前言

这是一本线性代数与矩阵分析的快速入门书，致力于40小时内，贯穿线性代数与矩阵分析的核心内容。作者目前是华中师范大学AI专业在读本科生，正在复习线性代数，学习矩阵论，最优化，机器学习等课程。深知数学公共课远不满足专业课的数学要求，故以开源的方式著此书。尽可能以小白的视角来帮助本科生理解线性代数与矩阵分析。

本书包含以下内容，矩阵的表示，线性变换，矩阵的真实大小——秩，本征值与本征向量（eigenvalue and eigenvector,有的书也叫特征值、特征向量，但是区别于机器学习里的特征向量feature vector），相似变换，迹，内积，范数，奇异值分解等。

本书由于注重培养直觉，不会很严谨，并且在正文里基本上不会有严谨的定义等。对于本书需要严谨的部分，将会以脚注和延伸版块的形式进行补充，但并不意味着你可以不读此部分。书中会有少量习题，并且是高质量的，难度适中，计算量小，部分习题有承前其后的作用，会单独标记出来。

希望这本快速入门书能帮助到你，如果你有任何关于内容的问题，请向我发邮件,我的邮箱是**2314181884@qq.com**。

我希望有更多的人来帮助我完成这本书，你可以在Github上的Linear-Algebra-and-Matrix-From-Students-Perspective仓库克隆代码，然后推送你修改或补充的代码到项目仓库里。

第一章:初识矩阵

1.1从线性方程组到矩阵

在初高中我们学过线性方程组，需要求解三元一次方程组，比如:

$$\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3=b_1 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3=b_2 \\ a_{31}x_1+a_{32}x_2+a_{33}x_3=b_3 \end{cases}$$

当需要描述的数字变得很多，比如需要求解七元一次方程组，那么写这么多数字与字母将变得不幸。为了简单，我们引入矩阵。这并不是一个高大上的概念，它很朴素，你可以理解为就是一张数表而已。

对上面的三元一次方程，我们用矩阵描述为$A\vec{x}=\vec{b}$。这个描述很像$ax=b$这个一元一次方程句号。相信你很容易猜到了A，$\vec{x}$与$\vec{b}$的含义了。

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \vec{x}=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \vec{b}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

其中A称为3x3矩阵，即3行3列矩阵。$\vec{x}$和$\vec{b}$都称为列向量，因为都是竖直的，向量也是矩阵，这两个向量是3x1矩阵，即3行1列矩阵。

1.2 $A\vec{x}=\vec{b}$的四种视角

除了1.1节给出的简化表达式$A\vec{x}=\vec{b}$，还能怎么简化呢？因为1.1简化得过于厉害，我们都似乎不知道该怎么计算这个线性方程组了。我们按照提取公共部分的想法，你看x_1，x_2，x_3在方程组里重复书写了好几次，按照这样的想法，我们可以把方程组简写成:

$$x_1\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}+x_2\begin{bmatrix} a_{12} \\ a_{22} \\ a_{23} \end{bmatrix}+x_3\begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

我们很容易看出这个表达式的意思，根据它进行计算，并且我们能从这个式子发现一些有趣的想法，等式左边是矩阵A的列[1]的线性组合，系数是x_1，x_2，x_3

有三种代数视角

视角一:以$\vec{b}$为中心

等式表达的是，向量$\vec{b}$是A的列的组合的产物，如果能找到组合系数$\vec{x}$，那么该线性方程组有解，可以为$\vec{x}$

视角二:以A为中心

先只看左边$A\vec{x}=x_1\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}+x_2\begin{bmatrix} a_{12} \\ a_{22} \\ a_{23} \end{bmatrix}+x_3\begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$，当$\vec{x}$取任意值，那么有无穷多个$A\vec{x}$的值，那么这个结果集合我们称作$A$的列空间，记作$C(A)$(Column Space of A)

因为是列的组合的结果集，所以叫做列空间，而不是其他什么空间。空间比结果集更好听，且有几何的意味。这里A的每一列都是3x1的矩阵，也就是向量。有列空间，当然也有行空间，但是这里暂时不讨论，需要的时候再讨论。如果$\vec{b}$在$C(A)$里，那么该方程组有解。

视角三:以$\vec{x}$为中心

$A\vec{x}=\vec{b}$好像是函数表达式，A操作$\vec{x}$得到$\vec{b}$。若看成$f(\vec{x})=\vec{b}$，那么输入$\vec{x}$，输出$\vec{b}$，函数关系A。可以把A理解为操作。另外向量本身是具有几何的意味，那么向量$\vec{x}$到向量$\vec{b}$的函数可以称作向量的几何变换。在代数上，$A\vec{x}=\vec{b}$是A把$\vec{x}$映射成$\vec{b}$

视角四:矩阵的本质是坐标系变换

我们想问$\vec{x}$是如何变换到$\vec{b}$的? 自然是通过A作用于$\vec{x}$得到$\vec{b}$。但是太抽象了，我们需要仔细挖掘一下具体是怎么变换的。那么看看$x_1\begin{bmatrix}a_{11}\\a_{21}\\a_{31}\end{bmatrix}+x_2\begin{bmatrix}a_{12}\\a_{22}\\a_{23}\end{bmatrix}+x_3\begin{bmatrix}a_{13}\\a_{23}\\a_{33}\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$吧，这具体了一些，但是依旧很抽象，我们换成具体的数好了。写个二元一次的例子叭。

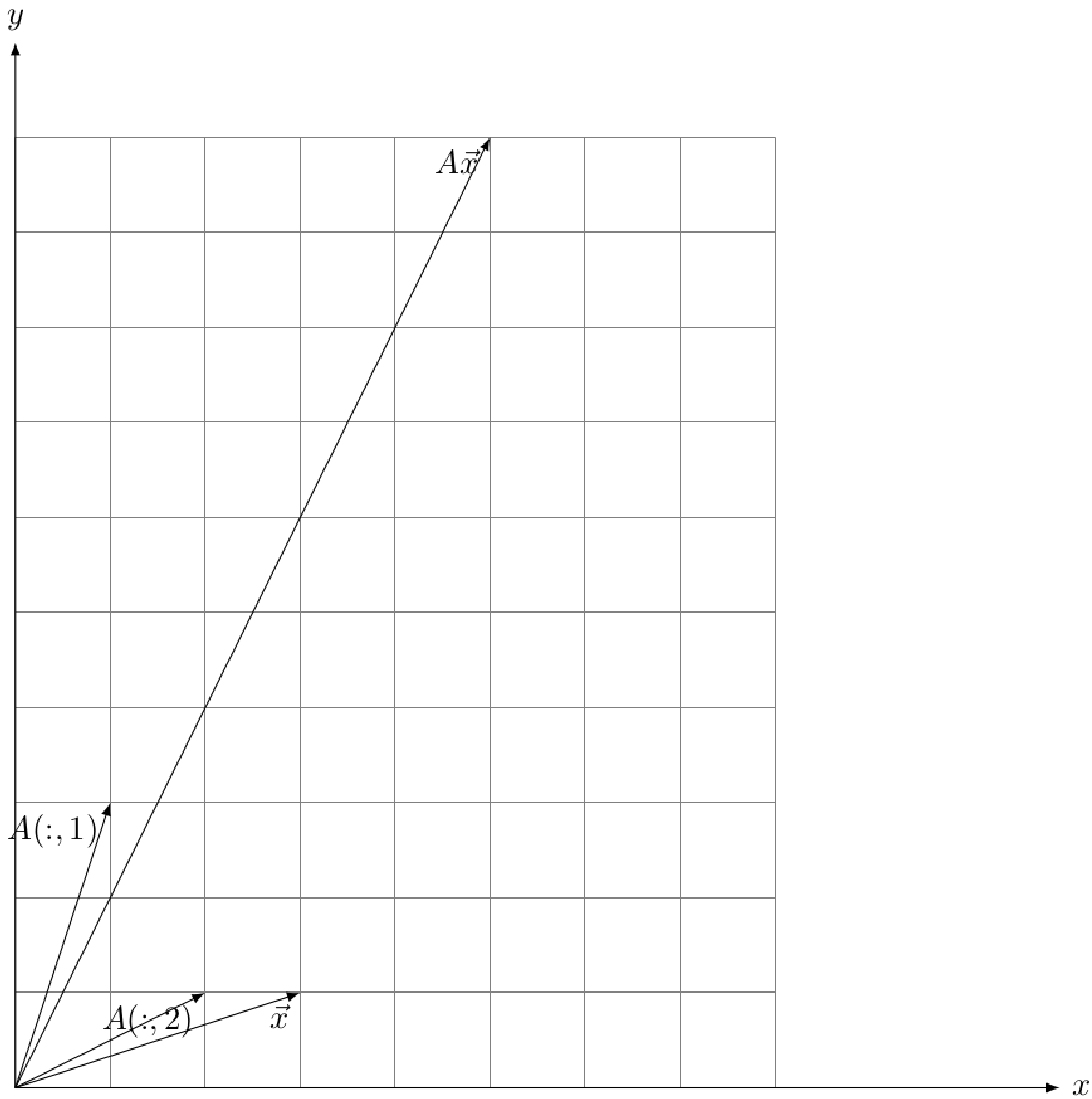
$$\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=x\begin{bmatrix}1\\0\end{bmatrix}+y\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}x\\y\end{bmatrix}$$

我们可以看出来这个矩阵很特殊，作用是啥也不干，是恒等变换。如果换成$\begin{bmatrix}2&0\\0&2\end{bmatrix}$，那么是把$\vec{x}$的长度扩长为原长度的2倍。一般的，矩阵$\begin{bmatrix}c&0\\0&c\end{bmatrix}$表示伸缩变换，扩长为原长度的$c$倍。

如果看看下面这个矩阵呢?

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

则$A\vec{x}=\begin{bmatrix}x+2y\\3x+y\end{bmatrix}$，这好像很难看出什么几何意义了，我们画画图试试看，并且给$\vec{x}$赋值为$(3,1)$。



图中$A\vec{x}$，$\vec{x}$，$A(:,1)$，$A(:,2)$似乎也看不出什么关系，但若结合线性组合看看

$$3\begin{bmatrix}1\\3\end{bmatrix}+1\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}5\\10\end{bmatrix}$$

答案似乎呼之欲出了，但可能还是有些模糊，看看$\begin{bmatrix}2&0\\0&2\end{bmatrix}\begin{bmatrix}3\\1\end{bmatrix}=3\begin{bmatrix}2\\0\end{bmatrix}+1\begin{bmatrix}0\\2\end{bmatrix}=\begin{bmatrix}6\\2\end{bmatrix}$

这两个矩阵方程有什么共同点呢？很明显，他们的系数$\vec{x}$是一样的，不同的是A的列不一样。其实答案已经出来了，只是我们习惯了平面直角坐标系，我们画的图都是基于平面直角坐标系的，这样就很不容易看出来了。我们认为坐标$\vec{x}$的是一样的，但是我们又知道他们是不一样的，因为坐标系不一样。对于图一，它的新

的坐标系是由$\begin{bmatrix}1\\3\end{bmatrix}$与$\begin{bmatrix}2\\1\end{bmatrix}$构成的，看看，如果我们把$\begin{bmatrix}1\\3\end{bmatrix}$当作"x轴"的基向量，$\begin{bmatrix}2\\1\end{bmatrix}$当作"y轴"的基向量，那么$P$点在新坐标系下的坐标为$\begin{bmatrix}3\\1\end{bmatrix}$，而在平面直角坐标系$\begin{bmatrix}1\\0\end{bmatrix}$与$\begin{bmatrix}0\\1\end{bmatrix}$下的坐标为$\begin{bmatrix}5\\10\end{bmatrix}$

小结:三种代数视角与几何视角的关系

视角一与几何视角

$\sum x_iA(:,i)$即A的列的线性组合，系数是x_i，那么新坐标系下点的坐标为x_i，新坐标系由$A(:,i)$充当基向量来构成。

视角二与几何视角

A的列空间$C(A)$由新坐标系的基$A(:,i)$组合而成。$C(A)$有多大，新坐标系的范围就有多大。

视角三与几何视角

$A\vec{x}=\vec{b}$即$f(\vec{x})=\vec{b}$可以看成是对标准基向量(例如$\begin{bmatrix}1\\0\end{bmatrix}$)组成的直角坐标系进行坐标系变换。变换前后，点在各自坐标系下的坐标是相同的，如果换算坐标系，那么新坐标系下的坐标在旧坐标系下的坐标与原坐标不一样，除非是恒等变换。

1.3坐标变换

我们最感兴趣的当然是几何视角了，但是对坐标系进行变换并不是我们想要关心的，我们希望关心同一坐标系下坐标是怎么进行变换的。

$\vec{x}$是怎么移动到$A\vec{x}$的呢？我们很自然的想到先把$\vec{x}$旋转到与$A\vec{x}$同一方向上，然后伸缩至与它相同的长度。当然也可以是先伸缩再旋转。那么A在旧坐标系对向量的作用可以看成是对向量$\vec{x}$进行"旋转"加"伸缩"。我们把进行旋转操作的矩阵记为R(rotation matrix)，进行伸缩操作的矩阵记为S(scaling matrix)。

如果用式子来描述R与S的性质，并且我们知道旋转操作是仅改变方向，伸缩操作是仅改变长度。那么旋转操作可以表达为$||R\vec{x}||$=$||\vec{x}||$[2]。而伸缩操作可以表达为$S\vec{x}=\lambda\vec{x}$，这不仅可以表示长度的关系，也可以表示方向是同向的。

想问问A与R，S有什么联系吗？我们知道A是R，S的组合操作，那用式子该怎么表示呢?是$A=R+S$还是$A=RS$呢？读者可能会从矩阵即函数的视角猜到应该是$A=RS$，因为函数就是操作，操作的组合表现为复合形式。并且由于这里R与S操作是相互独立的，先旋转再伸缩与先伸缩后旋转是等价的效果，所以这里有表达式$A=SR=RS$。注意$SR\vec{x}$表示先对$\vec{x}$进行R操作再S操作，要从右往左读[3]。

那我们怎么知道R与S呢？S是很容易知道的，因为伸缩矩阵前面已经知道形式了，只需要伸缩比例λ即可。R暂且不知道，先看看简单的S叭。

$\lambda = \frac{||S\vec{x}||}{||\vec{x}||}$，那么对于1.2中的例子，$||\vec{x}|| = \sqrt{3^2+1^2} = \sqrt{10}$，$||S\vec{x}|| = \sqrt{5^2+10^2} = 5\sqrt{5}$，所以$\lambda = \frac{5\sqrt{2}}{2}$。

旋转矩阵R似乎无从下手，我们想想旋转首先得知道旋转的角度有多大，以及旋转的方向是顺时针还是逆时针方向。这里我们可以根据向量的内积与长度的关系求出角度，对于这个例子，方向也很容易看出来，是逆时针。但是这两个参数知道了，用矩阵该怎么表示呢？有人说我可以进行逆向工程，因为知道了A与S，我可以求出R句号。但是这就涉及了矩阵的计算，属于本书后面章节的内容，目前我们是希望推导出R。

我们知道矩阵仅仅是一张为了简化书写的数表，完全可以通过其他方法得出来该矩阵。我们可以用坐标来推出矩阵R，因为R涉及到了旋转角度与旋转方向，我们可以联想到起同样作用的三角函数，我们用三角函数来表示坐标叭。一般的，我们有$\vec{x} = (||\vec{x}||\cos\alpha, ||\vec{x}||\sin\alpha)$，$A\vec{x} = (||A\vec{x}||\cos(\alpha+\beta), ||A\vec{x}||\sin(\alpha+\beta))$。其中$\alpha$是$\vec{x}$与$x$轴的夹角，$\beta$是$\vec{x}$与$A\vec{x}$的夹角，在三角函数里我们知道，$\beta > 0$代表逆时针旋转$\beta$度，反之是顺时针旋转。三角函数式的坐标表达包含了足够的信息，我们只需要看看坐标的关系即可。

我们令$x_1 = ||\vec{x}||\cos\alpha$，$y_1 = ||\vec{x}||\sin\alpha$，$x_2 = ||A\vec{x}||\cos(\alpha+\beta)$，$y_2 = ||A\vec{x}||\sin(\alpha+\beta)$。那么有
$x_2 = ||A\vec{x}||(\cos\alpha\cos\beta - \sin\alpha\sin\beta) = \lambda||\vec{x}||(\cos\alpha\cos\beta - \sin\alpha\sin\beta) = \lambda(x_1\cos\beta - y_1\sin\beta)$，$y_2 = \lambda||\vec{x}||(\sin\alpha\cos\beta + \cos\alpha\sin\beta) = \lambda(y_1\cos\beta + x_1\sin\beta)$。让我们提取出$x_1$，$y_1$得到关于$\beta$的矩阵。

那么$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \lambda \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$，我们可以看出$R = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}$句号。有人会问，那$S$呢，你总不可能用一个数$\lambda$代替一个矩阵吧!

我们知道伸缩变换$S\vec{x} = \lambda\vec{x}$，这样的话，
$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \lambda \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = SR\vec{x}$ [4]

之前说S与R这两个操作互不影响，是相互独立的，那么改变顺序也应该成立，我们把λ提到R的右边，有$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \lambda \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = RS\vec{x}$。可以动手用数据验证下1.2节的例子。

注意，并不是所有的矩阵A，B都有$AB = BA$。正好提个疑问，什么时候两个矩阵满足乘法交换率呢，两个就矩阵一定要相互独立吗？怎么判断两个矩阵是相互独立的呢？在后续章节，听我娓娓道来，不要急。

1.4矩阵乘法的本质计算

经过刚才的讨论,我相信大家一定很感兴趣AB是怎么计算的，但是请大家耐住性子,我骸骨不想直接告诉你怎么计算(就像通常教科书那样很复杂的公式),也不想很快告诉你怎么简单计算出来。我想先告诉你他的本

质,同过本质来理解矩阵乘法的计算。

1.2节的视角三告诉我们,矩阵就是函数,是操作。那么两个矩阵相乘就是函数的符合。几何视角告诉我们,矩阵操作是坐标系变换,新坐标系的基为矩阵的各列。那么矩阵相乘就是坐标系的多次变换。如果是以抽象的函数视角来看,我们很难得出计算方法,那么我们就从几何角度来理解矩阵相乘的是如何计算的。

在这里,我们不得不先通过1.2的例子理清一下坐标系问题。我们是默认坐标系是自然基下的坐标系,也就是像$\begin{bmatrix}1\\0\\0\end{bmatrix}$,$\begin{bmatrix}0\\1\\0\end{bmatrix}$,$\begin{bmatrix}0\\0\\1\end{bmatrix}$的基。相互正交,且长度为1的基称为标准正交基,如果对于每个基,只有一个位置为1,其余位置为0,那么称作自然基,也就是具有唯一性质的标准正交基。

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

上述矩阵的方程可以看成一共有4个列向量,都是二维自然基下的坐标。从坐标系变换的角度来看$\vec{x}=\begin{bmatrix}3\\1\end{bmatrix}$,首先他在自然基下的坐标为(3,1),经过$A$的作用,变换后的$A\vec{x}$在新坐标系(新的基为$\begin{bmatrix}1\\3\end{bmatrix}$与$\begin{bmatrix}2\\1\end{bmatrix}$)下的坐标为(3,1),$A\vec{x}$再转换到自然基下的坐标就是(5,10)。

那么对于$AB\vec{x}$我们最终得到的向量的坐标也是在自然基下表示的。下面，我们来更细致地看看，先令$B\vec{x}=\vec{b}$,表示自然基下的$\vec{x}$经过B操作，$B\vec{x}$在B的各列组成的坐标系下的坐标为$\vec{x}$,转换到自然基下的坐标就是$\vec{b}$。那么令$A\vec{b}=\vec{a}$,表示自然基下的$\vec{b}$经过A操作，$A\vec{b}$在A的各列组成的坐标系下的坐标为$\vec{b}$,转换到自然基下的坐标就是$\vec{a}$。综合来看,$AB\vec{x}=\vec{a}$就是,自然基下的$\vec{x}$依次进过B,A进行坐标系变换,最后在自然基下的坐标为$\vec{a}$句号。这其中的两次矩阵操作过程完全可以看成是一次矩阵操作,记为C。

说了这么多,似乎已经不耐烦了,那么现在进入核心部分。坐标系变换关注的就是基变换，$\vec{x}$在自然坐标系下的最终去向为与最终复合的坐标系的基(C的各列)有关,因为$AB\vec{x}=C\vec{x}$。也就是说,我并不需要关注$\vec{x}$具体为多少了,我只要知道C为多少,通过输入,经过C就知道最中的向量在自然基下的坐标。我们只需要追踪基是如何变换的就可以了,举个例子叭。

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

我们需要知道基$\begin{bmatrix}1\\1\end{bmatrix}$,$\begin{bmatrix}-2\\0\end{bmatrix}$是如何变换到$\begin{bmatrix}2\\1\end{bmatrix}$,$\begin{bmatrix}0\\-2\end{bmatrix}$,他们必须是经过了$A=\begin{bmatrix}0&2\\1&0\end{bmatrix}$才能得到$C$。$B$的基$B(:,1)=\begin{bmatrix}1\\1\end{bmatrix}$在自然基下的坐标为$(1,1)$,经过$A$,则$AB(:,1)=\begin{bmatrix}0&2\\1&0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$在A的基下的坐标为$(1,1)$,转换为自然基下的坐标就是1 $\begin{bmatrix}0\\1\end{bmatrix}+1\begin{bmatrix}2\\0\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$，同样的，$\begin{bmatrix}0&2\\1&0\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix}=\begin{bmatrix}0\\-2\end{bmatrix}$。这样子我们知道了,矩阵乘法是通过追踪基的变换来计算的。

一般地,我们有$AB=\begin{bmatrix}AB(:,1)&AB(:,2)&\dots&AB(:,m)\end{bmatrix}$。这里我们也能知道,$A$与$B$相乘必须A的列数与B的行数要一致,否则A不能与B的列相乘。我们还可以知道,对B的列数没有与A有关的要求,仅仅要求

与$\vec{x}$的行数相等。

1. 采用matlab里矩阵各行各列的表示，例如$A(:，i)$表示第i列，$A(i，:)$表是第i行 ↩
2. $||R\vec{x}||$=$||\vec{x}||$是旋转操作的必要的条件，这是式子看不出来方向是否改变，它仅仅是表示旋转后长度不变，还要加上方向的表达式才行，即$\cos < A\vec{x},\vec{x} > \neq 0$ ↩
3. 复合函数都是从右往左读，数学里$f(g(x))$也就是$f \circ g(x)$，内层函数是g，先内层再外层，符号表达式上，就是从右往左读。 ↩
4. 把$R\vec{x}$看作新的一个向量$\vec{y}$，那么$\lambda\vec{y} = S\vec{y}$，就有了推导结果的表达式 ↩