \end{vmatrix}$$

如果在图中，将会是这样：



每当我们改变输入向量 (x, y, z) 时，都会对应三维空间内的平行六面体的一个体积。这有什么用呢？我们的目的是寻找一个垂直与 u, v 的向量，并且还能够使得这个向量的模是 u, v 所形成的平行四边形的面积。讨论这个“体积”有什么用呢？

准备好，最精彩的部分要来了。我们把这个行列式展开，看一看会发生什么，像这样：

$$f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{vmatrix} x & u_1 & v_1 \\ y & u_1 & v_2 \\ z & u_1 & v_3 \end{vmatrix} = x \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} + y \left(-\begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix}\right) + z \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix}$$

我们考虑上面这个表达式的含义？我们能否看出一些我们之前学习过的东西？没错！点乘！看出来了吗？

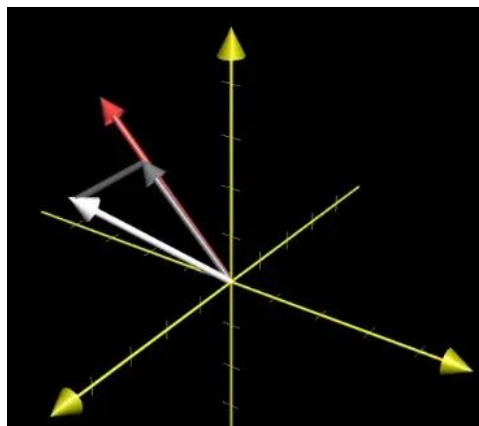
这个奇怪的函数，接受一个向量 (x, y, z) 然后输出一大坨，其实这个函数的本质就是一个点乘，我们设这个向量为 \mathbf{p} ，那么它的坐标将会是这样：

$$\begin{bmatrix} \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} \\ \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} \\ \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \end{bmatrix}$$

那么这个函数的点乘将会是这样：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} \\ \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} \\ \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \end{bmatrix} = \mathbf{u}, \mathbf{v}, \mathbf{x} \text{ 所围成平行六面体的体积}$$

你能理解这代表着什么吗？这代表着任何一个向量，和 \mathbf{p} 向量点乘，都将得到体积，这不正是我们想要找的向量吗？看下面这幅图：



红色的部分就是我们的 \mathbf{p} 向量， (x, y, z) 向着它投影，就代表这个平行六面体的高，红色向量的长度就代表着 \mathbf{u}, \mathbf{v} 所围成平行四边形的面积。

我们考虑，如果定义 \mathbf{u}, \mathbf{v} 叉乘可以定义成这样：

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

那么也就代表着我们可以写成这样：

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

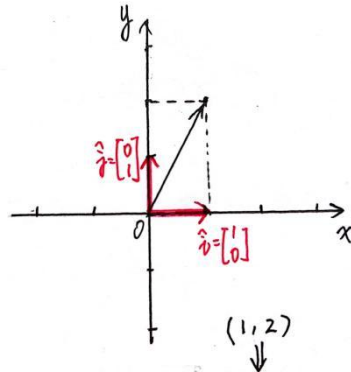
这也就是我们最后的叉乘公式的来历，像这样：

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

你是否已经要为这么美妙的公式落泪了呢？我已经感动哭了。

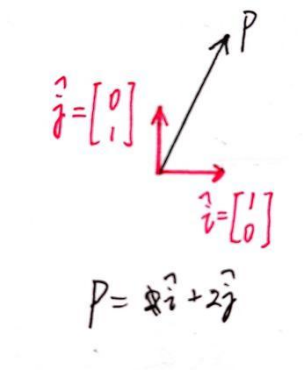
Chap3.8 基变换

下面，我们考虑基的变换。前面在矩阵乘法的时候，我们提到运用矩阵来操纵空间，但是无一例外，都是基于平面直角坐标系，例如我们之前的所有向量，都可以用坐标表示，像这样：



那么，我们可以把黑色的向量表示成向量 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 。

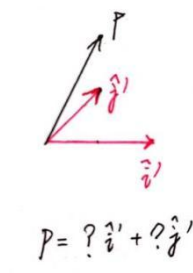
但是正如我之前所说的——向量是自由的。也就意味着向量并不依赖于坐标系，例如我们可以将整个坐标系去除掉，来讨论图中黑色向量的位置，像这样：



那么就可以写作： $1\hat{i} + 2\hat{j}$ 。

我们可以说这个基向量所代表的坐标系是“原始坐标”，在这个原始坐标中，由于向量 p 是 $1\hat{i} + 2\hat{j}$ ，那么我们可以记作，在原始坐标中，坐标为 $(1,2)$ 。

下面，我们保持 p 不动，依旧是这个向量，但是换一组基，像这样：

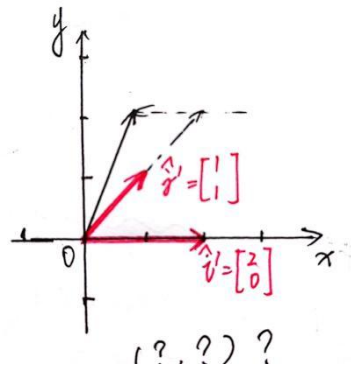


我们希望求解出 p 在这个空间当中的坐标，也就意味着我们必须求解出以下式子中 a ,

b 的值，像这样：

$$p = ai' + bj'$$

(a,b) 就是在这个空间中的“新坐标”。如何求呢？事实上不可能求出来，因为我没有告诉你任何有关于 i' 和 j' 的信息。这也就是说，我们必须建立原来的 i, j 与 i', j' 之间的关联，这也就意味着我们必须回归到平面直角坐标系中，这样才能构建起关联。像这样：



现在，我可以告诉你， i', j' ，在空间当中的坐标，像这样：

$$i = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow i' = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$j = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow j' = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

下面，为了加以区分，我们设“原始坐标”内的坐标是 (x, y) ，“新坐标”内的坐标是 (x', y') ，注意，这里的“坐标”不是“平面直角坐标系”中的坐标，而是用向量线性组合的“权”所组成的坐标。换句话说， p 可以表示成下面的两种形式：

$$p = \begin{bmatrix} 1 \\ 0 \end{bmatrix}x + \begin{bmatrix} 0 \\ 1 \end{bmatrix}y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$p = \begin{bmatrix} 2 \\ 0 \end{bmatrix}x' + \begin{bmatrix} 1 \\ 1 \end{bmatrix}y' = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

换句话说，我们可以通过矩阵，构建同一个空间之中“两个坐标”的关系，像这样：

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

下面我们就可以构建起来两个“坐标”之间的“关系”

换句话说，如果我们把“新坐标”的坐标乘上这个矩阵，就会得到原始空间内，描述这个向量的方式，像这样：

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

同理，如果我们把“原始坐标”的坐标乘上这个矩阵的逆，就会得到“新坐标”内，描述这个坐标的方式，像这样：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}$$

下面，我们考虑这样一个问题，如果在“原始坐标”内旋转 90° ，那么，如果用“新坐标”，如何描述这件事呢？

我们首先考虑如何将“原始空间”旋转 90° ，我们需要下面这个矩阵

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

比如，原始空间内的坐标是 $\begin{bmatrix} x \\ y \end{bmatrix}$ ，像这样：

$$\begin{bmatrix} x \\ y \end{bmatrix}_{\text{转}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

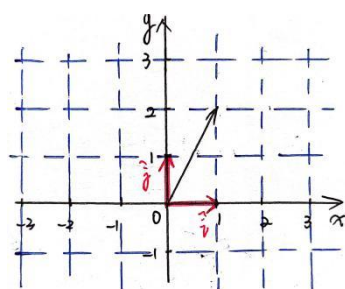
那么，这也就将整个原始空间旋转了 90° ，如果在空间当中，我们可以观察一下向量 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

（原始坐标的描述），将会被旋转到 $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ （原始坐标的描述）。下面，我们考虑 $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ 在

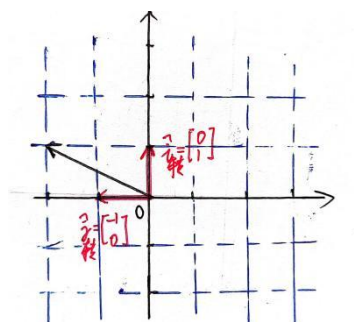
“新空间”中应该如何描述，也就意味着，运用刚才的理论，我们可以进行以下的操作，便可以获得在“新空间”内的坐标：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}_{\text{转}} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ 1 \end{bmatrix}$$

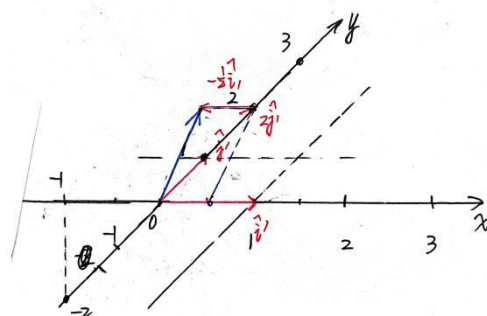
那么我们也得到了，在“新坐标”内的坐标。如果把这个过程写在图中，将会是这样，未经过旋转是这样：



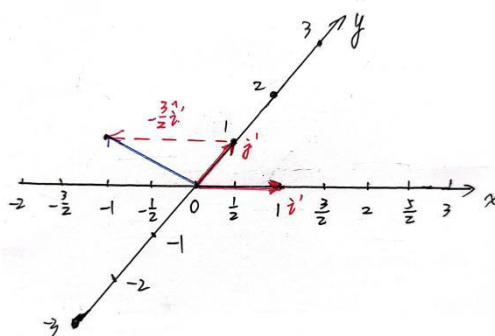
旋转后是这样：



那么在新坐标下看会是这样：



旋转后的新坐标会是这样：



下面，我们考虑任意一个旋转后的向量，在“新坐标”内的表达。首先在“原始坐标”中这很容易，我们只需要下面这个操作，就可以获得旋转整个空间后的向量的“原始坐标”，像这样：

$$\begin{bmatrix} x \\ y \end{bmatrix}_{\text{转}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

那么“新坐标”呢？是的，只需要我们乘一个逆就可以了，像这样：

$$\begin{bmatrix} x' \\ y' \end{bmatrix}_{\text{转}} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

但是，这里我们考虑的是从“原始坐标”到“新坐标”。下面我们考虑直接从“新坐标”

到“新坐标”，也就意味着，我们只需要将上面这个公式中的 $\begin{bmatrix} x \\ y \end{bmatrix}$ 换掉，像这样：

$$\begin{bmatrix} x' \\ y' \end{bmatrix}_{\text{转}} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

于是我们就获得了如何用“新坐标”描述“空间旋转”。

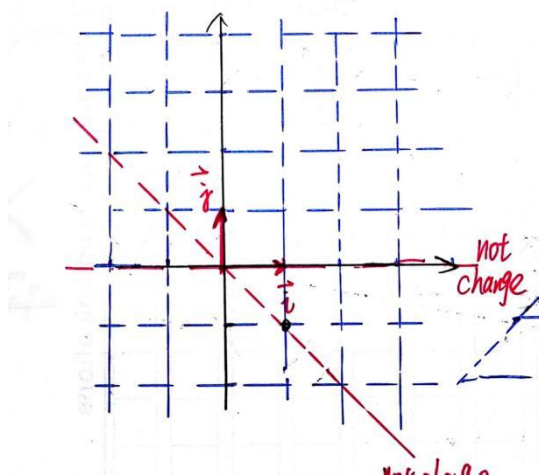
你或许会问，这很美吗？这有什么用？下一节你将会为此而感动落泪！

Chap3.9 特征向量与特征值的本质

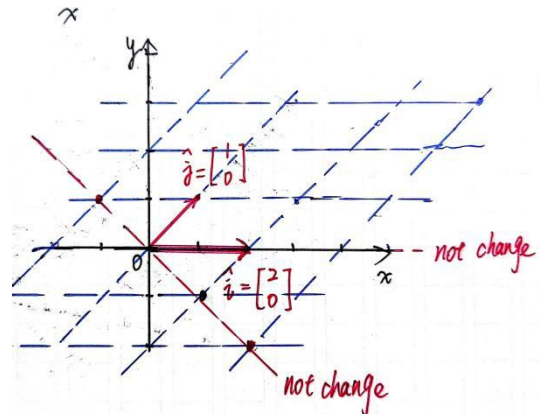
前面，我们讨论了矩阵的本质，就是操纵线性空间进行线性变换。那么我们看下面这个矩阵，我们要发现经过这个矩阵变换的“变中不变”的部分。

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

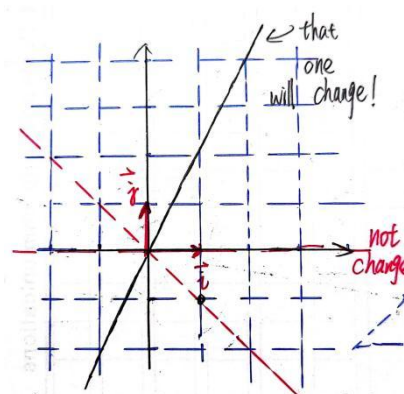
变化前，整个线性空间将会长成这个样子：



请关注 x 轴，和 $y=-x$ 这条对角线，这两条线在经过线性变换之后会产生神奇的效果。我们考虑任意一个向量后，经过这个矩阵的变换，都会或多会少得经过一些旋转，而这两条线却仍然保持在这条线上，像这样：



是不是很有趣呢，假若我们随便找另一条线，那么线性变换后肯定还是直线，但是一定偏离了原来的位置，像这样：



[illegible]

我们首先要将上面的几何变换翻译为数学语言，像这样：

这句话的含义就是 \mathbf{x} 向量经过矩阵 \mathbf{A} 的变换之后，还和向量 \mathbf{x} 共线。我们从这个角度认识一下上面的变换，像这样：

$$y = \lambda y$$

 $\lambda = 1$
$$\vec{x} = c \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

这就是特征向量和特征值。特征向量代表着经过矩阵线性变换之后依旧保持不旋转的向量，特征值代表着经过矩阵线性变换后，变化的长度是多少。在这个例子中，线性变换前后，这个向量不发生拉伸或者压缩。

$$A\vec{x} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \vec{x} = \lambda \vec{x}$$
$$(A - \lambda I)\vec{x} = 0$$

是的，到这里还没有完，请你注意 \mathbf{x} 向量前面是个什么？是的，是一个矩阵，而如果这个矩阵是一个正常的矩阵，也就意味着它的行列式不为 0，那么满足上面这个式子的向量 \mathbf{x} 必定只有零向量，还记得我们之前讨论“线性方程组通解”吗？

所以，如果有能满足这个表达式的 \mathbf{x} ，则必须要求前面的矩阵是奇异的，也就是前面的矩阵的行列式为 0，像这样：

$$\det(A - \lambda I) = 0$$

于是，我们就可以通过这种方法求出来对应的 λ （特征值），进而将 λ 回代，通过求解这样一个矩阵所代表的线性方程组，进而求解出我们需要的特征向量。

下面我们讨论特征根与矩阵的关系。我们先从简单的 2×2 矩阵开始研究，像这样，如果我们要找到这个矩阵的特征向量：

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

也就意味着，我们要找到满足下面的特征值：

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

也就意味着特征值需要满足以下的方程：

$$\lambda^2 - (a + d)\lambda + \begin{vmatrix} a & c \\ b & d \end{vmatrix} = 0$$

所以，如果我们这个矩阵有两个特征值，则必然满足以下的关系：

$$\lambda_1 \lambda_2 = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

$$\lambda_1 + \lambda_2 = a + d = \text{trace}$$

是的特征值之积就等于行列式，特征值之和就是对角线之和。这一结论不仅适用于 2×2 的矩阵，对于 $n \times n$ 的矩阵，这一想法依旧成立，需要用到高次的韦达定理即可。

你或许会问，为啥要学“特征向量”啊？

在回答这个问题前，我们先引入这样的一个模型。加入我们要计算下面这个矩阵的三次幂，那么如果我们手算，似乎还可以，像这样：

$$\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^3 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 0 & 1 \end{bmatrix}$$

那么如果是一百次幂呢？像这样：

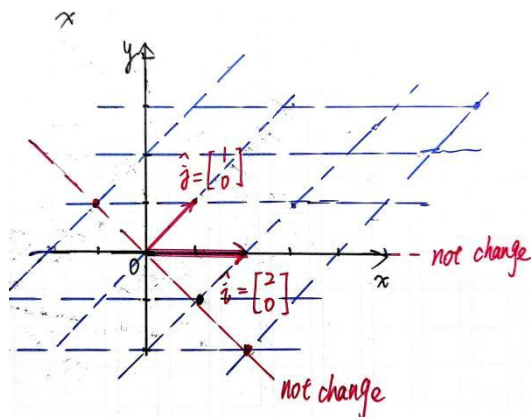
$$\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^{100}$$

你是否要吐血了呢？

注意，每次相乘一个矩阵，意味着将整个空间进行某种变换，空间中的某些向量的位置会改变，而如果我们在普通的“直角坐标系”下，这将会极其复杂，十分难以窥探变换的规律，这也正是上一节“基变换”发挥功能的时刻了！

一个矩阵，作用于一个空间，有些东西是保持不变的。这就是我们这一节讲的“特征向量”，特征向量在一个矩阵的作用后，在一定程度上是“不变的”，因为它只会作“拉伸”或者“压缩”！

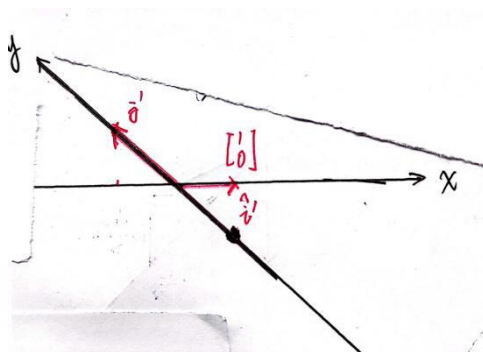
由前面的推理，我们可以知道，在经过这个矩阵的变换后，这两条线上的向量不发生旋转：



这也就意味着特征向量是：

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

有了这一想法后，我们不妨先考虑我们的特征向量所代表的坐标，在几何中，这个坐标系将会变成这样：



那么如果有矩阵作用于这个坐标系的话，这个坐标系只会有坐标轴的伸长和压缩，那么下一步，我们的思路，就应当是，把我们“原始坐标”的坐标，转化成“特征坐标”中的坐标。和上一节一样，我们设“原始坐标”为 $\begin{bmatrix} x \\ y \end{bmatrix}$ ，“特征坐标”为 $\begin{bmatrix} x' \\ y' \end{bmatrix}$ 。如果一个向量可以向下面这样描述：

$$p = \begin{bmatrix} 1 \\ 0 \end{bmatrix}x + \begin{bmatrix} 0 \\ 1 \end{bmatrix}y$$

那么，在“特征坐标”中，其描述方法将会是这样：

$$p = \begin{bmatrix} 1 \\ 0 \end{bmatrix}x' + \begin{bmatrix} -1 \\ 1 \end{bmatrix}y'$$

于是，我们就可以将“原始坐标”，转化成特征坐标了，像这样：

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

接下来，我们考虑在这个新的坐标系下进行 100 次变换？如何变换？首先，这个矩阵只会伸长或者压缩我们“新坐标”的基向量。这也就意味着，没经过一次变换，“新坐标”中的“坐标”，就会增大“特征值”倍。像这样：

例如进行一次上面的矩阵变换，新坐标的变化将会是这样的：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} 2x' \\ y' \end{bmatrix}$$

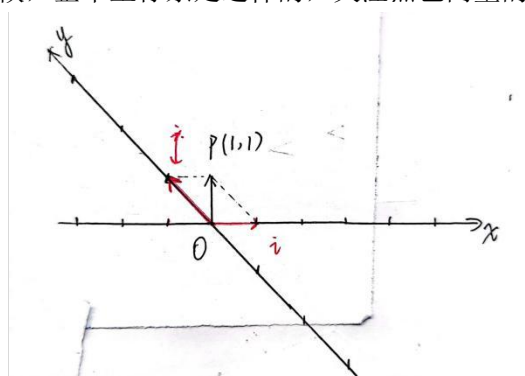
再进行一次上面的矩阵变换，新坐标的变化将会是这样的：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} 2x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} 2^2 x' \\ y' \end{bmatrix}$$

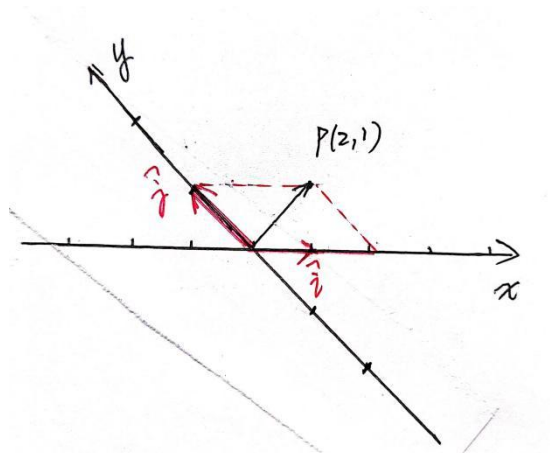
那么如果我们进行一百次矩阵变换，新坐标的变化将会是这样的：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} 2x' \\ y' \end{bmatrix} \Rightarrow \begin{bmatrix} 2^2 x' \\ y' \end{bmatrix} \Rightarrow \dots \Rightarrow 100 \text{次} \Rightarrow \begin{bmatrix} 2^{100} x' \\ y' \end{bmatrix}$$

上面，我们用坐标描述我们的思想，下面我们在几何中直观看到，刚刚发送了什么！没有经过变换的时候，整个坐标系是这样的，关注黑色向量的坐标表示：



经过一次变换后，将会是这样的。



前面是坐标和几何的表示，下面我们把这个过程用矩阵来表示，像这样：如果经历一次变换，那么在“新坐标”中，空间的变换将会是这样：

$$\begin{bmatrix} 2x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

如果再经历一次变换，那么在“新坐标”中，空间的变换将会是这样：

$$\begin{bmatrix} 2^2 x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^2 \begin{bmatrix} x' \\ y' \end{bmatrix}$$

如果经历 100 次矩阵变换，那么在“新坐标”中，将会变成这样：

$$\begin{bmatrix} 2^{100} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{100} \begin{bmatrix} x' \\ y' \end{bmatrix}$$

我们设在新坐标下进行变换的新矩阵称为 Λ 。像这样：

$$\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

如果我们把“新坐标”换成“原始坐标”，利用下面这个公式：

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

前面我们一直在用“等效”的想法，也就意味着，在新坐标下，用 Λ 进行的变换，等加成在原有坐标系下，用 A 进行的变换，也就意味着有以下的表达式成立：

$$A^n \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \Lambda^n \begin{bmatrix} x' \\ y' \end{bmatrix}$$

其中乘的矩阵代表着“坐标”之间的转换。

接下来，我们将上面的式子统一，像这样：

$$A^n \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \Lambda^n \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}$$

这也就意味着，我们只需求解下面的式子，就可以求解出 A 的幂指数了，像这样：

$$A^n = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \Lambda^n \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}$$

上面，我们一直在用几何的方式直观认识，下面我们从代数的角度再进行一次认识。

首先，我们把任意的一个矩阵，如果我们要计算它的 n 次幂，我们首先需要找到这个矩阵所对应的特征向量和特征值，如果这些特征向量不能作为一组可以线性表出整个空间的基向量，那么就不能使用以下的计算方法。

现在，假若我们很幸运地得到了全部由特征向量所组成的一个矩阵，我们称之为 S ，像这样：

$$S = [\vec{x}_1 \quad \vec{x}_2 \quad \dots \quad \vec{x}_n]$$

那么，这个矩阵其中的每一个 x ，就代表着一个特征向量。那么如果我们将它乘以 A ，将会变成这样：

$$AS = A[\vec{x}_1 \quad \vec{x}_2 \quad \dots \quad \vec{x}_n] = [A\vec{x}_1 \quad A\vec{x}_2 \quad \dots \quad A\vec{x}_n]$$

进一步地，如果我们利用特征向量的性质，那么将会有：

$$AS = A[\vec{x}_1 \ \vec{x}_2 \ \dots \ \vec{x}_n] = [A\vec{x}_1 \ A\vec{x}_2 \ \dots \ A\vec{x}_n] = [\lambda_1\vec{x}_1 \ \lambda_2\vec{x}_2 \ \dots \ \lambda_n\vec{x}_n]$$

是的，如果我们再进一步化简，提出一个类似于对角矩阵的矩阵，像这样：

$$AS = [\lambda_1\vec{x}_1 \ \lambda_2\vec{x}_2 \ \dots \ \lambda_n\vec{x}_n] = [\vec{x}_1 \ \vec{x}_2 \ \dots \ \vec{x}_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

于是，我们就获得了这样一个矩阵，我们和上面一样，管它叫做 Λ 。像这样：

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

于是我们就可以把上面复杂的式子写成这样：

$$AS = S\Lambda$$

如果我们两便同时在右边乘以 S 的逆，将会有：

$$A = S\Lambda S^{-1}$$

是的这和几何是殊途同归的，因为假若我们求解 A 的平方像这样：

$$A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$$

亲爱的读者，我不知道你是否已经看到“特征向量”的美妙了！假若再去一遍一遍地乘，仍然只需要将里面的对角矩阵进行 n 次方的计算即可，而对角矩阵的 n 次方很好计算。

Chap3.10 对称矩阵

首先，我们考虑任意一个对称矩阵的时候，都会有以下两个非常好的性质。举几个例子，例如第一个矩阵，像这样：

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

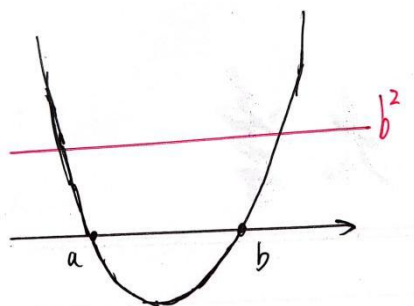
于是，这个矩阵的特征根，就像是下面这样，会得到：

$$\begin{vmatrix} a-\lambda & b \\ b & d-\lambda \end{vmatrix} = 0$$

那么也就意味着，有这件事情成立，像这样：

$$(a-\lambda)(d-\lambda) = b^2$$

如果画在图中，将会是这样：



可见， 2×2 的对称矩阵是一定有实数的特征根的，我们不再继续证明，而选择相信这一结论对 $n \times n$ 依旧成立。也就是说，我们有以下的结论：

对称矩阵的特征值都是实数

下面，我们考虑对称矩阵的另一个性质，也就是对称矩阵的特征向量都相互垂直。首先，我们考虑单位矩阵，是一个特殊的对称矩阵。那么它的特征向量，将会是所有的向量，因为经过单位矩阵后，空间并没有发生变化。我们可以选取空间中的一组向量，使得他们分别正交于彼此。

再例如，我们可以考虑任意的一个对称矩阵都可以进行以下的拆分：

$$S = Q \Lambda Q^{-1}$$

那么如果我们将这个矩阵转置

$$S^T = (Q^T)^{-1} \Lambda^T Q^T$$

并与原矩阵相乘，那么将会有：

$$S^T S = (Q^T)^{-1} \Lambda^T Q^T Q \Lambda Q^{-1}$$

那么观察一下，我们已经知道，对称矩阵转置依旧是对称矩阵，对称矩阵于对称矩阵相乘也依旧是对称矩阵，这也就意味着，如果 Q 是正交矩阵，那么上面的式子就可以化简成下面这样：

$$S^T S = (Q^T)^{-1} Q^{-1}$$

可以看到，等式两侧都是对称矩阵，这一结论大致应当是成立的。反之，如果不是正交矩阵，那么我们将会得到上面那一坨扭曲的式子，其正确性也不大清楚。由于我们学习的是线性代数而非高等代数，所以我们选择相信这个结论是正确的，也就意味着我们有以下的结论成立：

对称矩阵的特征向量可以相互垂直

接下来，如果我们认为以上的结论是正确的，并且用 Q 来代表一个正交矩阵的话，那么也就可以将任意的一个对称矩阵写成下面的形式：

$$S = Q \Lambda Q^T$$