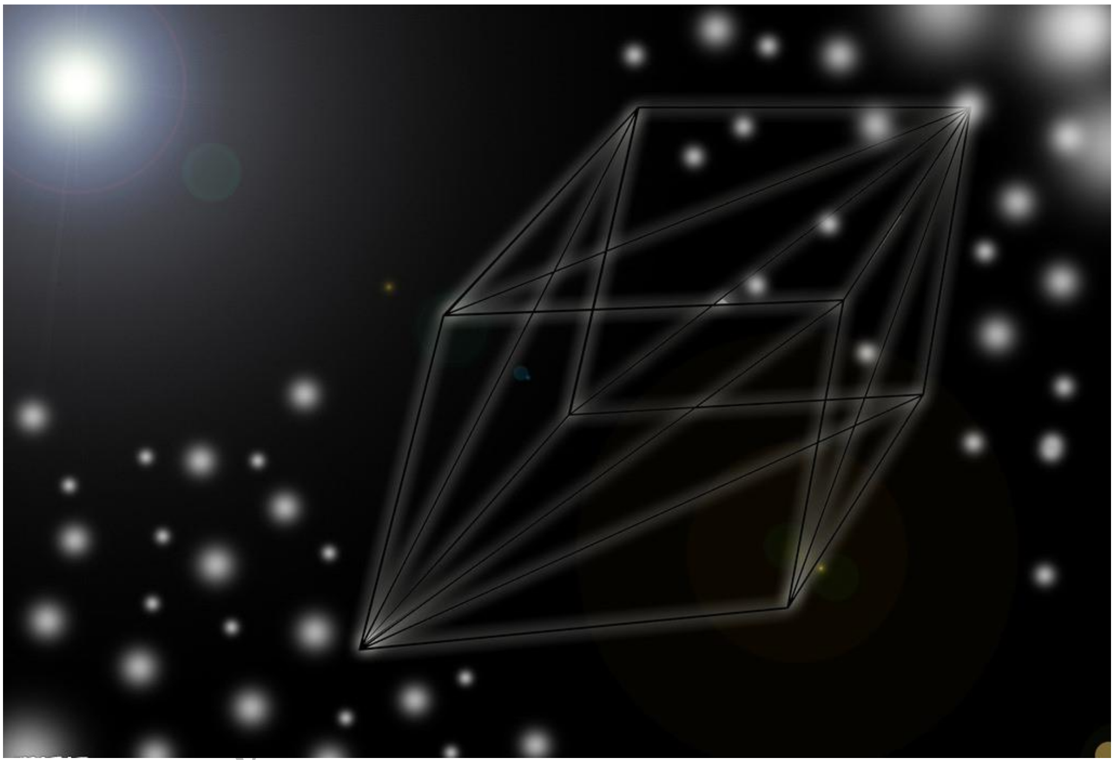
换个角度看线性代数



作者：风萧萧兮乱翻书

二〇一三年九月

前 言

首先声明，我不是数学老师，也不是一个数学高手，甚至连一个合格的数学爱好者都算不上，就算是一个数学草根吧。套用现在的网络流行词，顶多也就是个数学屌丝吧。

大学学习线性代数的时候，迫于考试的压力，强记了一堆枯燥的概念和晦涩的“定义”、“定理”、“性质”，毕业没几年就都还给老师了。印象当中，线性代数就是一堆矩阵的概念、定理、性质、变换之类的吧。矩阵到底是什么？矩阵的乘法为什么规定的那么奇怪？行列式究竟是一个什么东西？那个行列式的计算规则或许只在我脑海中留下了“怪异”这个形容词。相似矩阵怎么就相似了，到底相似在何处呢？哦，谢天谢地，我已经毕业了！偶尔想起线性代数的时候，我只会想，这门课为什么叫做“线性代数”呢？根据我的印象，它主要的工作是在折腾矩阵啊。或许应该叫做“矩阵概论”或者“矩阵折腾学”之类的更贴切吧。

近来，无意中在网上看到了网易公开课，（实在忍不住要为网易做得这件功德无量的公益事业叫好。当然现在新浪等网站也都在做，衷心感谢他们！）其中有麻省理工学院一个老教授 William Gilbert Strang 讲解线性代数的视频。好奇中，想看看世界顶级理工学府是怎么讲解线性代数的。看了几集，实在是有种拨云见日、茅塞顿开的感觉。（愤青们别喷我！或许我上的大学还不够好，或许我没有遇到更优秀的线性代数老师，或许我自身的理解能力有限，总之，我就事论事，别说我崇洋媚外。）虽然我至今没看全，也有很多没看懂。一是

风萧萧兮乱翻书

因为本人水平实在有限，二是老教授更注重引导，课堂上常略去些教材上的东西，而我没有他们的教科书，所以有时不免感觉突兀而跟不上趟（这算是我自我安慰的理由吧）。但是，还是鼓励和启发了我试图从另一个角度来重新看待线性代数。

突然有了些兴趣后，又在网上搜寻和拜读了一些专家、网友关于线性代数各个方面的精彩文章和论述，观看了一些教学视频，所有这些都让我受益匪浅，感觉对线性代数有了些感性的理解。本人虽水平有限，但无知者无畏，于是还是结合本人的一些思考，将自己对线性代数的粗浅理解整理了一些文字，以一个数学草根的角度，从“几何”、“图形”出发理解线性代数，希望能够对大家学习和理解线性代数有所帮助。

风萧萧兮乱翻书

目　录

风萧萧兮乱翻书

第一章　线性方程组和向量空间

§1 线性方程组的几何意义和解的情况

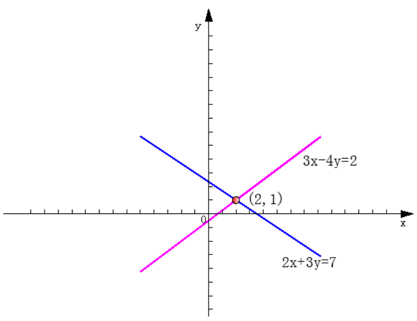
线性代数和其他的课程有些不一样，它好像没有一条固定的、大家公认的主线。有的教科书从矩阵开始讲，有的教科书从线性方程开始讲，还有的教科书从行列式开始讲，不过，上来就扯上向量空间的倒还真不多。

我们还是从熟悉的开始吧。先来看一个二元线性方程组 $\begin{cases} 2x+3y=7 \\ 3x-4y=2 \end{cases}$

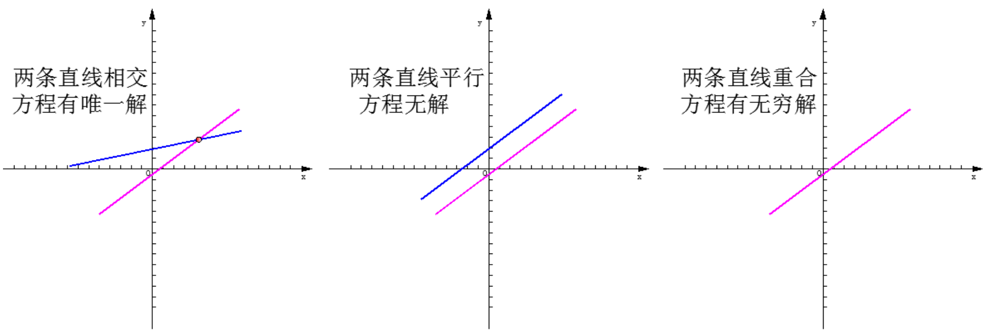
解这个方程组当然不难（你要不会解就别往下看了吧），容易解得 x=2，y=1

这个方程组在几何上有什么意义呢？

从平面解析几何的角度，二元线性方程组的两个方程分别代表二维平面上的一条直线，所求的解就是两条直线的交点。



由此推广到任意的两个二元一次方程组成的方程组，由图形可以直观的理解，当两个方程表示的两条直线平行时，两条直线无交点，方程无解；两条直线重合时，有无穷个交点，方程有无穷个解；除了上述两种情况外，两条直线必有一个交点，方程有唯一解。

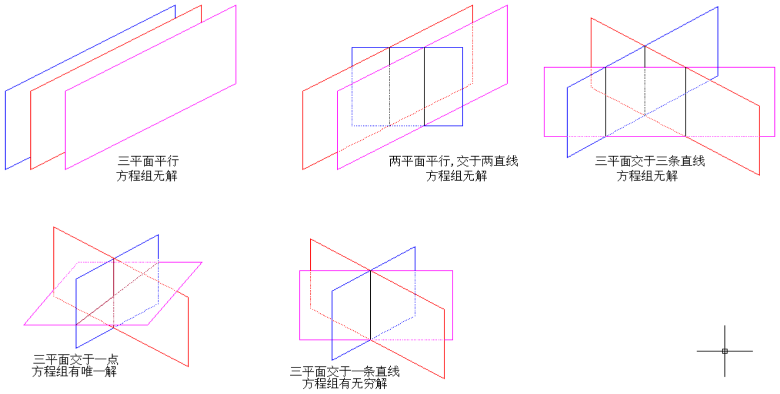


再来看一个三元线性方程组 $\begin{cases} 2x+3y+z=6 \\ x-y+2z=-1 \\ x+2y-z=5 \end{cases}$

当然也容易解得 y=1，x=2，$z=-1$

从空间解析几何的知识我们知道，上面方程组中的每个三元一次方程代表三维空间上的一个平面，所求的解就是三个平面的交点。三个平面的交点有如下图

所示几种情况：若三个平面都平行，没有交点，方程无解；若其中两个平面平行，第三个平面将与其分别交于两条平行直线，没有共同交点，方程无解；若三个平面交于三条平行直线（此时三个平面必同时垂直于某一平面），没有共同交点，方程无解；若三个平面交于一条直线，方程有无穷个解（无穷个交点）；除了上述情况外，三个平面将交于一点，方程有唯一解。

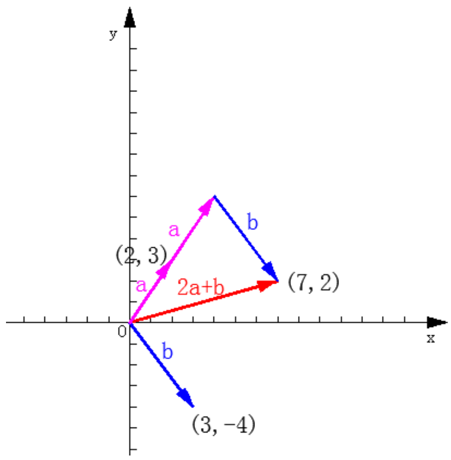


由上面的图形分析可以看出，三维空间的情况比二维空间已经复杂得多。按照这个思路推广下去看来是难以可持续发展了，何况四维空间超平面的相交情况我根本想象不出来，更别说画出来了。

此路不通，不妨换个思路吧。以上是从方程组的单个方程出发，理解方程组的几何意义。如果将方程组看做“一个方程”呢？比如将上面的二元一次方程组看做“一个方程”：

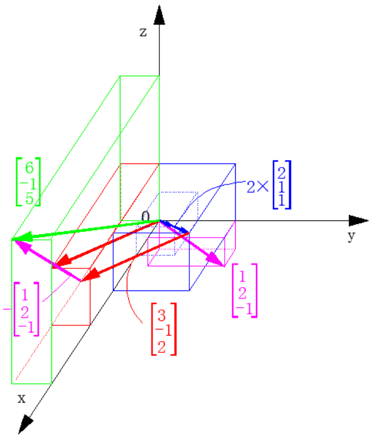
$$\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\begin{bmatrix}7\\2\end{bmatrix}$$

这个“方程”的几何意义可以理解为三个二维向量的线性关系的描述。如图所示，2 个（所求的 x）$\begin{bmatrix}2\\3\end{bmatrix}$向量与 1 个（所求的 y）$\begin{bmatrix}3\\-4\end{bmatrix}$向量相加得到向量$\begin{bmatrix}7\\2\end{bmatrix}$。



同样的，上面的三元一次线性方程可看做“一个方程”：

$$\begin{bmatrix}2\\1\\1\end{bmatrix}x+\begin{bmatrix}3\\-1\\2\end{bmatrix}y+\begin{bmatrix}1\\2\\-1\end{bmatrix}z=\begin{bmatrix}6\\-1\\5\end{bmatrix}$$

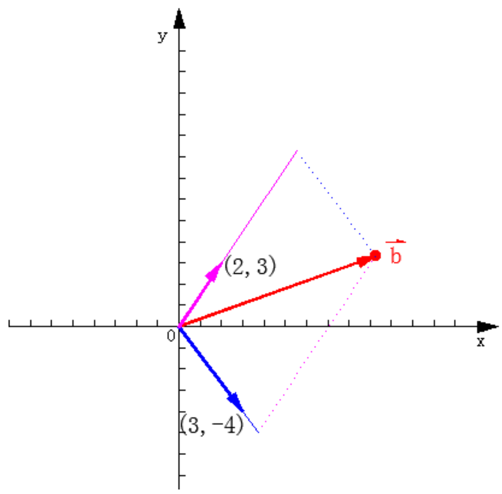


这个“方程”的几何意义可以理解为四个三维向量的线性关系的描述。2 个 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 向量与 1 个 $\begin{bmatrix}3\\-1\\2\end{bmatrix}$ 向量及-1 个 $\begin{bmatrix}1\\2\\-1\end{bmatrix}$ 向量的和为 $\begin{bmatrix}6\\-1\\5\end{bmatrix}$ 向量。

抱歉，三维空间的图画得实在不怎么样，好在并不难想象。四维的“图像”当然已经画不出来了。不过按照这个思路，抽象地想象四个四元一次方程组为四维向量的线性关系，甚至推广到 n 并不十分困难。

在数学和物理中都有学习到向量(矢量)的概念，相信大家应该比较熟悉了。在线性代数中，将 n 个有序数 a_1，a_2，…，a_n 所组成的数组称为 n 维向量。这 n 个数称为该向量的 n 个分量，第 i 个数称为第 i 个分量。n 维向量可写成一行，称之为行向量；也可以写成一列，称之为列向量。如前面例中的 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 就是一个三维列向量。若干个同维数的列向量（或行向量）所组成的集合叫做向量组。

再回到我们前面讨论的“方程” $\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\begin{bmatrix}7\\2\end{bmatrix}$，我们把等式右侧的向量改为 $\vec{b}$（当然 $\vec{b}$ 是一个二维向量），于是上式写成 $\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\vec{b}$。$\vec{b}$ 满足什么条件时，方程有解呢？



从图上容易看出，对于任意一个二维向量 $\vec{b}$，过 $\vec{b}$ 的顶点作 $\begin{bmatrix}3\\-4\end{bmatrix}$ 向量的平行线交于

风萧萧兮乱翻书

$\begin{bmatrix}2\\3\end{bmatrix}$向量，其交点所示向量当然与$\begin{bmatrix}2\\3\end{bmatrix}$向量方向相同（或相反），其大小（长度）与$\begin{bmatrix}2\\3\end{bmatrix}$向量的大小（长度）的比值就是方程中要求的 x（若方向相反则为负值）；同样，过$\vec{b}$的顶点作$\begin{bmatrix}2\\3\end{bmatrix}$向量的平行线交于$\begin{bmatrix}3\\-4\end{bmatrix}$向量，其大小（长度）与$\begin{bmatrix}3\\-4\end{bmatrix}$向量的大小（长度）的比值就是方程中的 y。也就是说对于二维空间中的任何一个二维向量，都能找出唯一的（x、y）与之对应。反过来说，通过$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量的线性组合，能够表达出整个二维空间的任意二维向量。我们姑且简单地说，通过$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量的线性组合可以构建一个二维空间（平面）。

定义：给定向量组 A:α_1，α_2，…，α_n和向量β，如果存在一组数λ_1，λ_2，…，λ_n，使得

$$\beta=\lambda_1\alpha_1+\lambda_2\alpha_2+\ldots+\lambda_n\alpha_n$$

则向量β是向量组 A 的线性组合，这时称向量β能由向量组 A 线性表示。

再来看看我们所取的$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量有什么特殊要求吗？或者说，两个二维向量需具备什么条件才能构建二维平面呢？从上面的分析和图像不难理解，只要所取的两个二维向量不在一条直线上（向量方向不相同或相反），那么通过这两个向量的线性组合就可构建一个二维空间。

如果两个二维向量在一条直线上，则两个向量在构建空间中所起的作用是“不独立”的，或者说只有其中一个向量是有效的。我们姑且称其中一个向量为“有效向量”，而另一个称为“冗余向量”。这时，这两个向量的线性组合只能构建二维空间中的一条直线，也就是其线性组合只能表示二维空间中这条直线中的二维向量。而且容易理解，用这两个共线向量线性表达这条直线上的某个二维向量，其线性组合的方式有无穷多个。

风萧萧兮乱翻书

这样我们再来分析一下二元一次方程组$\begin{bmatrix} a1 \\ a2 \end{bmatrix} x + \begin{bmatrix} b1 \\ b2 \end{bmatrix} y = \begin{bmatrix} c1 \\ c2 \end{bmatrix}$的解的情况。当$\begin{bmatrix} a1 \\ a2 \end{bmatrix}$与$\begin{bmatrix} b1 \\ b2 \end{bmatrix}$不共线时，对于任意的$\begin{bmatrix} c1 \\ c2 \end{bmatrix}$都有唯一解；当$\begin{bmatrix} a1 \\ a2 \end{bmatrix}$与$\begin{bmatrix} b1 \\ b2 \end{bmatrix}$共线时，如$\begin{bmatrix} c1 \\ c2 \end{bmatrix}$也在这条直线上，方程有无穷个解，如$\begin{bmatrix} c1 \\ c2 \end{bmatrix}$不在这条直线上，方程无解。

同样的道理，对于三个三维向量，当三个向量不在同一个二维平面（也就是不共面）时，它们可以构建三维空间。通过这三个向量的线性组合可以表达这个三维空间中的任一个三维向量。当三个向量共面但不共线时，“有效向量”为两个，其余一个为“冗余向量”，它们的线性组合只能表示这个三维空间中三个向量所在二维平面上的三维向量，也就是仅能构建这个面所在的二维空间；当三个向量共线时，“有效向量”为一个，其余两个均为“冗余向量”，它们的线性组合只能表示三维空间中这个这条线上的三维向量向量，也就是仅能构建这条线所在的一维空间。

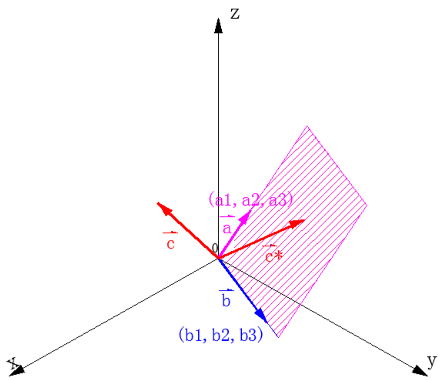
对于方程组$\begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix} x + \begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix} y + \begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix} z = \begin{bmatrix} d1 \\ d2 \\ d3 \end{bmatrix}$ 而言，$\begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix}$、$\begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix}$、$\begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix}$不共面时，对于任何$\begin{bmatrix} d1 \\ d2 \\ d3 \end{bmatrix}$方程有唯一解；$\begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix}$、$\begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix}$、$\begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix}$共面或共线时，当且仅当$\begin{bmatrix} d1 \\ d2 \\ d3 \end{bmatrix}$也在这个平面或直线上时方程有无穷个解，否则方程无解。

与此类推，对于 n 个 n 维向量而言，当 n 个向量不处于同一个 n-1 维空间（也称 n-1 维超平面）时， 这 n 个 n 维向量均为有效向量，可以构建整个 n 维空间。此时，对于 n 个 n 元一次方程而言，方程右侧向量为任意 n 维向量向量都有唯一解；当其中有 m（m＜n）个有效向量时，其余 n-m 个向量为冗余向量，则其仅可构建 n 维空间中的 m 维，也就是其线性组合只能表示 m 维空间所在的 n 维向量，当且仅当方程右侧向量也在这个 m 维空间时方程有无穷解，否则无解。

上面讨论的都是 n 个 n 元一次方程组成的方程组的情况。下面再来看看 m 个 n（n≠m）元一次方程组的情况。

以方程$\begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix} x + \begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix} y = \begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix}$为例，先讨论 m＞n 的情况：

与前面讨论的情况类似，两个三维向量$\begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix}$与$\begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix}$如果不共线，则其可构建三维空间中的一个二维平面，也就是其线性组合可以表示三维空间中的一个二维平面所在的三维向量。如下图所示，当$\vec{c}*$处于这个平面时，方程有唯一解，当$\vec{c}$不处于这个平面时，方程无解。类似的，如果两个三维向量共线，则其仅能线性表示这条直线所在的三维向量。当$\vec{c}$处于这条直线时，方程有无穷解，当$\vec{c}$不处于这条直线时，方程无解。



再以$\begin{bmatrix} a1 \\ a2 \end{bmatrix} x + \begin{bmatrix} b1 \\ b2 \end{bmatrix} y + \begin{bmatrix} c1 \\ c2 \end{bmatrix} z = \begin{bmatrix} d1 \\ d2 \end{bmatrix}$为例，讨论 m＜n 的情况：

三个二维向量$\begin{bmatrix} a1 \\ a2 \end{bmatrix}$、$\begin{bmatrix} b1 \\ b2 \end{bmatrix}$、$\begin{bmatrix} c1 \\ c2 \end{bmatrix}$都共线时，当然只有$\begin{bmatrix} d1 \\ d2 \end{bmatrix}$也在这条直线上时方程才有解，且为无穷个解，否则无解。三个二维向量不都共线时，取其中不共线的两个为有效向量就可线性表示这个二维平面（当然，再多的二维向量也不能构建三维空间），另一个向量就是可用两个有效向量线性表示的冗余向量。这时，无论$\begin{bmatrix} d1 \\ d2 \end{bmatrix}$取何值，方程都有无穷个解。

风萧萧兮乱翻书

§2 线性相关性和向量组的秩

前面我们的分析中，一个重要的判别影响条件是 n 个向量是否不处于同一个 n-1 维以下的空间。这种提法显然只是个定性的提法，如何进行数学的判断呢？

以三个不为零的三维向量$\vec{a}$、$\vec{b}$、$\vec{c}$为例：如果三个向量处于同一个二维空间（也就是共面但不共线），根据前面的分析，三个向量中必有一个是冗余向量，设这个冗余向量为$\vec{a}$（当然，冗余向量并不唯一，可以把任意两个不共线的向量看做有效向量，则剩余一个就为冗余向量）。我们还知道，由另两个有效向量的线性组合可以线性表达这个二维平面，所以这个冗余向量必可由另两个有效向量线性表示，即存在不全为零的 k_1，k_2，使得$\vec{a}=k_1\vec{b}+k_2\vec{c}$。也就是$\vec{a}+(-k_1)\vec{b}+(-k_2)\vec{c}=\vec{0}$。当然，如果三个向量处于同一个一维空间（共线），那么有效向量为一个，冗余向量为两个，同样存在不全为零的 k_1，k_2，使得$\vec{a}+(-k_1)\vec{b}+(-k_2)\vec{c}=\vec{0}$。对于三个三维向量处于同一个二维平面或一维直线中的情况，数学上称这三个向量是线性相关的。

如果三个向量不处于同一个二维空间，则三个向量均为有效向量，三个向量均不能相互线性表达。以$\vec{a}$为例，由于$\vec{a}$不在$\vec{b}$、$\vec{c}$所在二维平面，所以$\vec{a}$无法由$\vec{b}$、$\vec{c}$线性表达。对于这种三个向量不同处于任何同一个二维平面的情况，称三个向量是线性无关的。

数学上对线性相关和线性无关有更加严谨的定义：

定义：给定向量组$\alpha_1,\alpha_2,\ldots,\alpha_m$，如果存在不全为零的数 k_1，k_2，$\ldots k_m$，使

$$k_1\alpha_1+k_2\alpha_2+\ldots+k_m\alpha_m=0$$

*则称该向量组是**线性相关**的，否则称它**线性无关**。*

现在，我们可以从不同的角度来理解向量组的线性相关和线性无关的意义。

从空间的角度来理解，对于三个向量构成的向量组，三个向量不共面，也就是不同处于同一个二维空间时，向量组是线性无关的，当三个向量处于同一个平面（二维空间）或同一条直线（一维空间）时，向量组是线性相关的。推广到 n 个向量组成的向量组，当 n 个向量不同处于 n-1 维空间时，向量组是线性无关的，

风萧萧兮乱翻书

当 n 个向量同处于 n-1 维以下的空间时，向量组是线性相关的。从线性表达的角度理解，当一个向量组中的某一个或某几个向量可以由这个向量组中的其余向量线性表达时，这个向量组是线性相关的；如果其中任意向量都不能由其他向量线性表达，则这个向量组是线性无关的。从有效向量的角度理解，当一个向量组中的向量均为有效向量，冗余向量数量为 0 时，这个向量组是线性无关的；如果向量组中存在至少一个冗余向量，则这个向量组是线性相关的。

现在我们知道三个三维向量如果线性无关，那么这个向量组与原点可以构建三维空间。但它们如果线性相关，它们能够构建的是二维平面还是一维直线呢？或者说它们线性相关到什么程度呢？根据上一节的分析，我们可以定性的说，当这三个向量构成的向量组中有效向量为两个时，它们能够构建的是二维平面；有效向量为一个时，它们能够构建的是一维直线。（*有效向量为三个呢？别耍我了，当然就是整个向量组线性无关，可以构建三维空间了*）。

如何判断一个向量组有多少个有效向量呢？以 n 个向量构成的向量组为例。首先判断整个向量组是否线性无关，如果线性无关，则有 n 个有效向量；如果线性相关，则至少有一个可以由其他向量线性表达的冗余向量。将这个冗余向量去除后，判断剩余的 n-1 个向量是否线性相关。如果线性无关，则有 n-1 个有效向量；如果线性相关，则继续去除一个冗余向量。如此下去，直到找到 r 个有效向量组成的线性无关向量组。这个 r 个有效向量组成的线性无关向量组称为最大线性无关向量组，r 称为向量组的秩。当然，上面的分析是理论上的，实际操作中如何求一个向量组的秩后面还会讨论。

其实，前面所说的“有效向量”、“冗余向量”都是为了便于理解，本人“姑且”自定义的。数学上并没有这个提法。或许从现在开始你就可以忘记它们了。来看看教科书上的向量组的秩的定义吧：

定义　设有向量组 A，如果在 A 中能选出 r 个向量 α_1，α_2，…，α_r　满足：

（1）　向量组 A_0：α_1，α_2，…，α_r 线性无关

（2）　向量组 A 中任意 $r+1$ 个向量（如果 A 中有 $r+1$ 个向量的话），都线性相关

那么称向量组 A_0 是向量组 A 的一个最大线性无关向量组（简称最大无关组）；最大无关组所含向量个数 r 称为向量组 A 的秩，记作秩（A），R（A）或 R_A。

只含零向量的向量组没有最大无关组，规定它的秩为 0。

凤萧萧乱翻书

一个向量组的最大无关组一般并不唯一，因为每次去除的冗余向量（原谅我再次使用本人自定义的名称，希望你没有真的忘的这么快）并不唯一。如对于共面但不共线的三个三维向量而言，其中任意两个向量都可看做有效向量，冗余向量当然也就可看做其中的任意一个。

从前面的分析不难直观的理解，“秩”就是一个向量组中我们所称“有效向量”的个数，也就是这个向量组能够构建的空间的“维数”。

需要特别说明一下的是，虽然前面大部分讨论均以 n 个 n 维向量构成的向量组为例，但线性相关、线性无关和向量组的秩这些概念并不局限于 n 个 n 维向量构成的向量组，对于 m 个 n 维向量构成的向量组同样适用。

来看一个教科书上有关向量组的秩的定理：

定理　若向量组 A 能由向量组 B 线性表示，则有：R（A）≤R（B）；如果向量组 B 也能由向量组 A 线性表示，称 A 与 B 等价，且有：R（A）=R（B）。

这个定理从构建空间的角度应该比较容易理解。如果向量组 A 能由向量组 B 线性表示，那么向量组 B 能构建的空间显然涵盖了向量组 A 能构建的空间，也就是 R（A）≤R（B）；反过来，如果向量组 B 也能由向量组 A 线性表示，则有 R（A）≥R（B），两式都要成立，只有 R（A）=R（B）。

下面再运用“秩”的概念回头看看 m 个 n 元线性方程组的解集情况。

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

我们将 x_1， x_2， …， x_n 组成的列向量记为 x，其对应的系数组成的列向量组记为 A，方程组等式右边的常数项 b_1，b_2，…，b_n 组成的列向量记为 b。上面的方程组可以写成 A x =b。

我们已经知道，当 b 在 A 所能表达的空间时，方程组有解，否则无解。换一个说法就是，若 A 能表达的空间为 m 维，当且仅当 b 也在这个 m 维空间中时方程有解，否则方程组无解。用秩的概念描述就是，若 A 的秩（记为 R(A)）为 m，则将 b 加入到向量组 A 中构建的新的向量组（A，b）的秩（记为 R(A，b)）也为 m 时，即 R(A)= R(A，b)时，方程组有解。如新的向量组（A，b）的秩 R(A， b)=m+1 时，即 R(A)＜R(A，b)时方程组无解。对于有解的情况，当且仅当 A 所表达的空间维数与未知数的数量 n 相等，即 R(A)= R(A，b)=n 时方程组有唯一解。

下面是教科书中的相关定理描述：

定理　n 元线性方程组 A *x* =b

（1）　无解的充分必要条件是 R(A)＜R(A, b)；

（2）　有惟一解的充分必要条件是 R(A)=R(A, b)=n；

（3）　有无限多个解的充分必要条件是 R(A)=R(A, b)＜n.

风萧萧兮乱翻书

§3 向量空间

前面我们用了二维平面、三维空间和 n 维空间的概念来理解有关概念，从向量的角度，我们前面说的这些空间称为“向量空间”。先来看看教科书上的向量空间的定义：

定义：设 V 是 n 维向量的非空集合，如果 V 对向量的线性运算封闭，即：任取α、β∈V，k∈R（R 表示实数集），总有α+β∈V，kα∈V，就称 V 为向量空间。

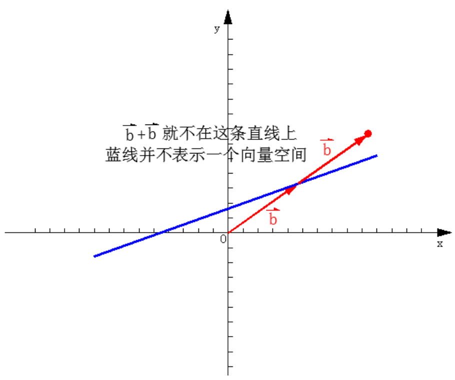
数学上的定义总是严谨、抽象而又枯燥。根据上面的定义可知，三维向量的全体是一个向量空间（通常记为$\mathbb{R}^3$），因为任意两个三维向量的和仍是三维向量，数λ乘三维向量也仍是三维向量。类似地，n 维向量的全体$\mathbb{R}^n$也是一个向量空间。其中，$\mathbb{R}^3$可以看做是几何空间中取定直角坐标系后，其全体三维向量的坐标构成的向量空间。实际上，就可以一般地理解为$\mathbb{R}^3$就是取定一个坐标系的整个三维空间。同样$\mathbb{R}^2$可以一般地理解为取定一个坐标系的二维平面。

特别的，仅由零向量构成的集合也是一个向量空间，称为零空间。因为，对于$\vec{0}$，加法和数乘计算的结果仍为$\vec{0}$，因此它对线性运算封闭。

定义：设 V、W 是两个向量空间，若 $W \subseteq V$，就称 W 是 V 的一个子空间。

任何一个向量空间 V 都必须包含零向量，否则将这个向量空间中的向量数乘 0，得到的零向量不在这个向量空间中，无法满足对线性运算封闭的条件。所以也可以说，零空间是任何一个向量空间的子空间。

需要注意的是，虽然我们可以借助几何中的空间概念来形象理解向量空间，但向量空间与我们几何中空间并不完全相同。



例如，并不是任一条直线、平面都可类比为向量空间。如图中蓝色直线并不代表一个向量空间，因为对于这条直线上的向量$\vec{b}$，乘以 k 或加上这条直线的向量后，新的向量并不在这条直线中了，也就是说不满足对线性运算封闭的要求。所以从几何的角度简单理解，仅有经过原点的直线、平面和空间才可类比为对应的向量空间。这与前面说的任何向量空间都须包含零向量也是一致的。对于三维空间而言，过原点的一个平面或一条直线就是三维空间的一个子空间。

从上面向量空间的定义中，还应该注意到一个概念：“线性运算”。所谓线性运算，就是加法和数乘或其结合，诸如α+3β就是对α、β的线性运算。线性代数研究的运算基本都是线性运算，我们也可以注意到“线性相关”等诸多概念实际也与线性运算密切相关。我想，这也许是线性代数的本质和名称的来源之一吧。

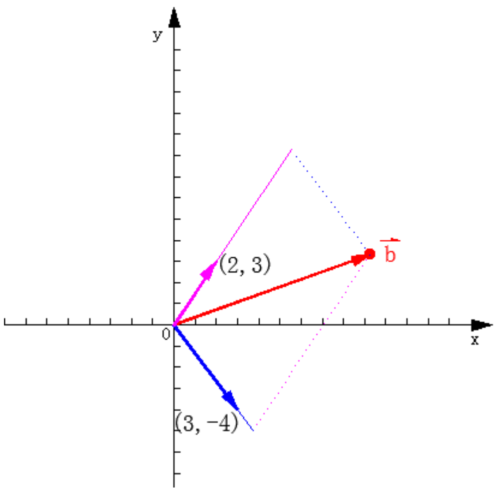
再回到向量，高中的时候老师就教过我们用坐标表示向量，也就是我们现在习惯用的有序数所组成的数组，如[1,2]表示一个二维向量，[1,2,3]表示一个三维向量。还记得用坐标表示向量的原始定义吗？来回顾一下高中数学教科书中平面向量的有关描述：

在平面直角坐标系中，分别取与 x 轴、y 轴方向相同的两个单位向量 i、j 作为“基底”，对于平面内的一个向量α，有且只有一对实数 x、y，使得

$$\alpha = x\,i + y\,j$$

这样，平面内的任意向量α都可由 x、y 唯一确定，我们把有序数（x，y）叫做向量α的坐标，记作α=（x，y）。显然，i=（1,0），j=（0,1）

把上面的式子写成我们刚熟悉的样子，α=$\begin{bmatrix}1\\0\end{bmatrix}x+\begin{bmatrix}0\\1\end{bmatrix}y$。有点感觉了吗？还记得我们第一节讨论的方程$\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\vec{b}$吗？我们已经知道，对于任意一个二维向量$\vec{b}$，都能找到唯一的有序数（x，y）与之对应。（x，y）的直观求法可通过过$\vec{b}$的顶点分别作两个向量的平行线与另一个向量相交，取交点所示向量的大小（长度）与对应向量的大小（长度）的比值确定。这样，我们可以认为通过$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量和原点可构建一个二维向量空间的坐标系，（x，y）就是向量$\vec{b}$在这个坐标系中的坐标！



风萧萧兮乱翻书

比较一下我们常用的平面直角坐标系。在平面直角坐标系中，我们怎么求一个二维向量的坐标（x，y）呢？以求横坐标 x 为例，我的老师教我的方法是：通过这个向量的顶点作 x 轴的垂线，如果垂点位于原点右侧，垂点到原点的长度就是所求的 x 坐标，如果垂点位于原点左侧，垂点到原点的长度取负值就是所求的 x 坐标。现在结合非直角坐标系的做法再换个描述方法：通过所求向量的顶点作 $\begin{bmatrix}0\\1\end{bmatrix}$向量(y 轴方向)的平行线与$\begin{bmatrix}1\\0\end{bmatrix}$向量所在直线（x 轴方向）相交，如交点所示向量与$\begin{bmatrix}1\\0\end{bmatrix}$向量方向相同，其长度与$\begin{bmatrix}1\\0\end{bmatrix}$向量长度的比值就是所求的 x，如交点所示向量与$\begin{bmatrix}1\\0\end{bmatrix}$向量方向相反，其长度与$\begin{bmatrix}1\\0\end{bmatrix}$向量长度的比值的负值就是所求的 x。当然上述两个方法对于我们常用的平面直角坐标系是一样的。现在我们知道，我们常用的平面直角坐标系只是一种特殊的坐标系，特殊在于它所取的两个“基底”垂直，且均为长度为 1 的基本向量。

根据前面的分析我们知道，两个不共线的 n 维向量（现在应该用“线性无关”这个描述了）的线性组合可以表示这两个向量所在二维向量空间中的任一 n 维向量。r 个线性无关的 n 维向量的线性组合可以表示 r 维向量空间内的任一 n 维向量。从坐标的角度我们在来描述一下：以这 r 个线性无关的向量作为“基底”，可构建一个 r 维向量空间的“坐标系”，通过这个“坐标系”可线性表示这个 r 维向量空间中的任一向量。在线性代数中，这个“基底”被称为“基”。

定义　设 V 为向量空间，如果 r 个向量$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r \in V$，且满足：

（1）　$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$ 线性无关

（2）　V 中任一个向量$\boldsymbol{\beta}$都可由$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$ 线性表示：

$$\boldsymbol{\beta}=\lambda_1\boldsymbol{\alpha}_1+\lambda_2\boldsymbol{\alpha}_2+\ldots+\lambda_r\boldsymbol{\alpha}_r$$

那么，向量组$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$ 就称为向量空间 V 的一组基（或一个基），r 称为向量空间 V 的维数，记为维（V）或 dim V，并称 V 为 r 维向量空间。

$(\lambda_1, \lambda_2, \cdots, \lambda_r)$称为向量$\boldsymbol{\beta}$在这组基下的坐标。如果向量空间 V 没有基，那么 V 的维数为 0。0 维空间只含一个零向量。

从几何空间的角度理解，向量空间的一组基就可以看做是这个空间的一个坐标系，这个坐标系有 r 个“独立坐标轴”（线性无关向量），就可以构建 r 维的空间，这个空间中的点都能由这个坐标线性表示。

现在我们知道，平面几何中，常用到直角坐标系体现在二维向量空间$\mathbb{R}^2$中，就是以$\begin{bmatrix}1\\0\end{bmatrix}$、$\begin{bmatrix}0\\1\end{bmatrix}$两个二维向量作为基的情况；同样，立体几何中，常用到直角坐标系体现在三维向量空间$\mathbb{R}^3$中，就是以$\begin{bmatrix}1\\0\\0\end{bmatrix}$、$\begin{bmatrix}0\\1\\0\end{bmatrix}$和$\begin{bmatrix}0\\0\\1\end{bmatrix}$三个三维向量作为基的情况。以此扩展下去，$\mathbb{R}^n$ 当然也存在一组基$\varepsilon_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$，$\varepsilon_2=\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}$，…，$\varepsilon_n=\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$，通常称$\varepsilon_1$，$\varepsilon_2$，…，$\varepsilon_n$为$\mathbb{R}^n$的自然基。

再提醒注意两点：1. 显然一个向量空间的基可以有很多组；2. 基是有序的，就像平面直角坐标系中，(x, y) 是有序的。如果更换一组基中向量的顺序，应该视为另一组基。

下面来看一个教学书上的例题兼定义：

设$\vec{a}$，$\vec{b}$为两个已知 n 维向量，则其线性组合组成的集合

$$L=\{x=\lambda\vec{a}+\mu\vec{b},\ \lambda、\mu\in R\}$$

是一个向量空间。因为若 $x_1=\lambda_1\vec{a}+\mu_1\vec{b}$，$x_2=\lambda_2\vec{a}+\mu_2\vec{b}$，*则有*

$$x_1+x_2=(\lambda_1+\lambda_2)\vec{a}+(\mu_1+\mu_2)\vec{b}\in L$$

$$kx_1=(k\lambda_1)\vec{a}+(k\mu_1)\vec{b}\in L$$

这个向量空间称为由向量$\vec{a}$，$\vec{b}$所生成的向量空间。

有点似曾相识的感觉吧？结合这个“例题”，我们再进一步应用相关“术语”概念整理、理解和完善一下前面的一些认识。

如果$\vec{a}$，$\vec{b}$（仅讨论两个向量都不为零的情况，下同）这两个向量是线性无关的二维向量，那么，它们“所生成”的向量空间可以看做是我们前面说的由其

风萧萧兮乱翻书

构建的“二维平面”，这个“二维平面”就是全部二维向量组成的二维向量空间，即：$\mathbb{R}^2$。$\vec{a}$，$\vec{b}$是这个二维向量空间的一组基。

如果$\vec{a}$，$\vec{b}$是相性相关的两个二维向量，那么它们“所生成”的向量空间是$\mathbb{R}^2$的一个一维子空间，在几何中就是两个向量所在的直线。$\vec{a}$或者$\vec{b}$可看做这个一维向量空间的一个基。

如果$\vec{a}$，$\vec{b}$这两个向量是不共线的三维向量，它们“所生成”的二维向量空间虽然仍可从几何角度看做是一个“平面空间”，但需要注意的是，这个“平面空间”与上面说的$\mathbb{R}^2$所示平面空间并不相同。它是三维空间中的一个平面，从向量空间的角度，生成的是$\mathbb{R}^3$的一个二维子空间。这个子空间中的向量都是三维向量。当然$\vec{a}$，$\vec{b}$仍是这个二维向量空间的一组基。这种情况下，$\vec{a}$，$\vec{b}$是三维向量，在这个向量空间中讨论的也都应该是三维向量，虽然它们生成的向量空间是二维向量空间！换个说法再次重复和强调一下，二维向量空间和$\mathbb{R}^2$并不等同，二维向量空间可能是 n 维向量空间中的二维子空间，里面的向量仍是 n 维向量。而$\mathbb{R}^2$是二维向量的总体组成的二维向量空间。

通过这个例子，我们应该知道注意区别向量的“维”和向量空间的“维”，它们并不一定是一致的。注意一下前面关于向量空间的基和坐标的定义中，没有对向量组（$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_r$）中向量的维数进行描述。这说明$\boldsymbol{\alpha}$可以是大于 r 维的任意维满足条件的向量。当然，由于需满足 r 个向量线性无关的要求，向量的维数不可能小于 r。

对于$\vec{a}$，$\vec{b}$共线及$\vec{a}$，$\vec{b}$为 n 维向量的情况就不再一一描述了。

§4 线性方程组的解集

前面我们讨论了线性方程组有唯一解、无穷解和无解的情况。线性方程组有解时，称线性方程组是相容的；线性方程组无解时，称线性方程组是不相容的。无解和唯一解的情况暂且不谈，方程组有无穷解的情况下，这些解（解的集合又称解集）之间有什么关系呢？或者说，能否用一个通式表示这个解集呢？

线性方程组的一般形式为：

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

当常数项 b_1，b_2，…，b_n 全为零时，方程组称为齐次线性方程组；当常数项不全为零时，称为非齐次线性方程组。

我们先讨论齐次线性方程组的解集情况。用我们熟悉的形式，先以

$\begin{bmatrix}2\\3\end{bmatrix}x_1+\begin{bmatrix}3\\-4\end{bmatrix}x_2+\begin{bmatrix}7\\2\end{bmatrix}x_3=0$ 为例。

我们容易知道，$\begin{bmatrix}2\\3\end{bmatrix}$、$\begin{bmatrix}3\\-4\end{bmatrix}$、$\begin{bmatrix}7\\2\end{bmatrix}$三个向量必定线性相关，但它们显然不共线，所以向量组的秩为 2。任取其中两个线性无关的向量作为一组基，可生成二维向量空间，也可以说，通过这组基可线性表示这个二维向量空间的任意二维向量。我们选取$\begin{bmatrix}2\\3\end{bmatrix}$、$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量作为一组基。既然$\begin{bmatrix}2\\3\end{bmatrix}$、$\begin{bmatrix}3\\-4\end{bmatrix}$两个向量可生成二维向量空间，而余下的向量$\begin{bmatrix}7\\2\end{bmatrix}$乘以任一实数仍是这个二维向量空间的向量，那么方程中$\begin{bmatrix}7\\2\end{bmatrix}$向量的系数 x_3 可以取任何实数，方程都有解。也就是说，其对应系数 x_3 是一个自由变量！一旦 x_3 取定某个数 k，我们都能找到对应的 x_1，x_2，使得$\begin{bmatrix}2\\3\end{bmatrix}x_1+\begin{bmatrix}3\\-4\end{bmatrix}x_2=-k\begin{bmatrix}7\\2\end{bmatrix}$，即方程的解。如取 k=1，即：$x_3$=k=1 时，可解得

风萧萧兮乱翻书

x_1=2，x_2=1，将（x_1，x_2，x_3）表示为列向量形式（称为“解向量”），即为 $x=\begin{bmatrix}2\\1\\1\end{bmatrix}$ 是方程组的一个解（称之为一个特解）。不难求出，取 x_3=k 时，x_1=2k，x_2=k。将其表示为列向量的形式为 $x=\begin{bmatrix}2x_3\\1x_3\\1x_3\end{bmatrix}=x_3\begin{bmatrix}2\\1\\1\end{bmatrix}$，或 $x=k\begin{bmatrix}2\\1\\1\end{bmatrix}$（k∈R）。这就是所求方程组的通解。

再举个例子：$\begin{bmatrix}2\\3\end{bmatrix}x_1+\begin{bmatrix}3\\-4\end{bmatrix}x_2+\begin{bmatrix}7\\2\end{bmatrix}x_3+\begin{bmatrix}8\\-5\end{bmatrix}x_4=0$。容易知道，对应向量组的秩仍为 2，仍取 $\begin{bmatrix}2\\3\end{bmatrix}$、$\begin{bmatrix}3\\-4\end{bmatrix}$ 两个向量作为一组基。同样，我们知道剩余两个二维向量 $\begin{bmatrix}7\\2\end{bmatrix}$、$\begin{bmatrix}8\\-5\end{bmatrix}$ 的任意线性组合仍在这个二维向量空间中，也就是说其对应的系数 x_3，x_4 均为自由变量。令 x_4=0，取 x_3=1，可求得 $x=\begin{bmatrix}2\\1\\1\\0\end{bmatrix}$ 为方程的一个特解，容易知道，$x=k_1\begin{bmatrix}2\\1\\1\\0\end{bmatrix}$ 为方程组的一组解；令 x_3=0，取 x_4=1，可求得 $x=\begin{bmatrix}1\\2\\0\\1\end{bmatrix}$ 也为方程的一个特解，同样，$x=k_2\begin{bmatrix}1\\2\\0\\1\end{bmatrix}$ 也是方程组的一组解。而且，$\begin{bmatrix}2\\1\\1\\0\end{bmatrix}$ 和 $\begin{bmatrix}1\\2\\0\\1\end{bmatrix}$ 必定线性无关（就像 $\begin{bmatrix}c_1\\0\end{bmatrix}$ 和 $\begin{bmatrix}0\\c_2\end{bmatrix}$ (c_1、c_2 均不为零)不可能在同一直线上同理）。

下面顺便插入一个解向量的性质：

解向量性质：若解向量 x=ξ$_1$，x=ξ$_2$ 是方程组 Ax=0 的解，则 x=kξ$_1$（k∈R），x=ξ$_1$+ξ$_2$，x=k$_1$ξ$_1$+k$_2$ξ$_2$（k$_1$，k$_2$∈R）都是该方程组的解。

这个性质很容易证明。仅以 x=k$_1$ξ$_1$+k$_2$ξ$_2$ 为例：

风萧萧兮乱翻书

$Ax= A(k_1\xi_1+k_2\xi_2)=K_1(A\xi_1)+ K_2(A\xi_2)$，根据已知条件可知，$A\xi_1=0$ ，$A\xi_2=0$，所以 $Ax=K_1(0)+ K_2(0)=0$。

回到上面的方程组，根据以上性质可知， $x=k_1\begin{bmatrix}2\\1\\1\\0\end{bmatrix}+k_2\begin{bmatrix}1\\2\\0\\1\end{bmatrix}$ 也是所求方程组的解，这就是所求方程组的通解。

由上面的分析我们可推广知道，对于 m 个 n 元线性方程组成的齐次方程组：

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=0,\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=0,\\\cdots\cdots\cdots\cdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=0,\end{cases}$$

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

若 A 的秩为 r，则可取其中 r 个线性无关的向量为一组基，可生成 r 维向量空间，且 A 中的列向量均在这 r 维向量空间中。若 n=r，前面已经分析过，方程有唯一解（显然解为全零）。若 n 大于 r，则必有 n-r 个自由变量，并可求出 n-r 个线性无关的解向量 ξ_1，ξ_2，…，ξ_{n-r}，$x=k_1\xi_1+ k_2\xi_2+\ldots+ k_{n-r}\xi_{n-r}$（$k_1$，$k_2$，…，$k_{n-r}\in R$）是所求方程组的通解。

如果把方程组的所有解向量视为一个集合，可以发现，这个集合对于线性运算封闭，结果都为$\vec{0}$。根据向量空间的定义，可以看出这个集合为向量空间，而且是个零空间。线性无关的 ξ_1， ξ_2， …， ξ_{n-r} 是这个零空间的最大无关组，或者说可看做这个零向量的一组基，这个零向量是 n-r 维零空间。

还需要说明的是，在上面求解的过程中，可以看出我们选取基和求特解的时候都存在不唯一性，因此可能得到的通解形式不唯一，但它们的集合是等价。从向量空间的角度来说，所求得的这个零空间的基并不一定相同，但这个零空间是同一个。

在此基础上，我们来讨论非齐次线性方程组。

先看看教科书上的两条性质：

性质：（1）设 $x=\eta_1$ 及 $x=\eta_2$ 是方程组 $Ax=b$ 的解，则 $x=\eta_1-\eta_2$ 是对应的齐次线性方程组 $Ax=0$ 的解。

（2）设 $x=\eta$ 是方程组 $Ax=b$ 的解，则 $x=\xi$ 是对应的齐次线性方程组 $Ax=0$ 的解，则 $x=\eta+\xi$ 仍是方程组 $Ax=b$ 的解。

上述性质应该比较明显，在此就不做证明了。根据这个性质我们可以想到，求得 Ax=b 的一个特解后，加上对应的齐次线性方程组 Ax=0 的通解就是所求的方程组 Ax=b 的通解。

风萧萧兮乱翻书

第二章　矩阵及其运算

§1 矩阵

1. 矩阵的定义

讨论线性代数当然离不开矩阵。先来看看矩阵的一系列定义：

定义 由 $m\times n$ 个数 $a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 排成的 m 行 n 列的数表

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

称为 m 行 n 列矩阵，简称 $m\times n$ 矩阵。为表示它是一个整体，总是加一个括弧，并用大写黑体字母表示它，记作：

$$A=\begin{pmatrix} a_{11} & a_{12}\cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2}\cdots & a_{mn} \end{pmatrix}$$

这 $m\times n$ 个数称为矩阵 A 的元素，简称为元，数 a_{ij} 位于矩阵 A 的第 i 行第 j 列，称为矩阵 A 的（i，j）元，以数 a_{ij} 为（i，j）元的矩阵可简记作（a_{ij}）或（a_{ij}）$_{m\times n}$，$m\times n$ 矩阵 A 也记作 $A_{m\times n}$

元素是实数的矩阵称为实矩阵，元素为复数的矩阵称为复矩阵。

行数与列数都等于 n 的矩阵称为 n 阶矩阵或 n 阶方阵。n 阶矩阵 A 也记作 A_n.

只有一行的矩阵 $A=(a_1\ a_2\ \cdots a_n)$ 称为行矩阵，又称行向量。为避免元素间的混淆，行矩阵也记作 $A=(a_1,\ a_2\ ,\ \cdots,\ a_n)$。

只有一列的矩阵 $B=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ *称为列矩阵，又称列向量。*

两个矩阵的行数、列数都相等时，就称它们是同型矩阵。如果两个同型矩阵 A、B 的对应元素都相等，那么就称两个矩阵相等，记作 A=B。

元素都是零的矩阵称为零矩阵。注意不同型的零矩阵是不同的。

这一系列定义好长啊。好在虽然长，倒不晦涩难懂。可是虽然定义不难懂，这好像也太简单了点。矩阵就是一个数表？这也太不“实质”了吧。

风萧萧兮乱翻书

教科书中通常还会举些诸如班级成绩单等日常生活中可看做矩阵的例子，在此就不多说了。

其实可以看出，我们前面已经使用矩阵了，列向量就是一个列矩阵，线性方程组写作 Ax=b 的形式，我们前面称 A 是“列向量组”，现在知道也是一个矩阵。

从向量的角度来看，一个 m×n 矩阵既可看做是 n 个 m 维列向量的组合，也可看做是 m 个 n 维行向量的组合。就像我们前面用到 Ax=b 中的 A 时，其实就是把他看成了列向量的组合。

所以，我们暂且可以将 m×n 矩阵看成是一个由 n 个 m 维列向量组成的向量组或由 m 个 n 维行向量组成的向量组。

2. 矩阵的加法和数乘

矩阵的加法和数乘相对都比较好理解，就不多解释了。直接上教科书定义：

定义 设有两个 $m\times n$ 矩阵 $A=(a_{ij})$ 和 $B=(b_{ij})$，那么矩阵 A 与 B 的和记作 A+B，规定为：

$$
\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}
$$

显然，矩阵加法必须在两个同型矩阵之间进行才有意义，和矩阵是两个同型矩阵对应元素的相加。根据这个定义，不难推出矩阵加法满足下列运算规律：

（1） A+B=B+A； (加法交换律)

（2） (A+B)+C=A+(B+C) (加法结合律)

定义 数λ与矩阵 A 的乘积记作 λA 或 Aλ，规定为：

$$
\lambda\boldsymbol{A}=\boldsymbol{A}\lambda=\begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix}
$$

矩阵数乘运算也比较简单，就是矩阵中所有元素都乘以这个数。根据这个定义，不难推出矩阵数乘满足下列运算规律(λ、μ为数)：

（1） （λμ）A=λ（μA）；

（2） （λ+μ）A=λA+μA；

（3） λ（A+B）= λA+λB

矩阵的加法和数乘运算统称为矩阵的线性运算。应该还记得，我们前面提高过线性运算。

$\lambda=-1$ 时，$\lambda A=-A=\begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix}$。显然，$A+(-A)=0$，$-A$ 称为矩阵 A 的负矩阵，并由此规定矩阵的减法为：$A-B=A+(-B)$.

3. 矩阵的转置和几种常用的特殊矩阵

定义 把矩阵 A 的行换成同序数的列得到的一个新矩阵，叫做 A 的转置矩阵，记作 A^T.

例如：$A=\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 3 & 5 \end{pmatrix}$，$A^T=\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \end{pmatrix}$。

矩阵的转置也是一种运算，根据定义，不难推出矩阵转置满足下列运算规律：

（1） $(A^T)^T=A$;

（2） $(A+B)^T= A^T+ B^T$;

（3） $(\lambda A)^T=\lambda A^T$;

（4） $(AB)^T= B^T A^T$.

线性代数中经常用到几种特殊的矩阵。

满足 $A^T=A$ 的矩阵称为对称矩阵或对称阵。显然，对称阵必须是个方阵，否则 A^T 与 A 根本不同型，更谈不上相等了。直观的理解，对称阵就是关于其“主对角线”（从左上角到右下角的直线称为主对角线，从右上角到左下角的直线称为副对角线，）对称的方阵。

除主对角线外，其余元素都为零的方阵称对角阵，对角阵也记作 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\ldots\lambda_n)$；若对角阵中对角线元素均相等，这个方阵称数量阵（也称纯量阵）；数量阵中，若对角线元素全为 1，这个方阵称为单位阵。可以看出数量阵是对角阵的特款，单位阵又是数量阵的特款。以下依次为对角阵、数量阵、单位阵的示例：

风萧萧兮乱翻书

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix} \quad \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

单位阵通常记作 E 或 **I**（很多字体的“**I**”都会写成“I”，这和 1 区分很不明显，而频繁改字体很麻烦。国内大部分教科书大部分也是用 E，我后面还是用 E 吧）。如为了强调是 n 阶单位阵，也可记作 E_n.

风萧萧兮乱翻书

§2 矩阵的乘法

矩阵的乘法当然是矩阵运算的一个重点。先看看教科书上矩阵乘法的定义吧：

定义　设 $A=(a_{ij})$ 是一个 $m\times s$ 矩阵，$B=(b_{ij})$ 是一个 $s\times n$ 矩阵，那么规定矩阵 A 与矩阵 B 的乘积是一个 $m\times n$ 矩阵 $C=(c_{ij})$，其中

$$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{is}b_{sj}=\sum_{k=1}^{s}a_{ik}b_{kj}$$
$$(i=1,2,\cdots,m;j=1,2,\cdots,n),$$

并把此乘积记作：$C=AB$

我不得不承认，这种定义对我的理解能力是个极大的挑战。当初预习功课到此定义时，我已接近崩溃，后经老师讲解，做习题对照，我才终于学会用两个手指对应指着两个矩阵的相应元素进行计算。唉，往事不堪回首啊……

1. 列线性组合法

好了，忘记那个梦魇吧。来回想一下我们最开始讨论过的方程组 $\begin{cases}2x+3y=7\\3x-4y=2\end{cases}$。开始我们把它写成了 $\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\begin{bmatrix}7\\2\end{bmatrix}$，现在我们再把它写成 Ax=b 的形式：

$\begin{pmatrix}2&3\\3&-4\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}7\\2\end{pmatrix}$。这个等式左侧就是个 2×2 的矩阵乘以 2×1 的矩阵啊！再换个写法，把我们求得的 x=2，y=1 写进去，$\begin{pmatrix}2&3\\3&-4\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix}=2\begin{pmatrix}2\\3\end{pmatrix}+1\begin{pmatrix}3\\-4\end{pmatrix}=\begin{pmatrix}7\\2\end{pmatrix}$。这下够清楚了吧。

再回想下讨论过的另一个线性方程组 $\begin{cases}2x+3y+z=6\\x-y+2z=-1\\x+2y-z=5\end{cases}$，我们曾写为

$\begin{bmatrix}2\\1\\1\end{bmatrix}x+\begin{bmatrix}3\\-1\\2\end{bmatrix}y+\begin{bmatrix}1\\2\\-1\end{bmatrix}z=\begin{bmatrix}6\\-1\\5\end{bmatrix}$，现在同样写成矩阵形式

$$\begin{pmatrix}2&3&1\\1&-1&2\\1&2&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=x\begin{pmatrix}2\\1\\1\end{pmatrix}+y\begin{pmatrix}3\\-1\\2\end{pmatrix}+z\begin{pmatrix}1\\2\\-1\end{pmatrix}=\begin{pmatrix}2x+3y+z\\1x+(-1)y+2z\\1x+2y+(-1)z\end{pmatrix}。$$

风萧萧兮公号翻书

现在我们可以知道，矩阵乘以一个列矩阵就是左侧矩阵的列向量与右侧列向量（列矩阵）的元素对应的线性组合，其结果是与左侧矩阵维数相同（行数相同）的一个列向量（列矩阵）！显然，只有左侧矩阵的列向量的个数（矩阵的列数）与右侧列向量（列矩阵）的维数（列矩阵的行数）相等时，它们才能有对应的线性组合相加，否则两个矩阵无法相乘。

当右侧的矩阵不是一个单列矩阵时，我们可以将其看做是几个列向量的组合，其对应第 i 列的列向量与左侧矩阵相乘的结果为乘积矩阵的第 i 列。

举个例子吧：

求：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}$

解：第一列：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ 2 \end{pmatrix}$

第二列：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = 0\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + 1\begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ 0 \end{pmatrix}$

得：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 11 & 5 \\ 5 & 3 \\ 2 & 0 \end{pmatrix}$

综上所述，矩阵相乘 C=AB 可看做是 A 中的列向量与 B 中各列相应元素线性组合的结果。将 B 中第 i 列的列向量记为 b_i，即 B=[b_1，…，b_i，…b_n]，则有 AB=[Ab_1，…，Ab_i，…Ab_n]。

2. 行线性组合法

我们前面主要是从列向量的角度出发研究相关问题。现在我们知道，一个 m×n 矩阵还可以看做是 m 个 n 维行向量的组合，那么从行向量的角度怎么理解矩阵乘法呢？与前面的做法类似，先举个行矩阵乘法的例子吧。

例：$\begin{pmatrix} 2 & 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix} = 2\begin{pmatrix} 1 & 0 \end{pmatrix} + 3\begin{pmatrix} 2 & 1 \end{pmatrix} + 1\begin{pmatrix} 3 & 2 \end{pmatrix} = \begin{pmatrix} 11 & 5 \end{pmatrix}$

风萧萧兮乱翻书

看出来了吗？首先一个行向量（行矩阵）乘以一个矩阵仍为一行（注意，与列矩阵乘法不同的是，行矩阵在左侧）。可以看做是右侧矩阵各行向量与左侧行向量（行矩阵）对应元素的线性相加。当然，与前面所说的一样，右侧矩阵的行数应与左侧行矩阵的元素数量对应，否则乘法不成立。

从行向量的角度再做一遍上面的例题：

求：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}$

解： 第一行：$\begin{pmatrix} 2 & 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}=2\begin{pmatrix} 1 & 0 \end{pmatrix}+3\begin{pmatrix} 2 & 1 \end{pmatrix}+1\begin{pmatrix} 3 & 2 \end{pmatrix}=\begin{pmatrix} 11 & 5 \end{pmatrix}$

第一行：$\begin{pmatrix} 1 & -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}=1\begin{pmatrix} 1 & 0 \end{pmatrix}-1\begin{pmatrix} 2 & 1 \end{pmatrix}+2\begin{pmatrix} 3 & 2 \end{pmatrix}=\begin{pmatrix} 5 & 3 \end{pmatrix}$

第三行：$\begin{pmatrix} 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}=1\begin{pmatrix} 1 & 0 \end{pmatrix}+2\begin{pmatrix} 2 & 1 \end{pmatrix}-1\begin{pmatrix} 3 & 2 \end{pmatrix}=\begin{pmatrix} 2 & 0 \end{pmatrix}$

得：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}=\begin{pmatrix} 11 & 5 \\ 5 & 3 \\ 2 & 0 \end{pmatrix}$

当然，结果是一样的。

所以，矩阵相乘 C=AB，还可看做是 A 中的各行相应元素与 B 中行向量线性组合的结果。将 A 中第 i 行的行向量记为 a_i，即 A=$\begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix}$，则有 AB=$\begin{pmatrix} a_1B \\ \vdots \\ a_iB \\ \vdots \\ a_nB \end{pmatrix}$。

3. 行×列法

矩阵乘法还能怎么理解呢？一个行向量（行矩阵）乘一个列向量（列矩阵）呢？不论从行向量的角度还是列向量的角度，其结果都是：

风萧萧兮乱翻书

$$\begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3$$

它的结果是一个 1×1 矩阵，也就是个数！仔细看看这个结果，这不是两个向量的内积吗？这也是那个晦涩的矩阵乘法 **C=AB** 的定义中教给我们 c_{ij} 的计算方法，也就是说，c_{ij} 是矩阵 A 的第 i 行与矩阵 B 的第 j 列的向量内积。

按这个思路再计算一遍 $\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}$

解：$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 2\times1+3\times2+1\times3 & 2\times0+3\times1+1\times2 \\ 1\times1-1\times2+2\times3 & 1\times0-1\times1+2\times2 \\ 1\times1+2\times2-1\times3 & 1\times0+2\times1-1\times2 \end{pmatrix} = \begin{pmatrix} 11 & 5 \\ 5 & 3 \\ 2 & 0 \end{pmatrix}$

还有其他的吗？前面试过了行乘列，何不试试列乘行呢？

4. 列×行法

先看个例子，$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{pmatrix}$，显然得到的是 3×2 的一个矩阵。两个矩阵各加一列和一行呢，如 $\begin{pmatrix} 2 & 1 \\ 3 & 3 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 12 & 9 \\ 10 & 10 \end{pmatrix}$，显然还是个 3×2 的矩阵，这两个 3×2 的矩阵有什么联系吗？这之间的变化是因为各新增了一列和一行引起的，不妨将新加的一列和新加的一行乘起来看看：$\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}\begin{pmatrix} 3 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 9 & 3 \\ 6 & 2 \end{pmatrix}$。哇，我发现了，$\begin{pmatrix} 5 & 5 \\ 12 & 9 \\ 10 & 10 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{pmatrix} + \begin{pmatrix} 3 & 1 \\ 9 & 3 \\ 6 & 2 \end{pmatrix}$。其实，仔细研究一下计算过程就能得出，矩阵乘法 AB 还可以看成矩阵 A 的第 i 列与矩阵第 i 行（i 从 1 到 A 的最后一列（B 的最后一行））的乘积的矩阵和。

还是同样的例子：求 $\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{pmatrix}$

风萧萧兮乱翻书

解： 第一列×第一行：$\begin{pmatrix}2\\1\\1\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix}=\begin{pmatrix}2 & 0\\1 & 0\\1 & 0\end{pmatrix}$

第二列×第二行：$\begin{pmatrix}3\\-1\\2\end{pmatrix}\begin{pmatrix}2 & 1\end{pmatrix}=\begin{pmatrix}6 & 3\\-2 & -1\\4 & 2\end{pmatrix}$

第三列×第三行：$\begin{pmatrix}1\\2\\-1\end{pmatrix}\begin{pmatrix}3 & 2\end{pmatrix}=\begin{pmatrix}3 & 2\\6 & 4\\-3 & -2\end{pmatrix}$

得：$\begin{pmatrix}2 & 3 & 1\\1 & -1 & 2\\1 & 2 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\2 & 1\\3 & 2\end{pmatrix}=\begin{pmatrix}2 & 0\\1 & 0\\1 & 0\end{pmatrix}+\begin{pmatrix}6 & 3\\-2 & -1\\4 & 2\end{pmatrix}+\begin{pmatrix}3 & 2\\6 & 4\\-3 & -2\end{pmatrix}=\begin{pmatrix}11 & 5\\5 & 3\\2 & 0\end{pmatrix}$

现在我们已经讨论了四种方法做矩阵乘法了，我们姑且称之为“列线性组合法”、“行线性组合法”、“行乘列法”和“列乘行法”。还有吗？

确实至少还有一种（其他还有没有我就实在不知道了）——分块法。

5. 分块运算

分块运算其实并不只是矩阵乘法的一种方法，它是矩阵运算的一个重要方法。

一个大的矩阵，可以在行之间、列之间插入一些贯穿始终的横线、纵线，将矩阵分割为若干小矩阵，这些小矩阵称为子块，被分割后的矩阵称为分块矩阵。

如：4×5 矩阵 A=$\begin{pmatrix}3 & 0 & 0 & 0 & 0\\0 & 3 & 0 & 0 & 0\\2 & 0 & 0 & 2 & 3\\4 & 2 & 1 & 0 & 2\end{pmatrix}$，可以分割为：

$$\left(\begin{array}{cc|ccc}3 & 0 & 0 & 0 & 0\\0 & 3 & 0 & 0 & 0\\\hline 2 & 0 & 0 & 2 & 3\\4 & 2 & 1 & 0 & 2\end{array}\right) \qquad \left(\begin{array}{cc|c|cc}3 & 0 & 0 & 0 & 0\\0 & 3 & 0 & 0 & 0\\\hline 2 & 0 & 0 & 2 & 3\\4 & 2 & 1 & 0 & 2\end{array}\right)$$

等等。

记 $A_1=\begin{pmatrix}2 & 0\\4 & 2\end{pmatrix}$，$A_2=\begin{pmatrix}0 & 2 & 3\\1 & 0 & 2\end{pmatrix}$，$E_2=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$，则 A 可表示为分块矩阵

风萧萧兮乱翻书

$A=\begin{pmatrix} 3E_2 & 0_{2\times 3} \\ A_1 & A_2 \end{pmatrix}$。可以看出，巧妙的运用分块矩阵可以将大矩阵分为若干含有零矩阵、数量阵或单位阵等特殊小矩阵的组合，这对于工程中经常遇到的大型而其中又有大块 0 块的矩阵而言，可以大大减少存储空间和简化计算。

下面简单介绍下分块矩阵的运算，当然也包括分块乘法。

1）分块矩阵的转置和数乘

分块矩阵的转置和数乘只涉及一个矩阵，因此对矩阵的分块方式没有要求。以含 4 个子块的分块矩阵为例：

转置：$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^T = \begin{pmatrix} A^T{}_{11} & A^T{}_{12} \\ A^T{}_{21} & A^T{}_{22} \end{pmatrix}$

数乘：$\lambda\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} \lambda A_{11} & \lambda A_{12} \\ \lambda A_{21} & \lambda A_{22} \end{pmatrix}$

2）分块矩阵的加法

根据矩阵加法的定义，两个同型矩阵才能相加。所以 A 与 B 必须同型，而且，对 A、B 的分块方式也需相同，以保证对应的子块矩阵同型。

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{pmatrix}$$

3）分块矩阵的乘法

根据矩阵乘法的定义，计算矩阵乘法 AB 时，A 的列数须与 A 的行数相等。运用分块矩阵乘法时，对 A、B 的分块也需保证相应分块矩阵满足行数等于列数的要求。进行分块矩阵的计算时，计算公式与将子块当做一个“数”的乘法类似。

即$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{21}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{21}+A_{22}B_{22} \end{pmatrix}$

好了，矩阵乘法我们所知的就这些了。

到此，我们应该似乎有点体会到矩阵乘法这样规定的奥妙了，更应该明白矩阵的乘法也是线性运算、线性组合的本质了。

来稍微归纳总结下矩阵的乘法吧。从前面讨论的每一种方法中我们也能再次的确认，只有 A 的列与 B 的行数量相等时，矩阵乘法 AB 才有意义，它们的乘

风萧萧兮乱翻书

积矩阵是一个行数与左矩阵 A 行数相等，列数与右矩阵 B 列数相等的新矩阵。即 $A_{m\times n}B_{n\times p}=C_{m\times p}$. 所以，在矩阵的乘法中必须注意矩阵相乘的顺序。AB（称为 A 左乘 B，或 B 右乘 A，或 B 被 A 左乘）与 BA 不能相混淆，AB 有意义时，BA 不一定有意义，即便有意义，一般情况下也并不相等。由此我们也知道了，矩阵乘法不满足交换律。

矩阵的乘法虽不满足交换律，但满足以下定律（在其乘法有意义的前提下）：

（1） $(AB)C=A(BC)$；（乘法结合律）

（2） $A(B+C)=AB+AC$；　　$(B+C)A=BA+CA$；　（乘法分配律）

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

（4） $(AB)^T=B^TA^T$；

$(A_1A_2\cdots A_s)^T=A_s^T A_{s-1}^T\cdots A_2^T A_1^T$

（5） $(\lambda E_m)A_{m\times n}=\lambda A_{m\times n}=A_{m\times n}(\lambda E_n)$

还有一个例子值得一提：设 $A=\begin{pmatrix}-2 & 4\\ 1 & -2\end{pmatrix}$，$B=\begin{pmatrix}2 & 4\\ -3 & -6\end{pmatrix}$，分别求 AB 和 BA。

容易解得:$AB=\begin{pmatrix}-2 & 4\\ 1 & -2\end{pmatrix}\begin{pmatrix}2 & 4\\ -3 & -6\end{pmatrix}=\begin{pmatrix}-16 & -32\\ 8 & 16\end{pmatrix}$,

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

这个例子在进一步告诉我们矩阵乘法不满足交换律的同时，还告诉我们两个不为零矩阵的矩阵相乘结果却可能是零矩阵。所以，对矩阵而言，由 AB=0，不能得出 A、B 中至少有一个为零矩阵的结论。也不能由 A（B-C）=0，且 $A\neq 0$ 就推出 B-C=0，从而得出 B=C 的结论。

有了矩阵的乘法，还可以定义矩阵的幂。设 A 是 n 阶方阵，定义

$$A^1=A，A^2=AA，\ldots，A^{k+1}=A^kA \quad（k 为正整数）$$

当然，只有 A 为方阵它的幂才有意义。容易知道方阵的幂计算满足：

$$A^{k+t}=A^kA^t\ ,\ (A^k)^t=A^{kt}$$

风萧萧兮乱翻书

§3 矩阵的秩

矩阵到底是什么？谁知道呢！也许从不同的角度，可以得出不同的涵义吧。那就先从我们熟悉的开始，把 $m\times n$ 矩阵看做是 n 个 m 维列向量的向量组，或者是 m 个 n 维行向量的向量组吧。

谈到向量组，当然离不开向量组的“秩”。既然矩阵既可以看做是列向量组成的向量组，也可以看做是行向量组成的向量组，自然而然地，矩阵就应该会有两个秩了？事实确实如此，矩阵的行向量组的秩称为矩阵的“行秩”，矩阵的列向量组的秩称为矩阵的“列秩”。

两个秩？似乎有点麻烦。好在教科书上的定理告诉我们，矩阵的行秩和列秩是相等的，也就是说矩阵的“两个秩”其实就是一个！这太好了！可为什么会这样呢？教科书上的证明有的是从行列式的角度证明的，有的是从矩阵初等变换的角度证明的，这些按照我们的编排还没讨论到，后面讨论到以后大家可以自己对照学习。当然，也有教科书干脆写“严格证明略”！其实我上学的时候挺喜欢看到“严格证明略”这样的字眼的，这样老师会告诉你，你只需要知道这条性质、定理的结论，没必要知道为什么会这样。最关键是考试不会考到它（此处省略感慨 n 个字）！

不过，我们还是尝试着结合我们前面讨论过的向量组的有关性质和矩阵乘法的理念来探究一下，顺便再复习、巩固下矩阵乘法“线性组合”的本质吧。

结合一个例子进行描述：对于一个 $m\times n$ 矩阵 A，设其列秩为 r。以 $A=\begin{pmatrix}1&2&2\\2&3&4\\3&4&6\end{pmatrix}$为例吧。举例的这个矩阵，从列向量的角度能够很明显发现第一列和第三列成比例，当然是线性相关的，可知列秩为 $r=2$。从行向量的角度很难一眼看出是否线性相关，暂且不管。取 A 中的最大无关列向量组构成个新的矩阵 B，由列秩为 r 可知 B 为 $m\times r$ 矩阵。本例中 $B=\begin{pmatrix}1&2\\2&3\\3&4\end{pmatrix}$。既然 B 的列是取自 A 中的最大无关列向量组，当然通过 B 中列向量的线性组合能够表达 A 中所有的列向量。还记得矩阵乘法“列线性组合法”吗？这就是说必然存在一个 $r\times n$ 矩阵 C，

风萧萧兮乱翻书

使得 BC=A。在这个例子中，C=$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}$。再换个角度，从“行线性组合法”的角度看 BC=A 这个等式。我们可以说 A 中的行向量都是 C 中行向量的线性组合。既然 A 的行向量是由 C 的行向量线性表示的，再回忆一下向量组的秩中，我们介绍的一个定理：“若向量组Ⅰ能由向量组Ⅱ线性表示，则有：R（Ⅰ）≤R（Ⅱ）”，所以 A 的行秩≤C 的行秩。而 C 只有 r 行，也就是只有 r 个行向量，C 的行秩当然最多也就等于 r 了。由此，我们得出一个重要结论：A 的行秩≤r，即 A 的行秩≤A 的列秩！类似地可以证明，A 的列秩≤A 的行秩。所以，矩阵的行秩和列秩只能相等！

教科书上的矩阵的秩的定义不尽相同，同济版的定义还使用了行列式的有关概念，我们还没对行列式进行讨论，而且那个定义实在拗口，在此就不摘录了。反正现在我们知道矩阵的行秩，矩阵的列秩，组成矩阵的行向量组的秩，列向量组的秩统统都是一个秩了。真是“爱就一个秩”啊！

如果一个矩阵的秩等于它的列数，那么称这个矩阵为列满秩矩阵；如果一个矩阵的秩等于它的行数，那么称这个矩阵为行满秩矩阵。如果一个矩阵的秩既等于它的列数，也等于它的行数，（这个矩阵当然肯定是个方阵），那么称这个矩阵为满秩矩阵。从向量组和向量空间的角度来说，对于一个 n 阶满秩矩阵，就是这个矩阵的行向量组线性无关，列向量组也线性无关，以其行向量组或列向量组都可以作为 n 维向量空间的一组基，其线性组合可生成整个 n 维向量空间。

结合上面的分析过程和结果，我们归纳一下矩阵秩的一些基本性质：

（1）$0 \leq R(A_{m\times n}) \leq \min\{m, n\}$；

（2）$R(A^T)=R(A)$；

（3）$\max\{R(A),R(B)\} \leq R(A, B) \leq R(A)+R(B)$

$R(A, B)$是把 B 的列放在 A 的后面产生的新的矩阵的秩。从向量组的秩的角度很容易理解，这个新矩阵的秩当然不小于 A、B 两个矩阵秩的任何一个，当然也不会大于 A、B 两个矩阵秩的和。严格的数学证明就不写了，大家可以参见教科书。（下同）

（4） $R(A+B) \leq R(A)+R(B)$

风萧萧兮乱翻书

A+B 是当然可以看做是 A、B 的线性组合，所以（A+B）的秩当然不会大于 A 的秩与 B 的秩的和。

（5） R(AB)≤min{R(A),R(B)}

这个性质我们在上面的分析过程中已经用了一部分了。从矩阵乘法的角度，AB 既可以看做 A 的列向量的线性组合，也可以看做 B 的行向量的线性组合，那么，AB 的秩当然不会大于 A 的列秩，也不会大于 B 的行秩。既然矩阵的列秩、行秩是一个秩，那么 AB 的秩当然不会大于 A 和 B 的秩的任意一个。

风萧萧兮乱翻书

§4 逆矩阵

前面我们讨论了矩阵的加法、减法和乘法运算，那矩阵有没有除法运算呢？我们回顾一下减法运算的引入，引入负矩阵的概念后，减法可看做是加法的一种变形，A−B 也就是 A+(−B)了。这与数的减法 6−2 在引入负号后变成 6+(−2)是一样的。你看减号和负号看上去都是一样的。那除法呢？6÷2 能化为乘法运算吗？当然可以，$6\div2=6\times\frac{1}{2}$ 嘛！你可能会说，这不就是变成乘以它的倒数嘛。是的，这当然很简单。跟据百度百科的描述，倒数是指数学上设一个数 x 与其相乘的积为 1 的数，记为 $\frac{1}{x}$ 或 x^{-1}，过程为"乘法逆"！ 而且，每个复数（实数）只有一个倒数。

矩阵有“倒数”吗？也就是说对于一个矩阵 A，会有另一个矩阵 B，使得 AB=1 吗？这当然太局限了，我们知道两个矩阵的乘法结果是一个矩阵，除非 A 是列矩阵，而 B 是个行矩阵，结果才会是个一维矩阵，也就是个数。不过，矩阵中有个特殊矩阵——单位阵 E，它的性质和 1 非常相像。1 乘任何数都等于这个数本身，而单位阵 E 乘一个矩阵（当然需满足矩阵乘法要求，有意义才行）也等于这个矩阵本身。那有没有一个矩阵 B，使得 AB=E 呢？这就是逆矩阵的概念：

定义 对于 n 阶矩阵 A，如果有一个 n 阶矩阵 B，使得

$$AB=BA=E$$

则说矩阵 A 是可逆的，并把矩阵 B 称为 A 的逆矩阵，简称逆阵。A 的逆阵记为 A^{-1}

看来还真有矩阵的“倒数”。和倒数一样，如果矩阵 A 是可逆的，那么 A 的逆矩阵是唯一的。因为：设 B、C 都是 A 的逆矩阵，AB=BA=EA，C=CA=E，于是有 B=BE=B(AC)=(BA)C=EC=C。

根据定义我们容易得出，如果 B 是 A 的逆矩阵，那么 A 也是 B 的逆矩阵，也就是 A、B 是互逆的。这与 2 的倒数是 $\frac{1}{2}$，$\frac{1}{2}$ 的倒数是 2 也是类似的。

当然，矩阵的乘法与数还是有些不一样，比如对于两个数 a、b，如果 a×b=1，根据乘法交换律必有 b×a =1，而我们前面讨论过矩阵乘法不满足交换律。可能

风萧萧兮乱翻书

存在两个矩阵 $A_{(m\times n)}$、$B_{(n\times m)}$，使得 $A_{(m\times n)}B_{(n\times m)}=E_m$，但 $B_{(n\times m)}A_{(m\times n)}$却没有意义。所以，我们这里讨论的逆矩阵仅针对方阵（非方阵的 $A_{(m\times n)}B_{(n\times m)}=E_m$ 的情况称左逆、右逆，这里不展开讨论）。对于方阵而言，如果 AB=E，则必有 BA=E。当然这不是说方阵就满足乘法交换律，而仅是逆矩阵的性质。

来看一下逆矩阵的几条运算性质：

（1） 若 A 可逆，则 A^{-1} 也可逆，且 $(A^{-1})^{-1}=A$；

（2） 若 A 可逆，数 $\lambda\neq 0$，则 λA 可逆，且 $(\lambda A)^{-1}=\frac{1}{\lambda}A^{-1}$；

（3） 若 A,B 为同阶矩阵且均可逆，则 AB 也可逆，且 $(AB)^{-1}=B^{-1}A^{-1}$

把性质 3 的证明思路稍微提一下。通常要证明两个矩阵是互逆矩阵，首先想到的是证明它们的乘积等于单位阵。$(AB)B^{-1}A^{-1}=A(BB^{-1})A^{-1}=AE\,A^{-1}=E$.

除了乘积的逆的性质中需要把两个矩阵的顺序颠倒一下以外，上面提的这些和倒数都很像吧，连符号 A^{-1} 都是类似的。我们知道，非零的实数都有倒数，那么是不是非零的矩阵都有逆矩阵呢？很不幸的是，这点逆矩阵和倒数的性质不一样了。

那么什么方阵才可逆呢？还是从前面讨论过的矩阵的乘法和向量空间中寻找答案吧。如果有 n 阶方阵 AB=E，我们知道 E_n 是 $\mathbb{R}^n$ 的自然基，那么 E 的秩必然就是 n。根据前面分析过的两个矩阵乘积 AB 的秩不大于两个矩阵的任何一个秩的性质，有 $R(E_n)\leqslant R(A)$，$R(E_n)\leqslant R(B)$。也就是说，A、B 的秩都不小于 n。而 A、B 都是 n 阶方阵，当然其秩不可能大于 n。所以，A、B 的秩都是 n。也就是说，只有 A、B 都是满秩方阵时，才能使得 AB=E。用数学语言描述就是，一个矩阵可逆的必要条件是这个矩阵满秩。

那么如果 A 满秩，它是否一定有逆矩阵呢？或者说，是充分条件吗？

还记得 Ax=b 这个方程吗？如果 A 满秩，也就是 A 中的列向量线性无关，我们知道，对于任一 n 维列向量，方程都有唯一解。分别取 E 的列向量，也就是分别令 $b_1=(1,0,\dots,0)^T$ (这是列向量 $\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$ 的另一种写法，大家应该不会疑惑吧)、$b_2=(0,1,0,\dots,0)^T$，…，$b_n=(0,0,\dots,1)^T$ 可分别求得 x 的解向量

风萧萧兮乱翻书

x_1，x_2，…，x_n，将这 n 个解向量组成矩阵 B，显然 AB=E。当然，如果你对矩阵乘法的线性组合的本质理解的比较透彻，上面的解释就更简单了。总而言之，如果 A 满秩，必然存在另一个矩阵 B，使得 AB=E。

综上所述，一个矩阵可逆的充分必要条件是这个矩阵满秩。而且，我们从理论上知道，这个矩阵的逆矩阵可以通过解 n 个 $Ax=\varepsilon_n$ 方程来求得。由此我们也知道，零矩阵没有逆矩阵（这点和倒数也一致）。

前面我们说过，对矩阵而言，由 AB=0，不能得出 A、B 中至少有一个为零矩阵的结论。也不能由 A（B-C）=0，且 A≠0 推出 B-C=0，从而得到 B=C 的结论。但如果 A 是可逆矩阵（既然可逆 A 当然就不是零矩阵），则由 Ax=0 只有零解可得，AB=0 必有 B=0 的结论。当然，如果有 A（B-C）=0，且 A 可逆，则可以推断 B-C=0，B=C。

不过，现在我们好像还是没有完全解答本节最初提出的问题，矩阵有除法吗？或者换个说法，对于矩阵 A、B、C 而言，若已知 A，C 并且有 AB=C，能求出 B 吗？我们知道，对于数 ab=c 而言，如果 a 有倒数（即 a≠0），等式两边同乘以 a 的倒数可得 b=c/a。如果 a 没有倒数（a=0），则只有 c=0 时，b 为任意数（无穷解），否则 b 无解。对于矩阵而言，如果 A 可逆，则可等式两边同时右乘 A^{-1}，得到 $A^{-1}AB= A^{-1}C$，即 $B= A^{-1}C$（需注意矩阵乘的方向）。如果 A 不可逆呢？这似乎有点麻烦。还记得矩阵乘法的“列线性组合法”吗，将 B 中第 i 列的列向量记为 b_i，则有 AB=[Ab_1，…，Ab_i，…Ab_n]，相应地将 C 中第 i 列的列向量记为 c_i，不就是相当于已知 A 和 c_i，要求 $Ab_i=c_i$ 中的 b_i 吗？根据前面讨论过的线性方程组解的情况可知，A 不可逆时，b_i 的解与 c_i 有关，可能无解或有无穷解。于是，当 b_1，…，b_i，…b_n 都有无穷解时，B 有无穷解，其中有一个无解，则 B 无解。

以上讨论的是方阵的情况，不是方阵的 AB=C 解的情况就不详细讨论了，有兴趣的也可以按此思路自己研究一下。

风萧萧兮乱翻书

§5 矩阵初等变换及应用

前面我们大量讨论的大部分是定性的分析、理解。可真给我们一个向量组，或者说给一个矩阵，我们怎么求得它的秩呢？有没有简单易行的办法呢？

还是从线性方程组开始吧。回到我们最开始讨论的$\begin{cases}2x+3y+z=6\\x-y+2z=-1\\x+2y-z=5\end{cases}$，现在知道还可以写成 Ax=b 的矩阵形式：$\begin{pmatrix}2&3&1\\1&-1&2\\1&2&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}6\\-1\\5\end{pmatrix}$。我们用中学代数中熟知的消元法不难解出答案。对其中的方程进行若干的数乘和加减（又是“数乘和加减”!），逐步消去未知数的系数（使得其为 0），最后形成一个阶梯型的方程组。过程我就不详细写了，倒不是因为我认为这太简单，更不是我认为这不重要（恰恰相反，矩阵初等变换对线性代数中的运算十分重要），也不是我想偷懒，关键是描绘这个过程需要的阶梯型方程组似的排版，对用 word 写这些的我来说实在是个不可能完成的任务。好在教科书上“矩阵的初等变换”相关章节通常都会以一个例子写出详细的过程，理解也不难。好了，我休息一会，你去看教科书吧！

……

看了教科书，你当然会知道，解线性方程组的过程只是对未知数的系数和常数项进行了运算，从矩阵的角度来说，就是对 A，b 组合的矩阵[A，b]进行变换。当然你还会知道矩阵[A，b]称为增广矩阵。你还应该知道对矩阵的行进行①换位变换（更换某两行的位置），②倍法变换（某行乘以一个不为 0 的数λ），③倍加变换（某行加上另一行的λ倍替换该行），统称矩阵的初等行变换（几种变换的称谓可能不完全一致，当然这应该不会影响理解）。类似地还有初等列变换。矩阵的初等行变换、初等列变换统称矩阵的初等变换。还有什么是行阶梯阵、行最简形什么的…

这一节你还是看教科书吧，我没什么可写的了。

……

好像有点不对啊，我们这一节开始说是讨论怎么用简单易行的方法求矩阵的秩啊，怎么支到教科书上“矩阵的初等变换”去了就不回来了。(╰_╯)

……

好吧。还是继续吧！ -_-!

风萧萧兮乱翻书

矩阵的初等行变换（以初等行变换为例，初等列变换类似）其实是对行向量进行线性变换，而且是可逆的（这里的可逆不是矩阵可逆，是变换过程可反推回去。如变换过程不可逆对于方程组而言就不是同解变换了。对了，这里还是再提醒一下，初等行变换过程中的换位变换、倍法变换都还好，倍加变换要特别注意是 i 行加上 j 行的λ倍替换 i 行，而不能替换 j 行）。也就是说初等行变换后矩阵的所有行向量都能用变换前矩阵的行向量线性表示，变换前矩阵的所有行向量也都能用变换后矩阵的行向量线性表示。两者的行向量组是等价的（顺便复习一下向量组等价的概念）。对于矩阵而言，也称两个矩阵行等价。（矩阵列等价及矩阵等价的概念类似。两个矩阵等价也可从矩阵初等变换的角度理解，即两个经有限次的初等变换可以相互变换的矩阵等价）。从秩的角度来说，就是矩阵的初等变换不会改变矩阵的秩。任何矩阵经过有限次的初等变换都可以变换成行阶梯阵和行最简形（希望你真的去看了教科书），而行阶梯阵或行最简形矩阵的秩就是非零行的个数！

关于行阶梯阵的秩就是非零行的个数可以再多说一下。

对于诸如$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$的行阶梯矩阵而言，有三个非零行。而且对于前三行的行向量而言，显然无法相互线性表示，所以三个非零行的行向量线性无关，矩阵的秩为 3.

一个矩阵的初等行变换既然是对矩阵行向量的线性变换，从矩阵乘法之“行线性组合法”的本质来理解，其实质可看做是这个矩阵右乘了另一个“变换方阵”。设一个 m×n 矩阵 A 经初等行变换后为 m×n 矩阵 B，实质可看做是 A 右乘了一个 m×m 的变换方阵 P，即 PA=B。由于矩阵初等行变换的过程是可逆的，B 经初等行变换可变回 A，也就是说有一定存在另一个 m×m 的变换方阵 Q，使得 QB=A。将 A=QB 代入 PA=B，有 P(QB)=B，即(PQ)B=B，于是有 PQ=E。由此可知，矩阵的初等行变换对应的变换矩阵是可逆的。把以上的分析综合一下：如果一个矩阵 A 能通过初等行变换变为 B，则一定存在一个可逆矩阵 P，使得 PA=B；如果存在一个可逆矩阵 P，使得 PA=B，那么 A 一定能通过初等行变换变为 B。

风萧萧兮乱翻书

同样，m×n 矩阵的初等列变换的实质是矩阵左乘一个 n×n 方阵。类似地可知，同理可知，如果一个矩阵 A 能通过初等列变换变为 B，则一定存在一个可逆矩阵 Q，使得 AQ=B；如果存在一个可逆矩阵 Q，使得 AQ=B，那么 A 一定能通过初等列变换变为 B。

用数学语言描述就是：

定理：设 A、B 为 m×n 矩阵，那么：

（1） A 和 B 行等价的充分必要条件是存在 m 阶可逆矩阵 P，使得 PA=B；

（2） A 和 B 列等价的充分必要条件是存在 n 阶可逆矩阵 Q，使得 AQ=B；

（3） A 和 B 等价的充分必要条件是存在 m 阶可逆矩阵 P 及 n 阶可逆矩阵 Q，使得 PAQ=B；

等一下！以前我们说过矩阵可以看成向量组，刚才我们又说矩阵初等变换的实质是乘另一个矩阵，那么矩阵不是也可以看做是一种线性变换？这个…，先留着它吧！

矩阵的初等变换的过程并不复杂，但确实是线性代数计算中简单而又最广泛的基础。其实很多线性代数的计算都会用到它，计算机相关线性计算的程序计算时相当多的部分也是依靠初等变换。除了求矩阵的秩以外，再以求逆矩阵为例窥其一斑。

一个方阵 A 可逆的话，经过初等变换必定可以变换成单位阵 E，显然它的逆矩阵 $B=A^{-1}$ 就是变换方阵。（顺便得到一个推论：方阵 A 可逆的充要条件是 A 和 E 等价）。那么对单位阵进行同样的变换，得到的矩阵就是 BE，也就是 B。换句话说，对 A 和 E 进行同样的初等变换，当 A 变换成 E 时，E 就变换成了 A 的逆矩阵。

举个例子，求 $A=\begin{pmatrix}1 & 0 & 0\\ 2 & 2 & 5\\ 0 & 1 & 3\end{pmatrix}$ 的逆矩阵。将 E 放在矩阵右侧，与 A 同步进行初等行变换。

$$\begin{pmatrix}1 & 0 & 0 & 1 & 0 & 0\\ 2 & 2 & 5 & 0 & 1 & 0\\ 0 & 1 & 3 & 0 & 0 & 1\end{pmatrix} \sim ③ \rightleftarrows ② \begin{pmatrix}1 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 3 & 0 & 0 & 1\\ 2 & 2 & 5 & 0 & 1 & 0\end{pmatrix} \sim ③-2\times①$$

风萧萧兮乱翻书

$$\begin{pmatrix}1&0&0&1&0&0\\0&1&3&0&0&1\\0&2&5&-2&1&0\end{pmatrix}\sim ③-2\times ② \begin{pmatrix}1&0&0&1&0&0\\0&1&3&0&0&1\\0&0&-1&-2&1&-2\end{pmatrix}\sim ②+3\times ③$$

$$\begin{pmatrix}1&0&0&1&0&0\\0&1&0&-6&3&-5\\0&0&-1&-2&1&-2\end{pmatrix}\sim ③\times(-1)\ \begin{pmatrix}1&0&0&1&0&0\\0&1&0&-6&3&-5\\0&0&1&2&-1&2\end{pmatrix}$$

经过一系列初等行变换，A 变换成了 E，E 变换成了$\begin{pmatrix}1&0&0\\-6&3&-5\\2&-1&2\end{pmatrix}$，这就是 A 的逆矩阵。

风萧萧兮乱翻书

第三章　行列式

§1 有向面积和有向体积

高等数学中我们学习过向量的点乘（内积、数量积）和叉乘（外积、向量积）等概念，前面讨论矩阵乘法的时候，也提到行向量乘列向量其实是两个向量的数量积（内积）。现在我们要讨论向量的向量积（外积）。

根据向量积的定义，两个向量$\vec{a}$、$\vec{b}$的向量积$\vec{a}\times\vec{b}$为一个新向量$\vec{c}$，$\vec{c}$的大小（模）$|\vec{c}|=|\vec{a}||\vec{b}|\sin\theta$，其中θ为两个向量的夹角，$\vec{c}$的方向为按照右手法则确定的既垂直于$\vec{a}$，又垂直于$\vec{b}$的方向。对于两个二维向量的外积，其向量积的大小可看做是平面上$\vec{a}$、$\vec{b}$为边的平行四边形面积。

以坐标形式表示，设有$\vec{a}=a_x\vec{i}+a_y\vec{j}$，$\vec{b}=b_x\vec{i}+b_y\vec{j}$两个二维向量（$\vec{i}$、$\vec{j}$表示 x、y 轴方向的单位向量，大家应该比较熟悉了）。向量积的计算符合分配率法则，所以可以按照展开式将其展开：

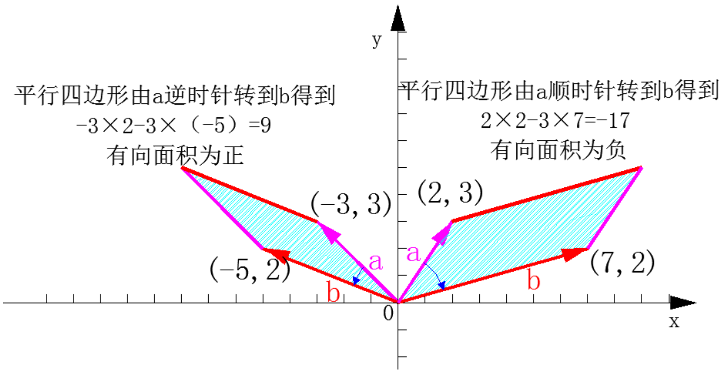
$$\vec{c}=\vec{a}\times\vec{b}=(a_x\vec{i}+a_y\vec{j})\times(b_x\vec{i}+b_y\vec{j})$$

$$=a_xb_x(\vec{i}\times\vec{i})+a_xb_y(\vec{i}\times\vec{j})+a_yb_x(\vec{j}\times\vec{i})+a_yb_y(\vec{j}\times\vec{j})$$

由向量积的定义可得，$\vec{i}\times\vec{i}=0$，$\vec{j}\times\vec{j}=0$。设$\vec{i}\times\vec{j}=\vec{k}$（$\vec{k}$显然是一个垂直于$\vec{i}$、$\vec{j}$的单位向量），则有$\vec{j}\times\vec{i}=-\vec{k}$（与$\vec{k}$方向相反的单位向量）。于是：

$$\vec{c}=\vec{a}\times\vec{b}=a_xb_y(\vec{i}\times\vec{j})+a_yb_x(\vec{j}\times\vec{i})=(a_xb_y-a_yb_x)\vec{k}$$

$\vec{c}$的模等于（$a_xb_y-a_yb_x$）的绝对值，从平面几何理解是$\vec{a}$、$\vec{b}$为边的平行四边形面积。也可将去掉绝对值符号后的表达式（$a_xb_y-a_yb_x$）看做$\vec{a}$、$\vec{b}$向量为边的平行四边形的“有向面积”，其值的正负取决于$\vec{k}$的正负（方向）。按照一般约定的右手法则，可理解为若这个平行四边形是由向量$\vec{a}$沿逆时针



风萧萧兮乱翻书

方向转到$\vec{b}$而得到的，有向面积为正值；若这个平行四边形是由向量$\vec{a}$沿顺时针方向转到$\vec{b}$而得到的，有向面积为负值。

如果仔细推敲一下两个二维向量的向量积，感觉是比较奇怪的。用$\begin{bmatrix} a_x \\ a_y \end{bmatrix}$、$\begin{bmatrix} b_x \\ b_y \end{bmatrix}$分别表示$\vec{a}$、$\vec{b}$，那么$\vec{c}$如何表示呢？我想应该是$\begin{bmatrix} 0 \\ 0 \\ a_x b_y - a_y b_x \end{bmatrix}$。也就是说，两个二维向量的向量积是一个三维向量！好吧，还是别管它了。我们现在更关心的是得出来的那个“数”。

推广一下，我们还是来看看三个三维向量的叉乘展开式。

设有$\vec{a}=a_x\vec{i}+a_y\vec{j}+a_z\vec{k}$，$\vec{b}=b_x\vec{i}+b_y\vec{j}+b_z\vec{k}$，$\vec{c}=c_x\vec{i}+c_y\vec{j}+c_z\vec{k}$三个三维向量。（$\vec{i}$、$\vec{j}$、$\vec{k}$表示 x、y、z 轴方向的单位向量）

$$
\begin{aligned}
\vec{a}\times\vec{b}\times\vec{c} &= (a_x\vec{i}+a_y\vec{j}+a_z\vec{k})\times(b_x\vec{i}+b_y\vec{j}+b_z\vec{k})\times(c_x\vec{i}+c_y\vec{j}+c_z\vec{k}) \\
&= a_xb_xc_x(\vec{i}\times\vec{i}\times\vec{i})+a_xb_xc_y(\vec{i}\times\vec{i}\times\vec{j})+a_xb_xc_z(\vec{i}\times\vec{i}\times\vec{k})+ \\
&\quad a_xb_yc_x(\vec{i}\times\vec{j}\times\vec{i})+a_xb_yc_y(\vec{i}\times\vec{j}\times\vec{j})+a_xb_yc_z(\vec{i}\times\vec{j}\times\vec{k})+ \\
&\quad a_xb_zc_x(\vec{i}\times\vec{k}\times\vec{i})+a_xb_zc_y(\vec{i}\times\vec{k}\times\vec{j})+a_xb_zc_z(\vec{i}\times\vec{k}\times\vec{k})+ \\
&\quad a_yb_xc_x(\vec{j}\times\vec{i}\times\vec{i})+a_yb_xc_y(\vec{j}\times\vec{i}\times\vec{j})+a_yb_xc_z(\vec{j}\times\vec{i}\times\vec{k})+ \\
&\quad a_yb_yc_x(\vec{j}\times\vec{j}\times\vec{i})+a_yb_yc_y(\vec{j}\times\vec{j}\times\vec{j})+a_yb_yc_z(\vec{j}\times\vec{j}\times\vec{k})+ \\
&\quad a_yb_zc_x(\vec{j}\times\vec{k}\times\vec{i})+a_yb_zc_y(\vec{j}\times\vec{k}\times\vec{j})+a_yb_zc_z(\vec{j}\times\vec{k}\times\vec{k})+ \\
&\quad a_zb_xc_x(\vec{k}\times\vec{i}\times\vec{i})+a_zb_xc_y(\vec{k}\times\vec{i}\times\vec{j})+a_zb_xc_z(\vec{k}\times\vec{i}\times\vec{k})+ \\
&\quad a_zb_yc_x(\vec{k}\times\vec{j}\times\vec{i})+a_zb_yc_y(\vec{k}\times\vec{j}\times\vec{j})+a_zb_yc_z(\vec{k}\times\vec{j}\times\vec{k})+ \\
&\quad a_zb_zc_x(\vec{k}\times\vec{k}\times\vec{i})+a_zb_zc_y(\vec{k}\times\vec{k}\times\vec{j})+a_zb_zc_z(\vec{k}\times\vec{k}\times\vec{k})
\end{aligned}
$$

分析一下上面的式子，三个含三分项的式子相乘，容易知道，可以展开为3^3=27 个分项。由于$\vec{i}\times\vec{i}=\vec{j}\times\vec{j}=\vec{k}\times\vec{k}=0$，因此上式中含有同样两个单位向量的分项均为 0，只剩下含有三个均为不同单位向量相乘的分项。由简单地排列组合

（相当于$\vec{i}$、$\vec{j}$、$\vec{k}$的排列组合）可以知道，其中有 3！=6 个不为零的分项。将等于 0 的分项去除，

上式= $a_xb_yc_z(\vec{i}\times\vec{j}\times\vec{k})+a_xb_zc_y(\vec{i}\times\vec{k}\times\vec{j})+a_yb_xc_z(\vec{j}\times\vec{i}\times\vec{k})+$

$a_yb_zc_x(\vec{j}\times\vec{k}\times\vec{i})+a_zb_xc_y(\vec{k}\times\vec{i}\times\vec{j})+a_zb_yc_x(\vec{k}\times\vec{j}\times\vec{i})$

需要注意的是，不能认为$\vec{i}\times\vec{j}=\vec{k}$，从而得出$\vec{i}\times\vec{j}\times\vec{k}=\vec{k}\times\vec{k}=0$ 的结论。类比两个二维向量的向量积的结果，应该把（$\vec{i}\times\vec{j}\times\vec{k}$）视为四维空间中与$\vec{i}$、$\vec{j}$、$\vec{k}$均垂直的一个新的单位向量$\vec{l}$。

设$\vec{l}=\vec{i}\times\vec{j}\times\vec{k}$，则有

$\vec{j}\times\vec{k}\times\vec{i}=\vec{k}\times\vec{i}\times\vec{j}=\vec{l}$，$\vec{i}\times\vec{k}\times\vec{j}=\vec{k}\times\vec{j}\times\vec{i}=\vec{j}\times\vec{i}\times\vec{k}=-\vec{l}$

由此得出：

$\vec{a}\times\vec{b}\times\vec{c}=$ （$a_xb_yc_z+a_yb_zc_x+a_zb_xc_y-a_xb_zc_y-a_yb_xc_z-a_zb_yc_x$）$\vec{l}$

有意思的是，得出的新向量的模正好是三维几何空间中以$\vec{a}$、$\vec{b}$、$\vec{c}$为棱的平行六面体的体积（如果你有兴趣可以算算看，三维空间算这个也不是太难，我就先闪了…）。参照前面的思路，可以认为括弧中的值

（$a_xb_yc_z+a_yb_zc_x+a_zb_xc_y-a_xb_zc_y-a_yb_xc_z-a_zb_yc_x$）

是这个平行六面体的有向体积。

在前面展开式的基础上，不难推出 n 个 n 维向量的向量积的表达式（说说而已，我可不想真来一遍）。由上述计算过程可以归纳出，n 个 n 维向量的向量积的展开式应含有 n^n 个向量，但其中不为零的向量应有 n!个。n 个 n 维向量的向量积当然还是一个向量，它的方向是 n+1 维空间中与这个 n 维空间垂直的方向（按推理也应该有两个，应设其中一个为正，另一个即为负）。更有意思的是，有人说得出的表达式（去掉那个单位向量表达式剩余部分）可看做是以这 n 个 n 维向量为棱的超多面体的有向体积。（这个？…好吧，详细证明略！你记住这个结论就行了(^_^)!）

这里还需要提一下表达式内的符号问题。也就是，在求两个向量积的过程中有：设$\vec{i}\times\vec{j}=\vec{k}$，则有$\vec{j}\times\vec{i}=-\vec{k}$；在求三个三维向量的向量积计算过程中，有

风萧萧兮乱翻书

设$\vec{l}=\vec{i}\times\vec{j}\times\vec{k}$，则有$\vec{j}\times\vec{k}\times\vec{i}=\vec{k}\times\vec{i}\times\vec{j}=\vec{l}$，$\vec{i}\times\vec{k}\times\vec{j}=\vec{k}\times\vec{j}\times\vec{i}=\vec{j}\times\vec{i}\times\vec{k}=-\vec{l}$。在求 n 个向量积的过程中怎么扩展？

这里要引入排列的逆序和逆序数概念。对于 n 个不同的元素，先规定各元素之间有一个标准次序，如上例中规定$\vec{i}$、$\vec{j}$、$\vec{k}$为标准排序，于是在这 n 个元素的任一排列中，当某两个元素的先后次序与标准次序不同，就说有一个逆序，如对调一下$\vec{i}$、$\vec{j}$，排成了$\vec{j}$、$\vec{i}$、$\vec{k}$，这就是一个逆序。一个排列中所有逆序的总数叫做这个排列的“逆序数”。逆序数为奇数的排列叫做“奇排列”；逆序数为偶数的排列叫做“偶排列”。

对于 n 个相互正交的单位向量$\vec{l}_1$，$\vec{l}_2$，…，$\vec{l}_n$，设$\vec{l}_{n+1}=\vec{l}_1\times\vec{l}_2\times\ldots\times\vec{l}_n$，则可以看做：当乘积中的排列为偶排列时，结果等于$\vec{l}_{n+1}$；当乘积中的排列为奇排列时，结果等于$-\vec{l}_{n+1}$。正如上例中，设$\vec{l}=\vec{i}\times\vec{j}\times\vec{k}$，则有$\vec{j}\times\vec{k}\times\vec{i}=\vec{k}\times\vec{i}\times\vec{j}=\vec{l}$，$\vec{i}\times\vec{k}\times\vec{j}=\vec{k}\times\vec{j}\times\vec{i}=\vec{j}\times\vec{i}\times\vec{k}=-\vec{l}$。

风萧萧兮乱翻书

§2 行列式

前面推导的向量积的表达式（去掉那个单位向量表达式剩余部分）有一个大名鼎鼎的术语——行列式！

行列式最初当然不是像我们上一节讨论的那样推出来。恕我孤陋寡闻，我确实还没看到有人这样推导行列式。也许这种“向量积的展开式”在数学上根本就是不严谨的吧。不严谨又咋样，管它呢，我是草根我怕谁，又不是写教科书，便于理解就是王道！不管怎么说，我们是第一个！所以我决定命名这种推导方法为“乱翻书法”！

行列式的概念最初是伴随着方程组的求解而发展起来的，行列式的提出可以追溯到十七世纪，最初的雏形由日本数学家关孝和与德国数学家戈特弗里德 莱布尼茨各自独立得出，时间大致相同。（不说这个了，再说成科学历史故事书了。）

从解方程组用到的消元法（我们现在知道，其实也就是对增广矩阵进行初等行变换）也能推导出行列式的表达式。以三元线性方程组

$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$为例（为简单起见，仅对系数矩阵进行变换）：

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\sim\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{pmatrix}\sim\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22}-a_{12}a_{21} & a_{11}a_{23}-a_{13}a_{21} \\ 0 & a_{11}a_{32}-a_{12}a_{31} & a_{11}a_{33}-a_{13}a_{31} \end{pmatrix}\sim$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22}-a_{12}a_{21} & a_{11}a_{23}-a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{pmatrix}$$

其中$\Delta=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}-a_{11}a_{23}a_{32}$，也就是三阶行列式表达式。你也可以看出第二行第二列位置上的值就是二阶行列式的表达式。不过这种推导好像不大直观，对四阶以上更难推广。

来看看教科书上行列式的定义吧：

定义：设有 $n\times n$ 个数，排成 n 行 n 列的数表

$$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix},$$

风萧萧兮乱翻书

作出表中位于不同行不同列的 n 个数的乘积，并冠以符号（-1）^{t}，得到形如

$$(-1)^t a_{1p_1} a_{2p_2} \cdots a_{np_n}$$

的项，其中 $p_1p_2\cdots p_n$ 为自然数 1,2，…，n 的一个排列，t 为这个排列的逆序数。由于这样的排列共有 n!个，因而形如上式的项共有 n!项。所有这 n!项的代数和称为 n 阶行列式。记作

$$D=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix},$$

简记作 $\det(a_{ij})$.

有意思的是，不同的教科书对行列式有不完全相同的定义。百度百科和维基百科则描述了一个更加“数学”的定义：$\det(A)=\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\prod_{i=1}^{n}A_{i,\sigma(i)}$。或许这才是行列式的原始定义，又或许是这个原始定义实在太“数学”了，不同的教科书试图从不同的角度对它进行了不是那么“数学”的描述。反正我看到这个定义已基本崩溃了。

关于行列式，我们首先要理清一个概念：行列式并不是一个矩阵，而是针对一个方阵中的元素进行若干运算的结果，行列式是一个数！

以行列式的形式再把二阶、三阶行列式写一下：

对于方阵 A=$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$，其对应的二阶行列式$|A|=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}$.

对于方阵 A=$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$，其对应的三阶行列式

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

上面列的二、三阶行列式计算公式在手算中经常可能用到，四阶以上的一般就不会这样算了。在计算机这么发达的今天，手算这个实在是没事找事。对于二、三阶行列式计算公式，教科书上一般都会介绍诸如“对角线法”等直观、简单的记忆法，在此就不讨论了。（一方面咱不鼓励应试教育，另一方面加斜线那种排版我实在排不出来。）

§3 行列式的性质

下面讨论行列式的一些性质。对于这些性质，我们在此一般不讨论详细的证明（如果想学习证明过程，教科书上都有），主要讨论从不同的角度来理解。

前面我们提到行列式的几何意义可以看做是行列式中列向量（或行向量）为棱的超平行多面体的有向体积，这点对于理解行列式的有关性质有十分重要的感性意义。还有另一种解释是方阵 A 的行列式就是线性变换 A 下的体积（面积）的伸缩因子。这两个解释一个是静态的体积概念，一个是动态的变换比例概念，但具有相同的几何本质。因为 A 可以看做 AE，也就是行列式 det(A)可以看做是 A 对单位正多面体（有向体积当然是 1）体积进行伸缩的比例。对于仅涉及一个方阵的行列式，我们可以主要利用静态体积的概念来讨论、理解行列式的有关性质，但对于涉及两个方阵的行列式的性质还需要从动态变换的角度来理解。

性质一：行列式和它的转置行列式相等

我们知道，行列式是针对方阵的一种运算。将这个方阵转置后的行列式称为原行列式的转置行列式。即$|A^T|$为$|A|$的转置行列式。

这个性质告诉我们，对一个方阵而言，用“乱翻书法”（不好意思，我再自吹一下）推导行列式，取行向量和列向量是一样的。其实仔细观察推导的过程和结果，也能发现这个结论。

从有向体积的角度，这个性质告诉我们方阵行向量为棱的超多面体和列向量为棱的超多面体的有向体积（当然也包括二维的面积和三维体积）相等。

从行列式的角度，有关行列式的性质中对列成立的，对行自然成立。反之亦然。

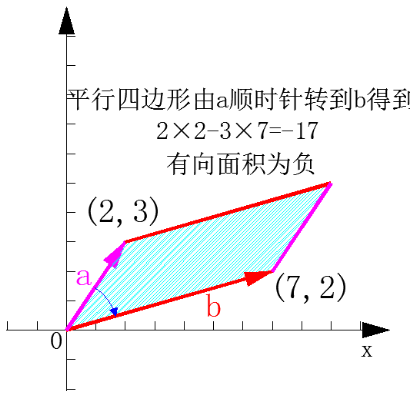
性质二：互换行列式的两列，行列式变号

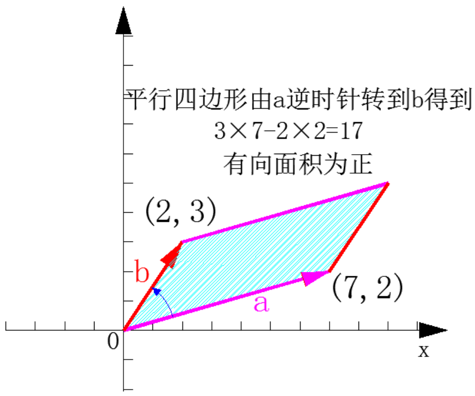
从前面的推导过程容易看出，互换行列式的两列，相当于排列的新的标准次序是原来标准次序的一个逆序。这使得各分项的符号与原标准次序中的符号正好相反，也就是整个值变号。

从向量积的角度，更换向量积中两个向量的顺序，得到的向量大小不变，方向反向。

从有向体积的角度，有向体积变号。

风萧萧兮乱翻书





性质三：行列式的两列相同，其值为零

由前一个行列式性质，交换这相同的两列后，行列式要变号。可由于这两列相同，交换后的行列式其实还是原行列式。所以这个行列式只能等于零。

从向量积的角度，这应该很清楚了，两个方向相同的其向量积等于零，这是向量积的基本性质。既然这两个向量的向量积为零，当然再乘其他向量的向量积结果还是零。

从有向体积的角度，如果有两列相同，即 n 维超多面体的两个棱重合，这个 n 维超多面体将褪变为 n-1 维超多面体，其有向体积当然为零。比如，三维的平行六面体褪变为二维平面，其体积当然为零；二维平行四边形褪变为一条直线，其面积也当然为零。

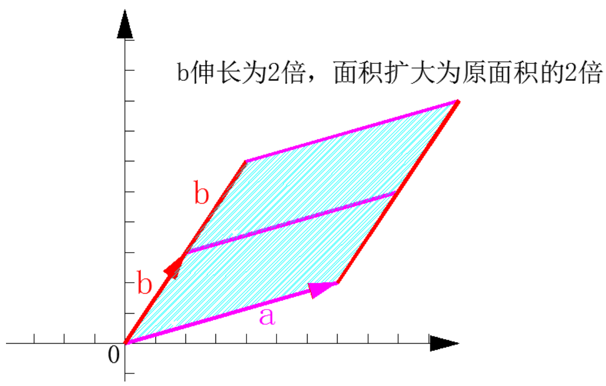
性质四：行列式中某两列元素成比例，其值为零

与前一性质类似，如有两列元素成比例，显然这两列所表示的向量在一条直线上，其有向体积为零。

性质五：行列式的某列中所有元素都乘以同一个数λ，等于用λ乘此行列式

由行列式的推导公式容易证明这个性质。

从有向体积的角度，n 维超多面体的其中一条棱伸长（或缩短）为原来的λ倍，有向体积也放大（或缩小）为原来的λ倍。以二维为例：如右图所示，当 b 伸长为原来的λ倍后，以其为边的平行四边形面积当然也就扩大为原面积的λ倍。

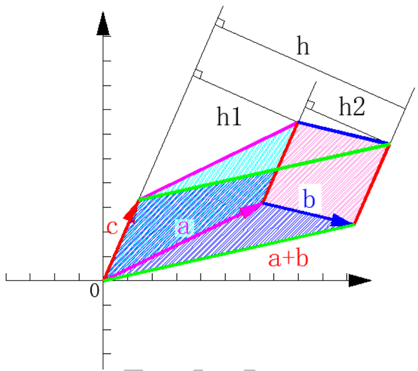


风萧萧兮乱翻书

性质六：若行列式的某列向量为两个列向量之和：$A_s=A_{s1}+A_{s2}$，则行列式等于两个行列式的和：

$$\left| A_1,\ \ldots A_{s1}+A_{s2},\ \ldots A_n \right| = \left| A_1,\ \ldots A_{s1},\ \ldots A_n \right| + \left| A_1,\ \ldots A_{s2},\ \ldots A_n \right|$$

从有向体积的角度，n 维超多面体的一条棱（向量）是另两个向量的向量和，这个 n 维超多面体的有向体积是将这条棱换位另两个向量为棱后的两个 n 维超多面体的有向体积之和。



以二维面积为例，如右图所示。以$\vec{c}$、$\vec{a}$为边的平行四边形面积为$\vec{c}$的长度×h_1，以$\vec{c}$、$\vec{b}$为边的平行四边形面积为$\vec{c}$的长度×h_2，以$\vec{c}$、$\vec{a}+\vec{b}$为边的平行四边形面积为$\vec{c}$的长度×h。显然，$h=h_1+h_2$。所以以$\vec{c}$、$\vec{a}+\vec{b}$为边的平行四边形面积是以$\vec{c}$、$\vec{a}$为边的平行四边形面积和以$\vec{c}$、$\vec{b}$为边的平行四边形面积之和。

性质七 把行列式的某一列的λ倍加到另一列上去，行列式值不变

从向量积的角度，以二维为例：

$$(\vec{a}+\lambda\vec{b})\times\vec{b}=[(a_x+\lambda b_x)\ \vec{i}+(a_y+\lambda b_y)\vec{j}]\times(b_x\vec{i}+b_y\vec{j})$$

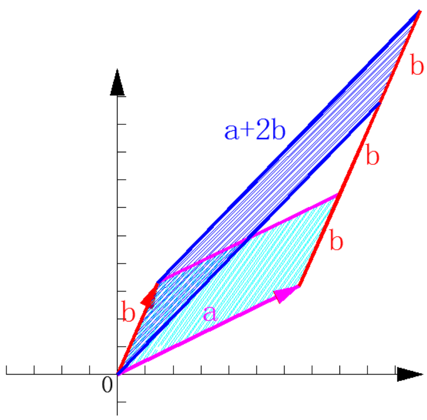
$$= (a_x+\lambda b_x)b_x\,(\vec{i}\times\vec{i})+(a_x+\lambda b_x)b_y\,(\vec{i}\times\vec{j})+(a_y+\lambda b_y)b_x\,(\vec{j}\times\vec{i})+(a_y+\lambda b_y)b_y\,(\vec{j}\times\vec{j})$$

$$=(a_x+\lambda b_x)b_y\ \vec{k}-(a_y+\lambda b_y)b_x\ \vec{k}=[(a_xb_y+\lambda b_xb_y)-(a_y b_x+\lambda b_yb_x)]\vec{k}$$

$$=[(a_xb_y-a_yb_x+\lambda b_xb_y-\lambda b_yb_x)]\vec{k}\ =(a_xb_y-a_yb_x)\vec{k}=\ \vec{a}\times\vec{b}$$

从上面的演算过程可以发现，+λ$\vec{b}$带入的新分项将相互抵消为零。

从有向体积的角度，n 维超多面体的一条棱（向量）换成这个向量和另一个棱向量的λ倍之和作为棱，有向体积不变。



以二维面积为例，如右图所示。以$\vec{a}$、$\vec{b}$为边的平行四边形面积为青色阴影区

域。将边$\vec{a}$换为$\vec{a}+\lambda\vec{b}$为边，由图可以明显看出$\vec{a}+\lambda\vec{b}$仍是在$\vec{a}+\vec{b}$所在的平行于$\vec{b}$的直线上，也就是相当于将这个平行四边形在底和高都不变的情况下拉伸为蓝色区域的平行四边形。显然，两者的面积仍然相等。

性质八 设 A、B 为两个 n 阶方阵，则有 $|AB|=|A||B|$

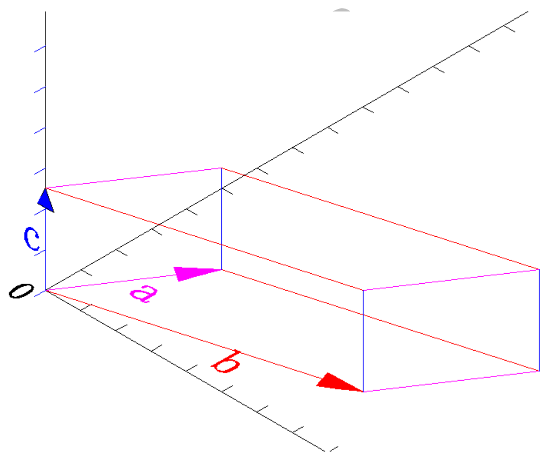
这个性质涉及到两个方阵，用静态的体积概念就不好解释了。用动态变换因子的概念可以理解为 A 对 B 进行变换后的体积等于 A 对 E 的体积变换因子和 B 对 E 的体积变换因子的积。

风萧萧兮乱翻书

§3 行列式按行（列）展开

前面我们说过 n 阶方阵对应的行列式可以看做对应 n 个 n 维向量为棱的 n 维超多面体的有向体积。这个“体积”怎么算呢？在三维空间中说到算体积，当然第一反应就是底面积乘高嘛。四维以上的 n 维空间呢，应该可以看做是其中 n-1 维的“体积”再乘剩下一维上的“高”吧（当然，有向体积需注意有个方向的问题）。

来个最简单的：设在二维平面上有两个二维向量$\vec{a}$、$\vec{b}$，分别表示为$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$、$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$，我们已经知道，以其为边的平行四边形的有向面积是$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$。现在在三维上垂直于平面正方向加入个长度为λ向量$\vec{c}$，表示为$\begin{bmatrix} \lambda \\ 0 \\ 0 \end{bmatrix}$，当然



在三维空间中，$\vec{a}$、$\vec{b}$也相应变成了$\begin{bmatrix} 0 \\ a_1 \\ a_2 \end{bmatrix}$、$\begin{bmatrix} 0 \\ b_1 \\ b_2 \end{bmatrix}$，以三个向量为棱的平行六面体的有向体积表示为行列式为$\begin{vmatrix} \lambda & 0 & 0 \\ 0 & a_1 & b_1 \\ 0 & a_2 & b_2 \end{vmatrix}$。而这个平行六面体（这时显然是个正六面体）的有向体积当然很好算，λ乘底面积嘛，即$\lambda\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$。于是我们得到了$\begin{vmatrix} \lambda & 0 & 0 \\ 0 & a_1 & b_1 \\ 0 & a_2 & b_2 \end{vmatrix}=\lambda\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$。我们还发现，只要保持$\vec{a}$、$\vec{b}$和高λ不变，拉动这个平行六面体，体积当然也不会发生变化（也可以从行列式性质七的角度相互印证和理

风萧萧

解)。也就是说，$\vec{c}$ 不一定取 $\begin{bmatrix}\lambda\\0\\0\end{bmatrix}$，取 $\begin{bmatrix}\lambda\\x\\y\end{bmatrix}$ 都可以，从行列式的角度都有 $\begin{vmatrix}\lambda & 0 & 0\\x & a_1 & b_1\\y & a_2 & b_2\end{vmatrix}$ $=\lambda\begin{vmatrix}a_1 & b_1\\a_2 & b_2\end{vmatrix}$。这个推导过程并不奇特，但这个结果倒挺有点意思的。推广一下，我们可以得到：$\begin{vmatrix}a_{11} & 0 & \cdots & 0\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & \ddots & \vdots\\a_{n1} & a_{n2} & \dots & a_{nn}\end{vmatrix}=a_{11}\begin{vmatrix}a_{22} & \cdots & a_{2n}\\\vdots & \ddots & \vdots\\a_{n2} & \cdots & a_{nn}\end{vmatrix}$，其中 $\begin{vmatrix}a_{22} & \cdots & a_{2n}\\\vdots & \ddots & \vdots\\a_{n2} & \cdots & a_{nn}\end{vmatrix}$ 可看做是 n-1 维的“有向体积”，a_{11} 就是剩下一维上的高。当然，从行列式定义的计算过程也容易验证这个等式。

得到上面这个等式，现在咱不管什么体积了（怎么有点卸磨杀驴的感觉？）。来看一个三阶行列式 $\begin{vmatrix}a_{11} & a_{12} & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}$，利用行列式的性质六，我们可以把它写成 $\begin{vmatrix}a_{11} & a_{12} & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}=\begin{vmatrix}a_{11}+0 & 0+a_{12} & 0+a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}=\begin{vmatrix}a_{11} & 0 & 0\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}+\begin{vmatrix}0 & a_{12} & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}$。对第二个行列式继续使用同样的方法继续拆分一下，得到：

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

。我们已经知道，$\begin{vmatrix}a_{11} & 0 & 0\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}=a_{11}\begin{vmatrix}a_{22} & a_{23}\\a_{32} & a_{33}\end{vmatrix}$，对于第二项 $\begin{vmatrix}0 & a_{12} & 0\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}$，利用行列式性质二，将第二列与第一列对换，可得 $\begin{vmatrix}0 & a_{12} & 0\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}=-\begin{vmatrix}a_{12} & 0 & 0\\a_{22} & a_{21} & a_{23}\\a_{32} & a_{31} & a_{33}\end{vmatrix}=-a_{12}\begin{vmatrix}a_{21} & a_{23}\\a_{31} & a_{33}\end{vmatrix}$（别忘了对换两行变号的哦）。对于第三项 $\begin{vmatrix}0 & 0 & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{vmatrix}$，当然可以类似于第二项将第三列与第一列对换，但得到的结果规律性就不够好了（因为结果中出现的

风萧萧兮乱翻书

$\begin{vmatrix} a_{22} & a_{21} \\ a_{32} & a_{31} \end{vmatrix}$打乱了元素在原行列式中的排列顺序。我们可以把第三列跟第二列对换，再把新的第二列（原第三列）跟第一列对换（你没有掐着手指头还在对换中吧？其实就是把第三列放最前面去啦！）当然，这时其实是完成了两次对换，所以

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

。最终，我们得到：

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

。当然，不难看出其中的

$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$是三阶行列式中将$a_{11}$所在的行、列去除后剩余的二阶行列式，它称作$a_{11}$的余子式。加上对换次数（回想一下逆排序、奇排序、偶排序吧）带来的符号变化（这个符号变化其实有很简单地规律），称为代数余子式。

来看一下教科书的定义：

定义：在 n 阶行列式中，把（i, j）元 a_{ij} 所在的第 i 行和第 j 列划去后，留下来的 n-1 阶行列式称为（i, j）元 a_{ij} 所的余子式，记作 M_{ij}；

记 $A_{ij}=(-1)^{i+j}$ ，A_{ij} 叫做（i, j）元 a_{ij} 的代数余子式。

将上面的推导扩展，就可以得到 n 阶行列式的按行（列）展开的定理：

定理：行列式等于它的任一行（列）的各元素与其对应的代数余子式乘积之和，即：

$$D=a_{i1}A_{i1}+a_{i2}A_{i2}+\cdots+a_{in}A_{in}\ (i=1, 2, \cdots, n);$$

或 $D=a_{1j}A_{1j}+a_{2j}A_{2j}+\cdots+a_{nj}A_{nj}\ (j=1, 2, \cdots, n)$。

这个展开式称“拉普拉斯展开式”（怪不得拉的很开）。当然，拉普拉斯说还可以按照几行（列）一起拉，我们就不管它了。

风萧萧兮乱翻书

§5 行列式的计算

给定一个 n 阶行列式，理论上当然可以按照行列式定义的方法去算出来行列式的值。但我们知道，这个计算式是 n!个乘项的和。也许有人会说，我们现在有拉普拉斯展开式啊。注意看这个“拉开式”，它把 n 阶行列式拉开为 n 个 n-1 阶行列式，还得把 n-1 阶行列式再拉开。一直拉下去，其实还是 n!个乘项之和（看来这个拉开式也没什么用处）。

对于一个 25×25 方阵而言（这在现代科技中的应用来说不算大的），就有 25！个（大约等于 1.55×10^{25}，或许应该可以读做“百万亿亿”吧，不知道我读对没有）乘法运算。如果一个超级计算机每秒钟可完成 1 万亿次的乘法运算，它算完这些乘法运算需要约一百五十五亿秒，大概约二百多年（算错了别喷我，我说过我数学不好的。反正很多年，反正我这一辈子不够算出这个行列式的）！用这种方法算大型行列式，估计电脑也会发疯的。

好在我们还有那么多行列式的性质。再回到行列式的性质中来吧。

首先看看行列式什么时候会等于零？或者应该说什么行列式才不等于零？根据行列式的性质三、性质四，结合有向体积的理解，你想到了什么？如果一个二阶方阵的两个列向量（或行向量，下同）共线，或者贴切的用术语说它们线性相关，那么它们为棱的“平行四边形”就褪变为一条直线，其有向面积（行列式）当然也就等于零。同样，如果一个三阶方阵的三个列向量线性相关，从几何角度来说就是共面，显然它们就无法构建一个平行六面体，其有向体积（行列式）也等于零。推广而言，对于一个 n 阶方阵，设它的秩为 r，那么这个方阵的列向量（或行向量）当然都在 r 维空间中，如果 r＜n 的话，以其为棱就不可能构建 n 维超多面体，也就是说这个 n 维超多面体的体积肯定为零。所以，只有 r=n，也就是方阵满秩的时候，其对应的行列式才不为零。反过来说，如果一个 n 阶方阵对应的行列式不为零，说明以这个方阵的 n 个列向量（或行向量）为棱可以“撑出”一个 n 维超多面体，或者说 n 个列向量不在同一个 n-1 维空间中，或者说 n 个列向量线性无关，或者说这可以以这 n 个列向量为基生成 n 维向量空间，或者说这个方阵是非奇异矩阵，或者说…你应该知道，这么多或者说都是等价的，还有最后的结论也是等价的——这个方阵满秩。由此，我们可以得出一个结论，n 阶方阵满秩是其对应的行列式不为零的充分必要条件。线性代数中，对于一个方

阵 A，如果 det(A)=0，则称 A 是奇异矩阵；如果 det(A)≠0，则称 A 是非奇异矩阵。

总结前面的一些概念，我们可以知道，对于 n 阶方阵 A 而言，以下说法等价：（1）A 的列向量组（行向量组）线性无关；（2）A 满秩；（3）A 可逆；（4）A 是非奇异矩阵；（5）A 经初等行（列）变换可变为单位阵 E；（6）A 与单位阵 E 等价；（7）A 的列（行）构成 $\mathbb{R}^n$ 的一组基；（8）方程 Ax=b 对 $\mathbb{R}^n$ 中的任意 b 都有唯一解；（9）线性齐次方程组 Ax=0 只有零解。而且，上述性质对于 A^T 同样成立。

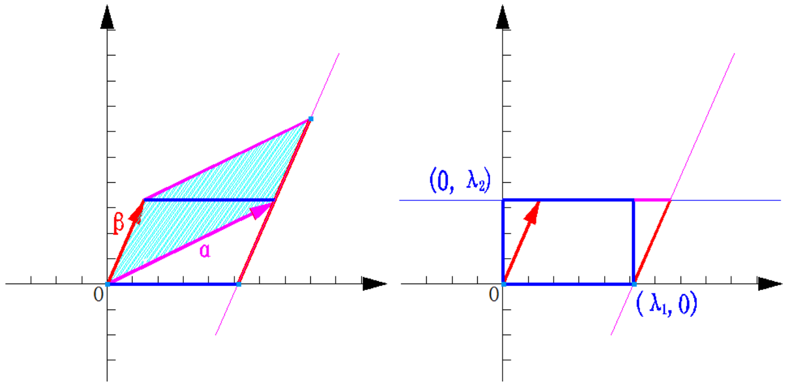
既然只有满秩的方阵其对应的行列式不为零，这好像轻松了不少。前面已经讨论过，可逆矩阵通过初等变换都可变换为单位阵 E。那么对一个方阵的进行初等变换的过程对其对应的行列式有什么影响呢？先看一下矩阵初等变换（以初等行变换为例）包含的三种变换模式：①换位变换（更换某两行的位置），②倍法变换（某行乘以一个不为 0 的数λ），③倍加变换（某行加上另一行的λ倍）。对照行列式的性质可知，换位变换将使行列式变号；倍法变换将同比例改变行列式；倍加变换对行列式无影响。

仔细分析一下可逆矩阵经初等行变换变成单位阵的过程中三种变换所起的作用，不难发现倍加变换的作用是把非对角线上的非零元素变换为零；换位变换的作用是变换行的位置，将出现在对角线上的零元素换成其他行的非零元素，将诸如 $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_3 \\ 0 & \lambda_2 & 0 \end{pmatrix}$ 的矩阵变成 $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$（这句话表达的有点别扭，但我实在不知道在书面上能用简单地几句话说明白。但只要做几次变换的例题，相信你能理解）；倍法变换的作用则是使对角线元素变为 1。由于倍加变换对行列式没有影响，换位变换改变行列式符号，所以，只需记录换位变换的次数，或者在进行一次换位变换时就将换位的某一行变次号，即能满足行列式不变的需求。这样，只通过倍加变换和换位变换，我们就可以将满秩矩阵变换为一个对角线阵。而对于对角线阵而言，其行列式当然就是对角线元素的积。

我们以二阶方阵为例，从几何的角度再来理解下。根据行列式“把行列式的某一列的 k 倍加到另一列上去，行列式值不变”的性质及其图形理解，我们前面已经知道，如图中所示，以α、β从向量的角度即是通过倍加变换使得α成为(λ_1, 0）两个二维向量为边的平行四边形的一边沿红线平移，平行四边形有向面积不变。

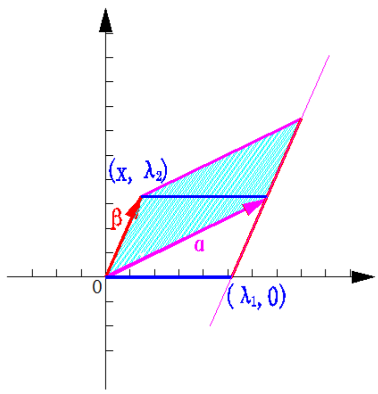
风萧萧兮乱翻书

既然这样，当然不妨将α平移到 x 轴上，从向量的角度即是通过倍加变换使得α成为（$\lambda_1, 0$）；然后，同样可以再沿上边平移至 y 轴，从向量的角度即是通过倍加变换使得β成为（$0, \lambda_2$）。显然，平行四边形的有向面积就是$\lambda_1 \times \lambda_2$。



同样的理解，对于 n 阶方阵，可以将对应 n 维超平行多面体的棱（列向量或行向量）经相应平移（倍加变换）至对应的ε_1，ε_2，…，ε_n方向，即在保持体积不变的情况下，将其平移为一个边分别为$\lambda_1, \lambda_2, \cdots, \lambda_n$的 n 维正超多面体。从向量的角度，也就是将对应列（行）向量变换为（$\lambda_1, 0, \cdots, 0$），（$0, \lambda_2, \cdots, 0$），…，（$0, \cdots, 0, \lambda_n$），方阵经变换成为对角阵。方阵行列式（有向体积）当然也就是对角线元素的积。

从前面的平行四边形的平移过程我们不难发现，其实将α平移为（$\lambda_1, 0$）后，即方阵经倍加变换为$\begin{pmatrix} \lambda_1 & x \\ 0 & \lambda_2 \end{pmatrix}$后，不需将其再平移为正四方形，已经可以计算出其面积了，因为β的 x 坐标值不会对面积产生任何影响。对应到方阵的变换而言，不需要将其变换为对角阵，仅需将其变换为形如$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & a_{nn} \end{pmatrix}$的三角阵，其对角线元素的积就是方阵的行列式的值。当然，从行列式展开式的角度，也容易得出三角阵对应的行列式为对角线元素之积的结论。而且，对于任何一个方阵而言，经初等行变换都可变换为行阶梯阵，而满秩方阵的行阶梯阵就是三角阵，而降秩方阵的行阶梯阵对角线元素将含有 0。因此，通过倍加变换和换位变换将方阵变换为行阶梯阵，从而以其对角线元素的积求出对应行列式更有普遍意义。当然，容易理解在进行行列式计算时也并不是不能使用倍法变换，有



风萧萧兮乱翻书

时提取某行的公因子能降低行列式的计算难度。

事实上，大多数计算机程序在进行行列式的计算时，通常也是采用通过初等行变换将方阵变换为行阶梯阵这种方式计算的。有人曾证明，一个 n×n 阶行列式利用行变换展开大约需要 $2n^3/3$ 次运算，对于一个 25×25 行列式而言，大约需要 1 万次运算，微型计算机也可在一秒内计算出来。（当然，对于纯手算而言，还是够呛。虽然手算有一些可以简化计算的技巧，但在计算机如此发达的今天，手算大型行列式几乎没有任何实际意义。）

以一个手算例子来作为行列式计算讨论的结束吧。求解$\begin{vmatrix} 1 & -4 & 3 & 4 \\ 2 & -8 & 0 & 12 \\ -3 & 0 & 1 & -2 \\ 3 & -9 & 5 & 10 \end{vmatrix}$：

首先通过倍加变换将第一列除第一个元素外的其他元素变换为零，可得到

$\begin{vmatrix} 1 & -4 & 3 & 4 \\ 2 & -8 & 0 & 12 \\ -3 & 0 & 1 & -2 \\ 3 & -9 & 5 & 10 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 0 & -6 & 4 \\ 0 & -12 & 10 & 10 \\ 0 & 3 & -4 & -2 \end{vmatrix}$，由于第二行第二个元素（对角线上的元素）为零，需进行一次行换位变换，行列式也因此需变号。这时，可将负号提出，也可将变换的某行都变次号。我们采用后者，将第二行与第四行交换并将原第四行变号，然后继续倍加变换消去下三角中的非零元素，得:

$\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & -3 & 4 & 2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -6 & 4 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & -3 & 4 & 2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -6 & 4 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & -3 & 4 & 2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 2 \end{vmatrix}$=1×（−3）×（−6）×2=36.

风萧萧兮乱翻书

§6 克莱姆法则

前面提到行列式的最初是从解线性方程组引出的，那当然还得看看它们之间到底有什么关系了。

行列式是仅对方阵的，所以我们来看 n 个 n 元线性方程组成的方程组 $Ax=\vec{b}$。我们前面已经知道，A 满秩时，对于任何$\vec{b}$，方程有唯一解；否则，方程仅在$\vec{b}$满足相关条件时（前面已经讨论的很清楚了，这里就不详细描述了）有无穷解，不满足相关条件时无解。但我们前面一直没有给出唯一解的公式。

设α_1，α_2，…，α_n是 A 的列向量，用ε_1，ε_2，…，ε_n表示单位阵 E 的列向量，将 x 替换ε_1，得到一个新矩阵 $E_{(x-1)}$，有 $A\,E_{(x-1)}=(b，\alpha_2，\cdots，\alpha_n)$。由行列式性质八，有 $\det(A)\det(E_{(x-1)})=\det(b，\alpha_2，\cdots，\alpha_n)$，现在我们把这个 $E_{(x-1)}$写全，来看看 $\det(E_{(x-1)})$是什么：$E_{(x-1)}=\begin{vmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & \ldots & 1 \end{vmatrix}$，很明显，$\det(E_{(x-1)})=x_1$。代入前式有 $\det(A)x_1=\det(b，\alpha_2，\cdots，\alpha_n)$，当 $\det(A)\neq 0$，于是有 $x_1= \det(b，\alpha_2，\cdots，\alpha_n)/\det(A)$。

为更清楚些，再求个 x_2。用将 x 替换ε_2，得到一个新矩阵 $E_{(x-2)}$，有 $A\,E_{(x-2)}=(\alpha_1，b，\alpha_3，\cdots，\alpha_n)$。$\det(A)\det(E_{(x-2)})=\det(\alpha_1，b，\alpha_3，\cdots，\alpha_n)$。其中 $E_{(x-2)}=\begin{vmatrix} 1 & x_1 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_n & \ldots & 1 \end{vmatrix}$，按第二行展开，可得 $\det(E_{(x-1)})=x_2$，（看来拉普拉斯还是有用的），当 $\det(A)\neq 0$，于是有 $x_2= \det(\alpha_1，b，\alpha_3，\cdots，\alpha_n)/\det(A)$。

这下应该比较清晰了，可以上定理了：

定理：如果 n 个 n 元线性方程组成的方程组的系数行列式 D 不等于零，那么方程组有唯一解：$x_j=\dfrac{D_j}{D}$ ($j=1, 2, \cdots, n$)；其中，D_j(是把系数行列式 D 中的第 j 列的元素用方程组右端的常数项代替后所得到的 n 阶行列式。

这个定理又称克莱姆法则，也有的教科书写克拉默法则，还有的译作克雷默。老外名字就是麻烦，一个 Cramer，翻译出这么多名字来。

对于一个由 n 个 n 元线性方程组成的方程组而言，使用克莱姆法则求这 n

风萧萧兮乱翻书

个未知数，需要计算 n+1 个 n 阶行列式，一般情况下是极为低效的，所以真正利用它来求解方程的并不多。但它指出了方程组的解与行列式的关系，构建两者之间的桥梁。

还记得讨论逆矩阵的时候，我们提到过逆矩阵可以通过解 n 个 $Ax=\varepsilon_n$ 方程来求得吗？如果一个方阵 A 可逆，令 $Ax=\varepsilon_1$，求出的 x 即为 A^{-1} 的第一列，依次令 $Ax=\varepsilon_2, \cdots, \varepsilon_n$，可以求出 A^{-1} 的全部 n 列。现在我们用克莱姆法则和行列式的有关概念再来讨论一下。

先来看 $Ax=\varepsilon_1$。为清楚起见，把它写全一点：$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$，

根据克莱姆法则，$x_1=\frac{|A_1|}{|A|}$，其中 $|A_1|=\begin{vmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix}$，这个行列式不陌生了，

它就等于 a_{11} 的代数余子式 A_{11}；$x_2=\frac{|A_2|}{|A|}$，其中 $|A_2|=\begin{vmatrix} a_{11} & 1 & \cdots & a_{1n} \\ a_{21} & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & 0 & \cdots & a_{nn} \end{vmatrix}$，按第二列

展开，容易看出是 a_{12} 的代数余子式 A_{12}。以此类推，可得 $x=\begin{pmatrix} \frac{A_{11}}{|A|} & \frac{A_{12}}{|A|} & \cdots & \frac{A_{1n}}{|A|} \end{pmatrix}^T$

同样，分别令 $Ax=\varepsilon_2, \cdots, \varepsilon_n$，可解出 A^{-1} 的共 n 个列向量。从而得出

$A^{-1}=\frac{1}{|A|}\begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$（注意第 i 行第 j 列是 A_{ji}），其中由代数余子式构

成的矩阵 $A^*=\begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$ 称为矩阵 A 的伴随矩阵，简称伴随阵。

于是，我们得到了又一个逆矩阵的计算公式 $A^{-1}=\frac{1}{|A|}A^*$。

风萧萧兮乱翻书

特殊地，对于二阶方阵 A=$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$而言，有 $A_{11}=a_{22}$，$A_{12}=-a_{21}$，$A_{21}=-a_{12}$ $A_{22}=a_{11}$，所以 $A^{-1}=\frac{1}{|A|}A^{*}=\frac{1}{|A|}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$。二阶矩阵的伴随矩阵有一个形象的记忆公式：主（对角线）换位，副（对角线）换号。

风萧萧兮乱翻书

第四章　线性变换

§1 线性空间和基变换

线性空间可以看做是向量空间的概念中对向量概念的进一步抽象和扩展。有的教科书上甚至没有区分这两个概念，而直接将两者视为同一个对象。先来看一下教科书上线性空间的定义。

定义：设 V 是一个非空集合，R 为实数域。如果对于任意两个元素 α，$\beta\in V$，总有唯一的元素 $\gamma\in V$ 与之对应，称为 α 与 β 的和，记作 $\alpha+\beta=\gamma$；又对于任一数 $\lambda\in R$ 与任一元素 $\alpha\in V$，总有唯一的一个元素 $\delta\in V$ 与之对应，称为 λ 与 α 的积，记作 $\delta=\lambda\alpha$；并且这两种运算满足以下八条运算规律（设 α，β，$\gamma\in V$；λ，$\mu\in R$）：

（1）$\alpha+\beta=\beta+\alpha$；（2）$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$；（3）$V$ 中存在零元素 0，对任何 $\alpha\in V$，都有 $\alpha+0=\alpha$；（4）对任何 $\alpha\in V$，都有 α 的负元素 $\beta\in V$，使 $\alpha+\beta=0$；（5）$1\alpha=\alpha$；（6）$\lambda(\mu\alpha)=(\lambda\mu)\alpha$；（7）$(\lambda+\mu)\alpha=\lambda\alpha+\mu\alpha$；（8）$\lambda(\alpha+\beta)=\lambda\alpha+\lambda\beta$。

那么，V 就称为（实数域 R 上的）向量空间（或线性空间），V 中元素不论其本来的性质如何，统称为（实）向量。

简言之，凡满足上述八条规律的加法及乘数运算，就称为线性运算；凡定义了线性运算的集合，就称向量空间。

由这个定义可以看出，线性空间的定义把向量的概念进行了扩展和延伸，向量不一定只是有序数组。这实在太抽象了点吧，看不懂就不管它了。

我们还是回到我们熟悉的向量空间中吧。

前面讨论过向量空间的维数、基、坐标等概念。我们知道，同一个向量空间可以取多组不同的基，而坐标是与基相伴而生的，在确定了某一组基的基础上才能讨论某个向量在这组基下的坐标，脱离了基，坐标也就没有了意义。

再回到我们最开始讨论的方程$\begin{bmatrix}2\\3\end{bmatrix}x+\begin{bmatrix}3\\-4\end{bmatrix}y=\begin{bmatrix}7\\2\end{bmatrix}$，我们曾说所求的（x，y），即（2，1）就是以$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$为基的坐标系中向量$\begin{bmatrix}7\\2\end{bmatrix}$的坐标。可我们说坐标是与基相伴而生的，$\begin{bmatrix}7\\2\end{bmatrix}$又是哪组基下的坐标呢？其实$\begin{bmatrix}7\\2\end{bmatrix}$是$\begin{bmatrix}1\\0\end{bmatrix}$、$\begin{bmatrix}0\\1\end{bmatrix}$这组自然基

风萧萧兮乱翻书

下的坐标，只是我们一般没有把它写出来而已。所以，上面的方程写成完整矩阵形式应该是$\begin{pmatrix}2 & 3\\3 & -4\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}7\\2\end{pmatrix}$。这个方程的意义从坐标角度的解读为：以$\begin{bmatrix}2\\3\end{bmatrix}$和$\begin{bmatrix}3\\-4\end{bmatrix}$为基的坐标系中向量$\begin{bmatrix}2\\1\end{bmatrix}$就是以$\begin{bmatrix}1\\0\end{bmatrix}$、$\begin{bmatrix}0\\1\end{bmatrix}$为基的坐标系中的向量$\begin{bmatrix}7\\2\end{bmatrix}$。

不知道你理解了没有？不妨再举个简单的例子：A、B 两个物体赛跑， A 的速度为 72，B 的速度为 20，哪个更快呢？也许现在小学的某些测试就会考这样的题。如果你没有头晕，当然会知道在没有约定速度“度量单位”的情况下，比较 72 和 20 这两个数字的大小并没有意义，甚至这两个数字也没有实际意义。而如果声明 A 的速度单位是“公里/小时”，B 的速度单位是“米/秒”，通过换算你会发现原来它们的速度是相等的。对于向量而言也是类似，单独讨论向量的坐标就像没有约定速度单位而讨论速度为 72 还是 20 一样没有意义。从这个意义上来说，基就是向量 “度量单位”的声明。有时为方便起见，我们可以约定国际单位“米/秒”为速度的“缺省单位”，这时认为凡是没有另行指定单位的都默认以“米/秒”为速度的单位，这样 20×2=40 才有实际意义，同时也会有 72（公里/小时）=20 的等式出现，其中 20 并不是没有单位，而是默认了。对于向量而言，E 就是这个默认的“缺省度量单位”。从这个意义上来说，$\begin{pmatrix}2 & 3\\3 & -4\end{pmatrix}\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}7\\2\end{pmatrix}$与“72（公里/小时）=20”有着类似的意义。

虽然“米/秒”是国际单位，但在实际生活中各行各业、各种情景中还是会结合具体情况选择常用的速度单位，就像汽车上常用“公里/小时”或“英里/小时”作为速度单位，舰船习惯用“节”(“海里/小时”)，航空研究惯用“马赫”，而天文研究却更常使用光速。同样，虽然 E_n是向量空间$\mathbb{R}^n$的自然基，但有的情况下我们仍需要使用其他不同的基。

我们知道，同一个速度取不同的单位时有不同的数值，同一个向量在不同的基下有不同的坐标。但只要都是速度单位，它们之间总是可以相互换算的。同样，同一个向量空间不同的基当然也应该可以相互换算。

设α是向量空间 V 的一个向量，其在基α_1，α_2，…，α_n下的坐标为x，在基β_1，β_2，…，β_n下的坐标为y，则有：

$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)x=(\beta_1,\beta_2,\cdots,\beta_n)y$，两边左乘$(\beta_1,\beta_2,\cdots,\beta_n)^{-1}$

即得：$y=(\beta_1,\beta_2,\cdots,\beta_n)^{-1}(\alpha_1,\alpha_2,\cdots,\alpha_n)x$。

令 $P=(\beta_1,\beta_2,\cdots,\beta_n)^{-1}(\alpha_1,\alpha_2,\cdots,\alpha_n)$，则有：$y=Px$；　$x=P^{-1}y$ 这就是两个基下的坐标转换公式。

同时由 $P=(\beta_1,\beta_2,\cdots,\beta_n)^{-1}(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 可得出

$(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_n)P$

$(\beta_1,\beta_2,\cdots,\beta_n)=(\alpha_1,\alpha_2,\cdots,\alpha_n)P^{-1}$

上式中的矩阵 P 称为从基$\alpha_1,\alpha_2,\cdots,\alpha_n$到$\beta_1,\beta_2,\cdots,\beta_n$的过渡矩阵。

定理：设 V_n 中的元素α，在基$\alpha_1,\alpha_2,\cdots,\alpha_n$下的坐标为$(x_1,x_2,\cdots,x_n)^T$，在基$\beta_1,\beta_2,\cdots,\beta_n$下的坐标为$(y_1,y_2,\cdots,y_n)^T$. 若两个基满足

$$(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_n)P$$

则有坐标变换公式

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}=P\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}，或\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}=P^{-1}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

为了更直观的类比理解，把上面的说法再以速度换算作一次类比：

有一个速度，在速度单位为“公里/小时”时数值为 72，在速度单位为“米/秒”时数值为 20，则有：

72（公里/小时）=20（米/秒），两边乘以$\frac{1}{20}$，得

（米/秒）=$\frac{1}{20}$×72（公里/小时），令 P=$\frac{1}{20}$×72，即有

（米/秒）=P（公里/小时）；（公里/小时）=P^{-1}（米/秒）

P 是“米/秒”到“公里/小时”的换算比例值，P^{-1}(这里表示的是 P 的倒数哦)是 “公里/小时”到“米/秒”的换算比例值。若一个速度在“米/秒”单位下的数值为 x，则其在“公里/小时”单位下的数值 y=Px；若一个速度在“米/秒”单位下的数值为 y，则其在“公里/小时”单位下的数值 x= P^{-1}y.

通过以上的类比，我们不难发现单位换算比例值 P 就是 1“米/秒”在“公里

风萧萧兮乱翻书

/小时”下的数值，同样的，基（α_1，α_2，…，α_n）到基（β_1，β_2，…，β_n）的过渡矩阵 P 也就是基（α_1，α_2，…，α_n）在基（β_1，β_2，…，β_n）下的坐标。

以上是一个简单的类比。当然，如果你是个吹毛求疵的人（也许该用“精益求精”更好一些），真的严格一一对应两个类比的话，你可能会发现有些步骤并不会完全相同，但这应该不会影响我们类比理解过渡矩阵、坐标换算等概念。或许很多时候过分的严谨正是我们理解数学的最大拦路虎，想想那些严谨却又枯燥、晦涩的数学定义，就能彻底打击像我这样的人的数学学习热情。来看一个定义：一般来说，在一个集合 F 上定义一个二元关系“+”，满足：(1) 交换律：对任意的 a，b ∈ F，a + b = b + a ∈ F；(2) 结合律：对任意的，a + (b + c) = (a + b) + c；(3) 单位元：存在一个元素 0 ∈ F，满足对任意的 a ∈ F，a + 0 = 0 + a = a；(4) 逆元：对任意的 a ∈ F，存在一个元素 (-a) ∈ F，满足 a + (-a) = 0。“+”称作定义在集合 F 上的加法。这是数学上“加法”的定义，的确够严谨、够深刻、够本质，但如果小学生学习加法前先学习这个定义，估计 99%的学生会对数学从此失去信心。

不知不觉扯远了，回来继续吧。

风萧萧兮乱翻书

§2 规范正交基

既然同一个向量空间 $\mathbb{R}^n$ 可以取不同的基，某些情况下，我们当然希望通过取得某些特殊的基来简化相关的运算。

自然基 E 当然是其中最普遍、最简便也最特殊一组基。前面我们提到过常用的直角坐标系有两个显著的特点——坐标轴两两垂直且单位长度为 1。对于作为基的向量组而言，我们要引入向量的长度和夹角的概念。

高等数学里对向量的数量积（内积、点乘）和向量积（外积、叉乘）的概念都有阐述，但一般只讨论到三维向量的情况。前面我们讨论行列式时曾扩展使用了 n 维向量积的概念（虽然我已经强调了那或许不够严谨），在讨论矩阵乘法中的列乘行中也曾提到数量积的概念。这些概念扩展到 n 维并不复杂，下面来看看相关一系列定义：

定义：设有 n 维向量

$$x=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix},\quad y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$$

令 $\langle x,y\rangle = x_1y_1+x_2y_2+...+x_ny_n$*，称* $\langle x,y\rangle$ *为向量 x 与 y 的内积。*

定义：令 $\|x\|=\sqrt{\langle x,x\rangle}=\sqrt{x_1^2+x_2^2+...+x_n^2}$*，* $\|x\|$ *称为 n 维向量 x 的长度（或范数）。当* $\|x\|=1$ *时，称 x 为单位向量。*

有些教科书（例如同济版）将内积表示为 $[x,y]$，我们这排版容易跟向量使用的中括号搞混，就采用 $\langle x,y\rangle$ 了。

定义：设有 n 维向量 x≠0，y≠0 时，

$$\theta=\arccos\frac{\langle x,y\rangle}{\|x\|\ \|y\|}$$

称为向量 x，y 的夹角。当 $\langle x,y\rangle=0$ *时，称向量 x 与 y 正交。若 x=0，则其与任何向量都正交。*

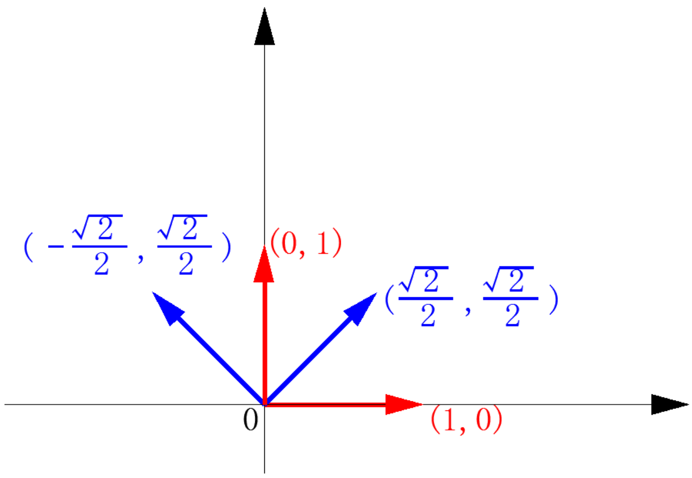
根据上面的定义，我们可以用数学语言描述一下自然基的两个重要特点：组成自然基的向量组两两正交且均为单位向量。

风萧萧兮乱翻书

如果不含零向量的向量组 α_1，α_2，…，$\alpha_r \in \mathbb{R}^n$，且两两正交，则称之为正交向量组，如果其中各向量均为单位向量，则称其为规范正交向量组（有的教科书称之为“标准正交向量组”）。容易证明，两两正交的向量组线性无关，因此，两两正交的不含零向量的向量组 α_1，α_2，…，$\alpha_r \in \mathbb{R}^n$ 是 r 维向量空间（$\mathbb{R}^n$ 的 r 维子空间）的一组基。

规范正交向量组组成的矩阵称为正交矩阵，不难证明正交矩阵 A 满足 $A^TA=AA^T=E$，即 $A^{-1}=A^T$。当然 $A^{-1}=A^T$ 也是个正交矩阵。

当向量空间的一组基为正交向量组时，称这组基为正交基；如果这组基为标准正交向量组，称这组基为规范正交基（有的教科书称为“标准正交基”）。显然，自然基 E 就是一组规范正交基。当然，E 并不是唯一的规范正交基。可以想象将 E 的向量组整体旋转任意角度得到的新的向量组仍是一个规范正交基。以二维平面为例：从坐标轴的角度理解，可将 E 视作直角坐标系，如图，将 E 的列向量组（坐标轴）同时旋转一个角度，得到得显然还是一组规范正交基。如 $\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$ 的列向量组就是 $\mathbb{R}^2$ 的一组规范正交基，它可以看做是将 E 的列向量组同时旋转 45° 得到的。$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 的列向量组也是一组规范正交基，它可以看做是将 E 的列向量组旋转θ角得到的。此外，将 E 经行（列）换位变换后得到类似 $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 的列向量组也是一组规范正交基，注意它和 E 并不等同



（我们前面曾强调过构成基的向量组是有序的）。当然，将它整体旋转得到的新的向量组也仍是规范正交基。反过来说，任何正交规范基也都可看做是 E 或者 E 经行（列）换位变换后得到正交规范基（诸如 $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$、$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$、$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ 等）

风萧萧兮乱翻书

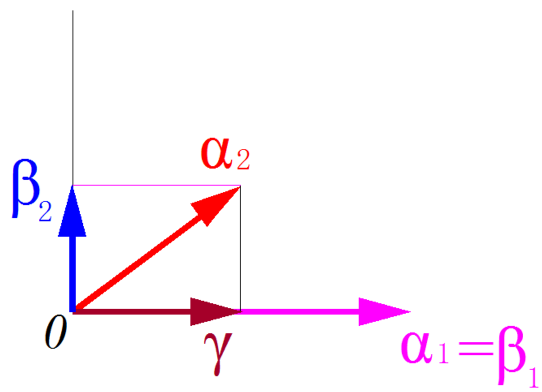
经某种旋转得到的结果。

使用规范正交基会给向量的很多计算带来便利，比如求一个向量在规范正交基下的坐标就比较方便。

有时候，我们会遇到作为一组基的向量组 α_1，α_2，…，α_n 中部分向量正交，而另一部分不正交的情况，这时我们就希望找到一些正交向量替代其中不正交的向量，从而组成一组正交基。当然，极端的情况就是向量组中的向量都不正交，我们希望保留其中的一个向量，而将其他向量都“正交化”。

先来看两个向量的情况。如图所示，有 α_1，α_2 两个向量，现要保持 α_1 不变，将 α_2 正交化，其实也就是想求得与 α_1 垂直的一个新向量。最直观的想法有两个，一个是将 α_2 旋转至垂直方向，另一个就是求 α_2 在垂直方向的投影向量。旋转的方法最直观，但要想求出这个新向量显然要用到大量的三角函数之类的东西（如果你有兴趣可以试试看，说不定能推导出一个新的正交化的公式）。我们下面来讨论求投影向量的方法。从图中可以直观的看出，α_2 与它在 α_1 上的投影向量 γ 及它在 α_1 垂直方向的投影向量 β_2 之间存在关系：

$\alpha_2=\beta_2+\gamma$。如果我们能求得 γ，自然就能求出 β_2。如何求出 γ 呢？显然，它的方向与 α_1 相同，关键是它的长度（模）是什么？希望你还记得高等数学中向量的数量积（内积）还有一种表达式：$\langle\alpha_1,\alpha_2\rangle=|\alpha_1||\alpha_2|\cos\theta$，其中 $|\alpha_1|$、$|\alpha_2|$ 是向量的长度（模），θ 是两个向量的夹角。$|\alpha_2|\cos\theta$ 不就是 α_2 在 α_1 上的投影长度！于是我们求得 γ 的长度 $|\gamma|=\dfrac{\langle\alpha_1,\alpha_2\rangle}{|\alpha_1|}$。下面再将 α_1 单位化，也就是除以自身的长度，就得到 α_1 方向的单位向量 $\dfrac{\alpha_1}{|\alpha_1|}$。求得



$\gamma=\dfrac{\langle\alpha_1,\alpha_2\rangle}{|\alpha_1|}\dfrac{\alpha_1}{|\alpha_1|}=\dfrac{\langle\alpha_1,\alpha_2\rangle}{|\alpha_1|\ |\alpha_1|}\alpha_1$。$|\alpha_1||\alpha_1|$ 也可以写成内积形式 $\langle\alpha_1,\alpha_1\rangle$，于是可写成：

$\gamma=\dfrac{\langle\alpha_1,\alpha_2\rangle}{\langle\alpha_1,\alpha_1\rangle}\alpha_1$，最终得出：

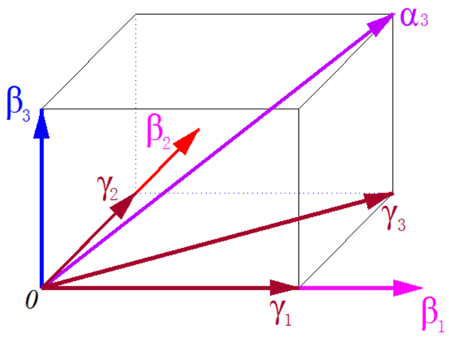
风萧萧兮乱翻书

$$\beta_2=\alpha_2-\gamma=\alpha_2-\frac{\langle\alpha_1,\alpha_2\rangle}{\langle\alpha_1,\alpha_1\rangle}\alpha_1$$

有了上面的讨论基础，下面再来看看三个向量的情形。如图所示，设β_1、β_2是我们已经得到的两个正交向量，现在需要将第三个向量α_3正交化。与前面的讨论类似，我们需要找到α_3在与β_1、β_2都正交方向，也就是与β_1、β_2所在平面正交方向的投影向量β_3。图中，γ_3为α_3在β_1、β_2所在平面的投影向量，γ_1为α_3在β_1上的投影向量，γ_2为α_3在β_2上的投影向量。由图不难理解几个向量之间的关系：

$\alpha_3=\beta_3+\gamma_3$；$\gamma_3=\gamma_1+\gamma_2$

即得到：$\beta_3=\alpha_3-\gamma_3=\alpha_3-\gamma_1-\gamma_2$



而一个向量在另一个向量上的投影向量我们已经知道求解方式了。

$$\gamma_1=\frac{\langle\beta_1,\alpha_3\rangle}{\langle\beta_1,\beta_1\rangle}\beta_1$$

$$\gamma_2=\frac{\langle\beta_2,\alpha_3\rangle}{\langle\beta_2,\beta_2\rangle}\beta_2$$

于是最终得到$\beta_3=\alpha_3-\frac{\langle\beta_1,\alpha_3\rangle}{\langle\beta_1,\beta_1\rangle}\beta_1-\frac{\langle\beta_2,\alpha_3\rangle}{\langle\beta_2,\beta_2\rangle}\beta_2$

按照这个方法，我们可以得到把α_1，α_2，…，α_r规范正交化的一个公式：

$$\beta_1=\alpha_1$$

$$\beta_2=\alpha_2-\frac{\langle\alpha_1,\alpha_2\rangle}{\langle\alpha_1,\alpha_1\rangle}\alpha_1$$

$$\beta_3=\alpha_3-\frac{\langle\beta_1,\alpha_3\rangle}{\langle\beta_1,\beta_1\rangle}\beta_1-\frac{\langle\beta_2,\alpha_3\rangle}{\langle\beta_2,\beta_2\rangle}\beta_2$$

…………

$$\beta_r=\alpha_r-\frac{\langle\beta_1,\alpha_r\rangle}{\langle\beta_1,\beta_1\rangle}\beta_1-\frac{\langle\beta_2,\alpha_r\rangle}{\langle\beta_2,\beta_2\rangle}\beta_2-\ldots-\frac{\langle\beta_{r-1},\alpha_r\rangle}{\langle\beta_{r-1},\beta_{r-1}\rangle}\beta_{r-1}$$

将上述β_1，β_2，…，β_r分别单位化，也就是除以自身的长度$\|\beta_i\|$，就得到了一组规范正交基。

风萧萧兮

上述这个过程称为施密特（Schimidt）正交化过程。这推导过程看起来没什么高深的地方吧，看来发现个定理、公式什么的并获得命名也不是什么高不可攀的事情。

风萧萧兮乱翻书

§3 线性变换

再来看最开始讨论的方程$\begin{pmatrix}2 & 3\\ 3 & -4\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}7\\ 2\end{pmatrix}$（How old are you？翻译过来就是“怎么老是你？”）这次我们看方程的等式左边$\begin{pmatrix}2 & 3\\ 3 & -4\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}$，把$\begin{pmatrix}x\\ y\end{pmatrix}$当做输入的话，输入一个二维向量，就能得到一个唯一确定的输出二维向量。你想到了什么？函数 y=f(x)！在定义域中任取一个自变量 x，按照对应法则 f，总有唯一确定的值 y 与之对应。当然，我们这个式子中输入的是向量，输出的也是向量，所以从严格的函数定义而言，它不算是一个函数，但显然是个映射。（回去翻翻映射的定义）。而且，从矩阵乘法的本质理解，它是对向量进行某种线性运算的结果，因此是一种线性映射。下面先看看线性映射的定义：

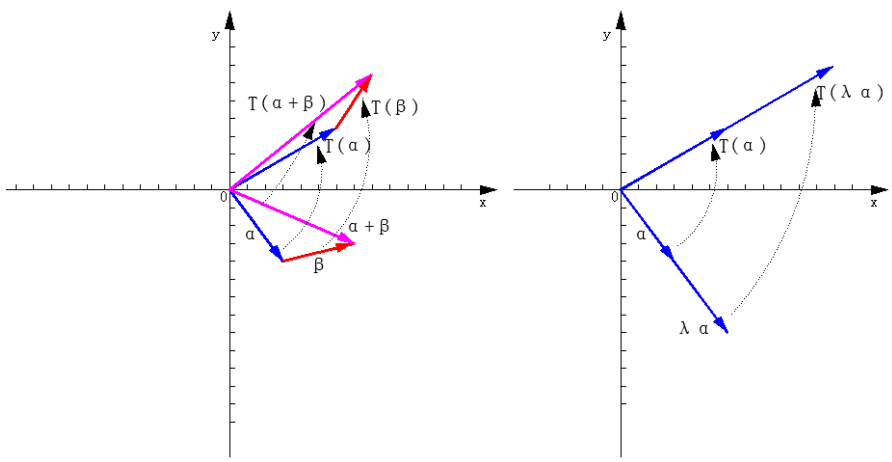
定义：设 V_n，U_m 分别是 n 维和 m 维线性空间，T 是一个从 V_n 到 U_m 的映射，如果映射 T 满足：

（1） 任给α_1，$\alpha_2\in V_n$（从而$\alpha_1+\alpha_2\in V_n$），有 $T(\alpha_1+\alpha_2)=T(\alpha_1)+T(\alpha_2)$；

（2） 任给$\alpha\in V_n$，$\lambda\in R$（从而$\lambda\alpha\in V_n$），有 $T(\lambda\alpha)=\lambda T(\alpha)$.

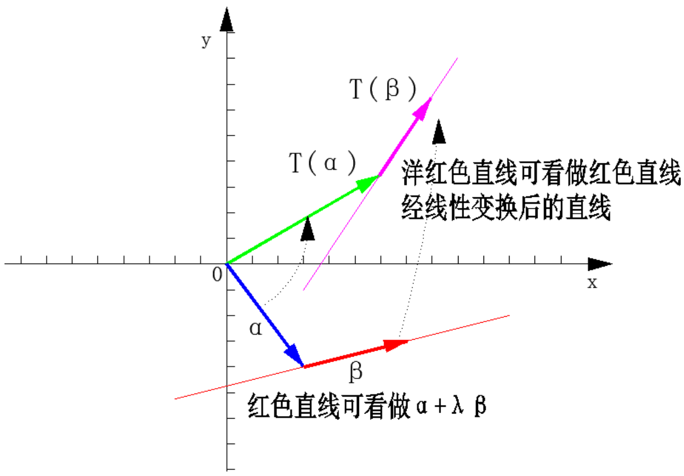
那么，T 就称为从 V_n 到 U_m 的线性映射，或称为线性变换。

线性映射的两个条件字面解释就是两个向量的和向量经线性变换的结果与两个向量分别变换后在相加的结果相等；向量数乘后经线性变换的结果与变换后再数乘的结果相等。如图所示，向量α、β经线性映射 T 后分别映射为向量 T(α)、T(β)，那么同一映射对向量α+β的映射结果 T(α+β)等于 T(α)+T(β)。同样，对λα的映射结果 T(λα)等于λT(α)。



风萧萧兮乱翻书

从几何的角度可以将线性映射直观的理解为空间的一条直线经某种线性映射后的结果仍是一条直线。如果这条直线过原点，当然可以将这条直线上的点看做某个向量α及λα（$\lambda \in R$）的顶点的集合，显然，经过某种线性映射 T，它将映射为 T(α)及λT(α)的顶点的集合，也就是另一条过原点的某条直线。对于不过原点的直线，总可以看做某两个向量α+λβ（$\lambda \in R$）的顶点的集合，如图所示，λ取某一个实数，就可以表示α+λβ这条直线上的某个点。对于任一个线性变换 T 都有 T（α+λβ）= T(α)+λT(β)，这当然还是一条直线。所以，可以直观的理解，一条直线经任一线性变换 T，仍为一条直线。当然这条直线的角度、方向会随着 T 的不同而不同。



前面已经说了，由矩阵与一个列向量的乘法 Ax 可以看做是一种线性变换，反过来，向量空间中的任一线性变换也可以写成矩阵的形式。（这里暂不做具体证明阐述，结合后面的讨论应该能明白。同济版教科书上也有具体说明）。也就是说，矩阵可以看做是一种线性变换。

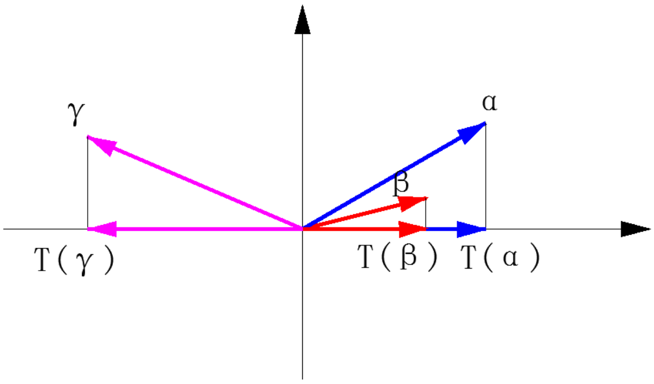
当 A 为 m×n 矩阵时，根据矩阵乘法可知对于 Ax，输入一个 n 维列向量，输出的是一个 m 维列向量，这时可看做是向量空间 V_n 到向量空间 V_m 的一个线性映射；当 A 为 n 阶矩阵时，根据矩阵乘法可知对于 Ax，输入一个 n 维列向量，输出的仍是一个 n 维列向量，这时可看做是向量空间 V_n 到其自身的一个线性映射。（有的教科书称前者为线性映射，后者为线性变换）。下面我们仅讨论后一种情况。

先来看几个特殊的线性变换矩阵：

例一：$A=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$

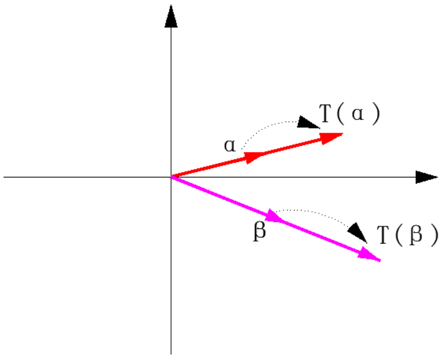
风萧萧兮乱翻书

由$\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x\\0\end{pmatrix}$，可以看出这个矩阵作用于任何二维向量$[x \quad y]^T$，都会仅保留其中的 x，而使得 y 变为零。从图形上来说，对于任一输入向量$[x \quad y]^T$，将输出其在 x 轴的投影向量$[x \quad 0]^T$。我们可以把 A 看做为一个投影变换矩阵。类似的，$\begin{pmatrix}0&0\\0&1\end{pmatrix}$是一个将输入向量投影至 y 轴的投影变换矩阵。



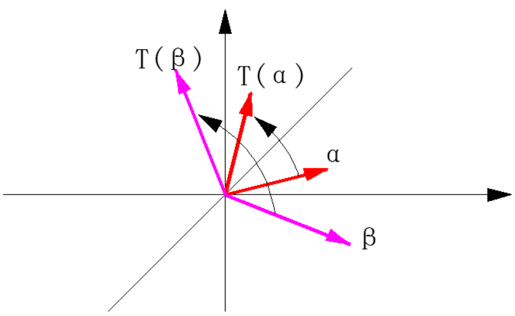
例二：A=$\begin{pmatrix}\lambda&0\\0&\lambda\end{pmatrix}$

由$\begin{pmatrix}\lambda&0\\0&\lambda\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\lambda\begin{pmatrix}x\\y\end{pmatrix}$，可以看出这个矩阵作用于任何二维向量$[x \quad y]^T$，都会使其变换为输入向量的λ倍。从图形上来说，对于任一输入向量$[x \quad y]^T$，将对其同方向拉伸λ倍。我们可以把其看做为一个拉伸变换矩阵。可以看出这个矩阵乘一列向量相当于一个常数λ乘这个列向量，这也是这种矩阵称为数量阵的原因吧。



例三：A=$\begin{pmatrix}0&1\\1&0\end{pmatrix}$

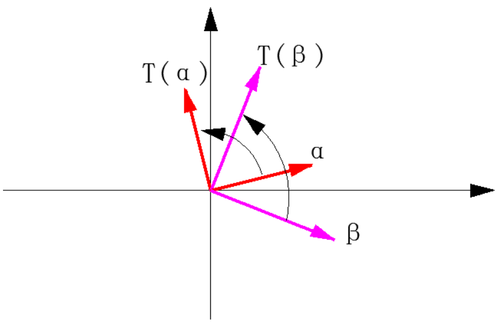
由$\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}y\\x\end{pmatrix}$，从输入向量到输出向量的变化可以看出，其实这个变换是将向量进行了行换位变换。从几何的角度，把它们画在图上可以发现，A 变换是一种对称变换，



对任一输入向量都将其沿$\begin{pmatrix}1\\0\end{pmatrix}$、$\begin{pmatrix}0\\1\end{pmatrix}$的对称轴进行对称。我们可以把其看做为一个对称换位变换矩阵。

例四：$A=\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}$

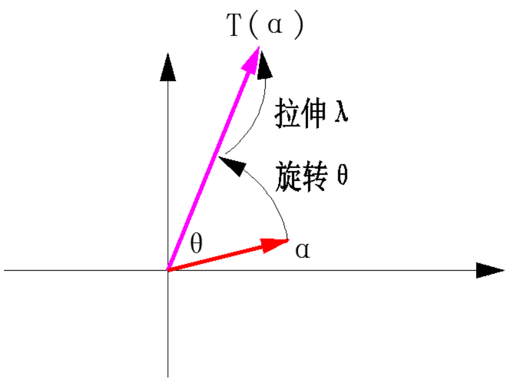
由$\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}-y\\x\end{pmatrix}$，仔细观察这个输入向量和输出向量，把它们画在图上可以发现，通过 A 的变换，其实是将输入向量逆时针旋转了 90°。更普遍一些：对于 $A=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$，其作用是将输入向量逆时针旋转 θ 角。我们可以把其看做为一个旋转变换矩阵。



看到上面两个例子，你一定会联想到前面讨论过的正交阵吧。前面我们说，正交阵（正交规范基）都可以看做是 E 或者 E 经行（列）换位变换后得到正交规范基经某种旋转得到的结果。现在结合上面的讨论，将矩阵的行换位变换看做是一种对称换位变换的话，可以说任何一个正交阵都是 E 经某种对称换位变换和某种旋转变换后得到的。同时，从线性变换的角度我们又可以说，一个正交阵是一种旋转变换和对称换位变换的结合，这种变换线性代数中又称“正交变换”。对于一个正交阵 A，$AA^T=E$ 从线性变换的角度可以理解为对 A^T 进行 A 这种正交变换的结果是 E。显然，对一个正交阵进行正交变换得到的结果仍会是一个正交阵，即：对于两个正交阵 A、B，则 AB 仍是一个正交阵。

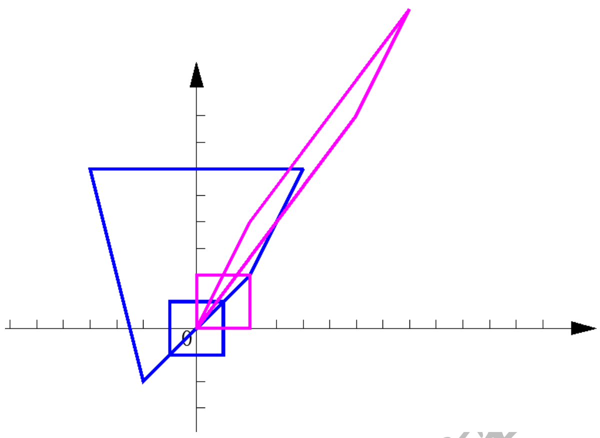
例五：$A=\begin{pmatrix}\lambda & 0\\0 & \lambda\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$

两个矩阵相乘可看做是两个线性变换的叠加。如$\begin{pmatrix}\lambda & 0\\0 & \lambda\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$可看做是对输入向量旋转一个角度后，再拉伸λ倍。



前面是一些简单而又特殊的变换示例，对于一个一般的矩阵而言，它作用于输入向量可能会对其同时进行旋转、拉伸、对称换位等变换，这使得矩阵的变换作用并不能简单地进行描述。当然，如果一个方阵 A 可逆，它的逆矩阵从线性变换的角度来说就是 A 的变换的逆变换，就是把 A 所做的旋转、拉伸、对称换位等等变换全部反过来做一遍。

来看一个普通矩阵对一个正方形的变换效果。我们已经知道，一条直线经线性变换仍为一条直线，所以，只需要看正方形的四个顶点变换的情况就可以了解正方形变换后的图形了。如图所示为 $A=\begin{pmatrix}1&3\\2&4\end{pmatrix}$ 分别对一个以 $\begin{pmatrix}1\\1\end{pmatrix}$、$\begin{pmatrix}-1\\1\end{pmatrix}$、$\begin{pmatrix}-1\\-1\end{pmatrix}$、$\begin{pmatrix}1\\-1\end{pmatrix}$ 为顶点的正方形（蓝色）和一个以 $\begin{pmatrix}2\\0\end{pmatrix}$、$\begin{pmatrix}2\\2\end{pmatrix}$、$\begin{pmatrix}0\\2\end{pmatrix}$、$\begin{pmatrix}0\\0\end{pmatrix}$ 为顶点的正方形（洋红色）变换的情况。由此可以看出，确实很难简单地描述其变换的整体情况。



两个矩阵相乘 AB 当然可以看做是对 B 的列向量分别进行 A 变换的组合，就是我们前面讨论矩阵乘法时的列线性组合法 $AB=[Ab_1, \ldots, Ab_i, \ldots Ab_n]$。也可以从行线性组合法的角度，看做对 B 进行相应的行线性变换，如

$$AB=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}\begin{pmatrix}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{pmatrix}$$

其实是将 B 的“第一行×a_{11}+第二行×a_{12}+第三行×a_{13}”的结果替换第一行，将 B 的“第一行×a_{21}+第二行×a_{22}+第三行×a_{23}”的结果替换第二行，并以此类推。当然，这种行线性变换与前面讨论过的初等行变换不能混为一谈，初等行变换可看做是行线性变换的一种特例。

风萧萧_乱翻书

§4 特征向量和特征值

通过上一节的讨论，我们知道除了对某些特殊矩阵，其对应的变换有简单的几何意义外，对于多数矩阵却不一定有对应的简单几何意义。那么要广泛地讨论某个一般矩阵所对应的线性变换对所有输入向量的变换情况是非常困难的。这种情况下，人们对一般线性变换中某些经变换后仍保持方向不变的特殊输入向量产生了兴趣。在实际生活的工程技术应用中，这种研究也有着巨大的现实意义。

对于一个矩阵 A 对应的线性变换，输入一个向量$\vec{x}$，对应会得出另一个向量$\vec{y}$=A$\vec{x}$。有没有某些特殊的输入向量$\vec{x}$，使得输出向量$\vec{y}$与输入向量$\vec{x}$方向相同呢？如果$\vec{y}$与$\vec{x}$方向相同，从向量的角度显然就是$\vec{y}=\lambda\vec{x}$（$\lambda\in R$）。也就是说，人们对是否存在某个特殊输入向量$\vec{x}$，使得 Ax=λ$\vec{x}$感兴趣。（当然，既然讨论的是输出向量$\vec{y}$等于输入向量$\vec{x}$的λ倍，当然输入向量和输出向量是同维数的，也就是 A 为方阵。）

要找到特殊输入向量$\vec{x}$，显然就是要求得 Ax=λ$\vec{x}$的解。从矩阵运算的角度，可将前式写成(A-λE) $\vec{x}$=0，其中(A-λE)显然也是个矩阵。这就是我们前面讨论过的齐次方程组。根据前面对齐次方程组的讨论情况，方程有非零解的充分必要条件是系数行列式┃A-λE┃等于零。（零向量当然是满足方程的解，但这个解的实际意义不大，所以我们要找的是非零解）。┃A-λE┃=0，即

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}=0$$

是以λ为未知数的一元 n 次方程。

在线性代数中，根据前式所求得的λ称为矩阵 A 的特征值，满足方程的特殊非零输入向量$\vec{x}$称为 A 的对应于特征值λ的特征向量，其对应的λ值称为特征值，┃A-λE┃=0 称为矩阵 A 的特征方程，将┃A-λE┃“拉开”得到的关于λ的 n 次多项式称为矩阵 A 的特征多项式。

根据多项式的性质， 这个关于λ的 n 次多项式在复数范围内总有 n 个根（m 重根算 m 个根），且有其根λ_1，λ_2，…，λ_n满足（这个多项式性质就不详细解释了，

风萧萧兮乱翻书

我也解释不大清楚）：

（1）$\lambda_1+\lambda_2+\ldots+\lambda_n=a_{11}+a_{22}+\ldots+a_{nn}$

（2）$\lambda_1\lambda_2\ldots\lambda_n=|A|$

其中 $a_{11}+a_{22}+\ldots+a_{nn}$ 也称为方阵 A 的“迹”。与行列式一样，方阵的迹也是方阵的一个数字特征。

回想一下我们曾讨论的旋转变换矩阵 $\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$，其作用是将输入向量旋转 90°输出，从这个角度理解当然不可能有输入向量和输出向量是同方向的。可根据上面的多项式性质，这个方阵对应的特征多项式有两个解，也就是仍存在对应的特征值λ。这是怎么回事呢？

我们按特征多项式解一下看看：其特征方程为 $\begin{vmatrix}0-\lambda & -1\\ 1 & 0-\lambda\end{vmatrix}=0$，展开为特征多项式为 $\lambda^2+1=0$，即 $\lambda^2=-1$. 显然在实数范围内λ无解（这与我们几何解释相通），但别忘了，在复数范围内它有两解：$\lambda_1=i$ $\lambda_2=-i$. 且它们确实满足 $\lambda_1+\lambda_2=0+0=0$，$\lambda_1\lambda_2=|A|=-1$. 由此还可解得对应 $\lambda_1=i$，$\lambda_2=-i$ 的特征向量为复向量 $\begin{bmatrix}1\\ -i\end{bmatrix}$、$\begin{bmatrix}1\\ i\end{bmatrix}$。

靠！×××（此处省略若干脏字）！线性代数里出现了虚数！这实在是晴天霹雳！在出现虚数的情况下，前面讨论的所有的几何意义都不知道怎么涵盖了！甚至连 $\begin{bmatrix}1\\ i\end{bmatrix}$ 这个复数二维向量都不知道怎么图形表示！

好在非数学专业的教科书并没有深入涉及复数矩阵的概念，我们就无视它吧！下面还是来看两个有实数根的方阵例子，并求出其特征值和特征向量。

例一：求 A=$\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}$ 的特征值和特征向量

其特征方程为 $\begin{vmatrix}3-\lambda & 1\\ 1 & 3-\lambda\end{vmatrix}=0$，特征多项式为 $(3-\lambda)^2-1=\lambda^2-6\lambda+8$，令 $\lambda^2-6\lambda+8=0$ 解得 $\lambda_1=2$，$\lambda_2=4$.

当 $\lambda_1=2$ 时，其对应特征向量满足 $\begin{pmatrix}3-2 & 1\\ 1 & 3-2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$，显然 $\begin{pmatrix}x_1\\ x_2\end{pmatrix}$ 有无穷

风萧萧兮乱翻书

解，$\begin{pmatrix}1\\-1\end{pmatrix}$是其对应于特征值$\lambda_1$=2 的一个特征向量，同方向的 k$\begin{pmatrix}1\\-1\end{pmatrix}$（k≠0）也都是其对应于特征值$\lambda_1$=2 的特征向量。

当λ_2=4 时，其对应特征向量满足$\begin{pmatrix}3-4 & 1\\ 1 & 3-4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，显然$\begin{pmatrix}x_1\\x_2\end{pmatrix}$也有无穷解，$\begin{pmatrix}1\\1\end{pmatrix}$是其对应于特征值$\lambda_2$=4 的一个特征向量，同方向的 k$\begin{pmatrix}1\\1\end{pmatrix}$（k≠0）也都是其对应于特征值$\lambda_2$=4 的特征向量。

从方阵线性变换的角度也可以理解，若$\vec{x}$是一个方阵 A 的特征向量，即 A$\vec{x}$=λ$\vec{x}$，即有 A(k$\vec{x}$)=λk$\vec{x}$，所以同方向的 k$\vec{x}$也是 A 的对应于特征值λ的特征向量。当然，对于同一特征值所对应的同方向的特征向量而言，一般我们可只取其中的一个进行典型研究。

例二：求 A=$\begin{pmatrix}3 & 1\\ -1 & 1\end{pmatrix}$的特征值和特征向量

其特征方程为$\begin{vmatrix}3-\lambda & 1\\ -1 & 1-\lambda\end{vmatrix}$=0，特征多项式为

（3-λ）（1-λ）+1=λ^2-4λ+4=（λ-2）^{2}，令其等于零解得λ_1=λ_2=2。这时，特征多项式有两个重根。

其对应特征向量满足$\begin{pmatrix}3-2 & 1\\ -1 & 1-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$，解得$\begin{pmatrix}1\\-1\end{pmatrix}$是其一个特征向量。由于特征值重根，所以仅有这一个方向的特征向量。

现在我们知道，对一个 n 阶方阵而言，在复数范围内其对应的特征向量有 n 个，但其中可能有复数特征向量（虽然我们表示无视它），可能有重根的特征向量。下面我们着重关注有 n 个不重复（互异）的实数根的情况。

如果一个 n 阶方阵 A 有 n 个互异的实数特征值时，它们对应的 n 个特征向量是否可能存在相同的情况呢？任取方阵 A 的两个互异特征值λ_1、λ_2，设其对应特征向量为$\overrightarrow{x1}$、$\overrightarrow{x2}$，于是有 A$\overrightarrow{x1}$=$\lambda_1\overrightarrow{x1}$，A$\overrightarrow{x2}$=$\lambda_2\overrightarrow{x2}$。如果$\overrightarrow{x1}$=$\overrightarrow{x2}$，则有$\lambda_1\overrightarrow{x1}$= A$\overrightarrow{x1}$=A$\overrightarrow{x2}$=$\lambda_2\overrightarrow{x2}$=$\lambda_2\overrightarrow{x1}$，即$\lambda_1\overrightarrow{x1}$=$\lambda_2\overrightarrow{x1}$得$\lambda_1$=$\lambda_2$。这与$\lambda_1$、$\lambda_2$互异的假设矛盾。所以互异

风萧萧兮乱翻书

的特征值对应的特征向量不可能相等。由于与$\overrightarrow{x1}$同向的向量都是λ_1对应的特征向量，与$\overrightarrow{x2}$同向的向量都是λ_2对应的特征向量，所以容易证明互异的特征值对应的特征向量也不可能共线。用线性代数的语言来说就是两两互异的特征值对应的特征向量两两线性无关。

要注意，我们虽然已经讨论了两两互异的特征值对应的特征向量两两线性无关，这只是说明这 n 个互异的特征值对应的特征向量都不共线，但并没有证明它们不共面，更没有证明它们是线性无关组。当然，有了这个开端，我们自然希望能够推导下去。

任取 A 的三个互异特征值λ_1、λ_2、λ_3，设其对应特征向量为$\overrightarrow{x1}$、$\overrightarrow{x2}$、$\overrightarrow{x3}$，于是有 $A\overrightarrow{x1}=\lambda_1\overrightarrow{x1}$，$A\overrightarrow{x2}=\lambda_2\overrightarrow{x2}$，$A\overrightarrow{x3}=\lambda_3\overrightarrow{x3}$。假设$\overrightarrow{x1}$、$\overrightarrow{x2}$、$\overrightarrow{x3}$线性相关，那么三者可相互线性表示，即有$\overrightarrow{x3}=\mu_1\overrightarrow{x1}+\mu_2\overrightarrow{x2}$（$\mu_1$、$\mu_2\in R$）。

由此有：$A\overrightarrow{x3}=A(\mu_1\overrightarrow{x1}+\mu_2\overrightarrow{x2})=\mu_1(A\overrightarrow{x1})+\mu_2(A\overrightarrow{x2})=\mu_1\lambda_1\overrightarrow{x1}+\mu_2\lambda_2\overrightarrow{x2}$.

由 $A\overrightarrow{x3}=\lambda_3\overrightarrow{x3}$ 又可得：$A\overrightarrow{x3}=\lambda_3(\mu_1\overrightarrow{x1}+\mu_2\overrightarrow{x2})=\lambda_3\mu_1\overrightarrow{x1}+\lambda_3\mu_2\overrightarrow{x2}$.

于是有：$\mu_1\lambda_1\overrightarrow{x1}+\mu_2\lambda_2\overrightarrow{x2}=\lambda_3\mu_1\overrightarrow{x1}+\lambda_3\mu_2\overrightarrow{x2} \Rightarrow (\mu_1\lambda_1-\mu_1\lambda_3)\overrightarrow{x1}=(\mu_2\lambda_3-\mu_2\lambda_2)\overrightarrow{x2}$

这得出了$\overrightarrow{x1}$、$\overrightarrow{x2}$线性相关的结论，这与我们前面证明的$\overrightarrow{x1}$、$\overrightarrow{x2}$两两线性无关的结论矛盾。所以$\overrightarrow{x1}$、$\overrightarrow{x2}$、$\overrightarrow{x3}$线性相关的假设前提不成立，即$\overrightarrow{x1}$、$\overrightarrow{x2}$、$\overrightarrow{x3}$线性无关得证！

按照数学归纳法和同样的推理，可以相应得证明 n 个互异的特征值对应的 n 个特征向量组成的特征向量组线性无关。也就是说，如果 n 阶方阵 A 有 n 个互不相等的特征值，那么一定能找到其 n 个线性无关的特征向量。

风萧萧兮乱翻书

§5 相似矩阵

我们先来总结下前面已经讨论过的，从不同的角度对方阵 A 及其与列向量 x 的乘积 Ax 的几种不同的涵义解读：首先，矩阵可以看做是列（行）向量组，矩阵与列向量的乘积可看做是矩阵的行与列向量相应元素的线性组合；其次，矩阵还可以看做一种线性变换，Ax 是这种线性变换作用于列向量 x 的一个输出向量；如果构成矩阵 A 的列向量组线性无关，则矩阵 A 可以看做是一组基，Ax 可以看做是声明在 A 这组基下的坐标为 x 的一个向量。

现在我们要把上面的描述综合到一起运用了。如果 A、B 是两个变换矩阵，P 是一组基，A(Px)可以看做是变换 A 作用于 P 这组基下的坐标为 x 的向量，P(Bx)则可以看做是变换 B 作用于自然基下坐标为 x 的向量在 P 这组基下的坐标。如果，对于任何 x，两者的结果都相等，即 A(Px) =P(Bx)，那么可以看出 A、B 其实是分别作用于 P 这组基下和 E 这组基下的同一变换，这时，我们称 A、B 为相似矩阵。

由 A(Px) =P(Bx) $\Leftrightarrow$ AP=PB $\Leftrightarrow$ A=PBP^{-1}，或者我们可以从 A=PBP^{-1} 再来探究一下变换的情况。对于一个向量 x，Ax 可以看做是 A 变换作用于 x 的结果，PBP^{-1}x 则可以看做（从右往左）先将 x 转换为 P^{-1} 这组基下的坐标，然后对其作用 B 变换，最后将结果再转换回 E 这组基下的坐标。而由 Ax=PBP^{-1}x 知道对于任何 x 两种运算结果都相等，由此可以知道 A、B 是作用于不同基下的同一变换。

还是以速度做个类比吧，A、B 都是对速度进行的一种线性变换，等式左边表示 A 对于 x（米/秒）的变换结果，等式右边表示先将 x（米/秒）转化为 y（公里/小时），然后对 y（公里/小时）进行 B 变换，最后将变换的结果 By 再转化为（米/秒）。两者的结果始终相等，这当然说明 A、B 是分别作用于（米/秒）和（公里/小时）同一变换。

来看看相似矩阵的定义：

定义：设 A，B 都是 n 阶矩阵，若有可逆矩阵 P，使 P^{-1}AP=B，则称 B 是 A 的相似矩阵，或者说矩阵 A 和 B 相似。对 A 进行运算 P^{-1}AP 称为对 A 进行相似变换，可逆矩阵 P 称为把 A 变成 B 的相似变换矩阵。

再来个实例巩固下吧：设 A=$\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$，P=$\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix}$，可得 P^{-1}=$\begin{pmatrix}3 & -2\\ -1 & 1\end{pmatrix}$，B=

风萧萧兮乱翻书

$P^{-1}AP=\begin{pmatrix}3 & -2\\ -1 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix}=\begin{pmatrix}3 & 6\\ -1 & -2\end{pmatrix}$。看起来，$A=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$和 $B=\begin{pmatrix}3 & 6\\ -1 & -2\end{pmatrix}$实在是没什么“相似”之处。

来看看两个变换的作用吧。任取一个 $x=\begin{pmatrix}2\\ 1\end{pmatrix}$，A 是我们前面已经熟悉的投影变换矩阵，$Ax=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix}=\begin{pmatrix}2\\ 0\end{pmatrix}$；来看看 B 变换的作用，B 变换是针对 P 这组基下的，先求 $x=\begin{pmatrix}2\\ 1\end{pmatrix}$在 P 下的坐标，由$\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix}y=\begin{pmatrix}2\\ 1\end{pmatrix}$，求得 $y=\begin{pmatrix}4\\ -1\end{pmatrix}$，B 作用于 y 的变换 $By=\begin{pmatrix}3 & 6\\ -1 & -2\end{pmatrix}\begin{pmatrix}4\\ -1\end{pmatrix}=\begin{pmatrix}6\\ -2\end{pmatrix}$，也就是经 B 变换后得到的输出向量在 P 为基下的坐标为$\begin{pmatrix}6\\ -2\end{pmatrix}$。这个向量在 E 下的坐标呢？$\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix}\begin{pmatrix}6\\ -2\end{pmatrix}=\begin{pmatrix}2\\ 0\end{pmatrix}$！当然，你任取 x，都会得到同样相等的结果。因为 A、B 本就是分别以 E 和 P 为基下的同一变换。

既然，相似矩阵是在不同基下的**同一线性变换**，那么它们当然应该具有相同的跟线性变换有关的性质，比如我们上节讨论的线性变换矩阵的特征值、特征向量、特征方程、迹等等（证明就不描述了，其实只要证明其特征方程相同，其他的自然就相同了。特征方程的证明也比较容易，教科书上都有）。而且，它们的行列式和秩也相同。（记得所有特征值的积等于行列式吗？）

既然，相似矩阵是在**不同基下**的同一线性变换，那么对于线性变换矩阵 A，它的相似矩阵当然不是唯一的，另取一组基就应该能得到一个这组基下的 A 的另一个相似矩阵，而这一系列相似矩阵当然互相还是相似的，也都具有相同的特征值、特征向量等性质。

既然这样，我们当然希望能够取得某组特殊的基，使得在这组基下的特殊相似矩阵能够极其方便地求出特征值。对角阵显然就是这样一个能够方便求得特征值的特殊矩阵。

对于一个方阵 A，如果能够找到一组基 P，使得其相似矩阵 $P^{-1}AP$ 为一个对角阵 Λ，就称为把矩阵 A 对角化。

风萧萧兮乱翻书

如何找到这样一组基 P 呢？由 $P^{-1}AP=\Lambda \Rightarrow AP=\Lambda P$，将 P 写成列向量组（$p_1$，$p_2$，…，$p_n$）形式，可得

$$A(p_1, p_2, \ldots, p_n)=\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}(p_1, p_2, \ldots, p_n)$$

$$=(\lambda_1 p_1, \lambda_2 p_2, \ldots, \lambda_n p_n)$$

也就是 $A p_i = \lambda_i p_i$ (i=1，2，…，n)。这里 p_i 是 P 的第 i 个列向量。从这个式子看，p_i 不就是 A 的特征向量吗？当然，现在要求 P 是一组基，也就是说，P 的列向量组要是线性无关的，所以如果能找到 A 的 n 个线性无关特征向量，就可以得到这个 P。

定理：n 阶矩阵 A 与对角阵相似（即 A 能对角化）的充分必要条件是 A 有 n 个线性无关的特征向量。

上一节我们已经知道，如果 A 的 n 个特征值互异，那么就能找到 A 的 n 个线性无关的特征向量（当然包括可能出现的烦人的复数）。但反过来说，A 如果没有 n 个互异的特征值（特征值有重根），却不能断定 A 一定没有 n 个线性无关的特征向量。最简单的反例是 E_n，显然 E_n 的 n 个特征值是 n 重根，都是 1。可显然所有 n 维向量都是 E_n 的特征向量。所以，矩阵 A 有 n 个互不相等的特征值是 A 能对角化的充分条件但却不是必要条件。同济版的教科书上说，一个 n 阶矩阵具备什么条件才能对角化，这是一个较复杂的问题，我们对此不进行一般性的讨论。

同济版的教科书上的下一句话是：“而仅讨论当 A 为对称阵的情形”。这句话对我们的讨论而言极其重要，因为它随后还证明实对称阵的特征值为实数！（证明我就不摘录了，看到复数我头晕。）

前面我们已经知道，n 阶实对称阵 A 的有 n 个实特征值（包括重根）。我们先看看不相等的特征值对应的特征向量的关系。设 λ_1、λ_2 是对称阵 A 的两个不相等的特征值， p_1、 p_2 为其对应的特征向量，根据特征向量的关系有：

$$A p_1=\lambda_1 p_1 \quad , \quad A p_2=\lambda_2 p_2$$

因为 A 为对称阵，及 $A^T=A$，则有：$(A p_1)^T= p_1^T A^T=p_1^T A$，同时 $(A p_1)^T=(\lambda_1 p_1)^T=\lambda_1 p_1^T$，于是有：$\lambda_1 p_1^T= p_1^T A$。两边乘 p_2，得：

$$\lambda_1 p_1^T p_2 = p_1^T A p_2 = p_1^T(\lambda_2 p_2) = \lambda_2 p_1^T p_2$$

即：$(\lambda_1-\lambda_2) p_1^T p_2=0$。根据前提假设 $\lambda_1 \neq \lambda_2$，故 $p_1^T p_2=0$，即 p_1 与 p_2 正交。

由此我们得出结论一，对称阵 A 的互异的特征值对应的特征向量两两正交。由此，若 n 阶对称阵 A 有 n 个互异的特征值，即可获得 n 个两两正交的特征向量构建一个正交基 P，使得 A 对角化，即 $P^{-1}AP=\Lambda$。

下面我们来看看重根的情况。教科书上有一个定理：

定理： 设 A 为 n 阶对角阵，λ 是 A 的特征方程的 r 重根，则矩阵 A−λE 的秩等于 n−r，从而对应特征值 λ 恰有 r 个线性无关的特征向量。

根据这个定理，我们先来看看两重根的情况。设 λ_1，λ_2，…λ_n 是对称阵 A 的 n 个特征值，其中 $\lambda_1=\lambda_2$ ，其余特征值均互异。依据上面的定理，仍有对应特征值的 n 个线性无关的特征向量 p_1，p_2，…，p_n，其中，p_3，p_4，…，p_n 两两正交，且同时与 p_1，p_2 正交。也就是说，除了不能确定 p_1，p_2 正交外，其余特征向量均两两正交。能否证明 p_1，p_2 也正交呢？可实际上我们求得的 p_1，p_2 并不一定正交。那么能否找到一个特征向量 p，替换 p_1 或 p_2，使得 n 个特征向量均正交呢？

设 $\lambda=\lambda_1=\lambda_2$，根据特征向量的关系有：$A p_1=\lambda p_1$，$A p_2=\lambda p_2$。⇒

$$A(p_1+p_2) = Ap_1+A p_2=\lambda p_1+\lambda p_2=\lambda (p_1+p_2)$$

这说明 p_1、p_2 的和向量 (p_1+p_2) 也是对应于特征值 λ 的特征向量。既然这样，向量 $p_1+(p_1+p_2)=2p_1+p_2$ 及 $p_2+(p_1+p_2)=p_1+2p_2$ 也都是对应于特征值 λ 的特征向量。再结合 μp_1，μp_2（$\mu \in R$）也为 A 的特征向量的情况，继续推导下去不难得出 $p=k_1 p_1+k_2 p_2$（$k_1, k_2 \in R$）都是 A 对应于特征值 λ 的特征向量。这说明，p_1、p_2 所在平面（二维向量子空间）的向量都是对应于特征值 λ 的特征向量。

前面我们已经谈到，p_3，p_4，…，p_n 都与 p_1，p_2 正交，而 p_1，p_2 线性无关，这说明它们都与 p_1、p_2 所在平面（二维向量子空间）正交，即与 p_1、p_2 所在二维向量子空间的所有向量正交（这是立体几何的知识，不用解释了吧）。于是我们只要在这个平面（二维向量子空间）中选择一个与 p_1 正交的向量 p 代替 p_2（甚至可以任取两个正交向量代替 p_1，p_2），即可获得 n 个两两正交的特征向量构建一个正交基 P，使得 A 对角化，即 $P^{-1}AP=\Lambda$。

与此类似，如果 A 的特征值有三重根 λ，则对应 λ 有三个线性无关特征向量，这三个线性无关特征向量为基的三维向量子空间中的向量均为 A 的对应特征值 λ

风萧萧兮乱翻书

的特征向量，且其余特征值对应的特征向量与这个三维向量子空间正交。在这个三维向量子空间中选取三个两两正交的特征向量，与其余特征值对应的特征向量即可构建一个正交基 P，使得 A 对角化，即 $P^{-1}AP=\Lambda$。同理可推广至 r 个重根。

综上所述，可得出下面的定理：

定理：设 A 为 n 阶对称阵，则必有正交阵 P，使 $P^{-1}AP= P^{T}AP=\Lambda$，*其中 Λ 是以 A 的 n 个特征值为对角元的对角阵。*

实际操作中，将一个对称阵 A 对角化，一般先求出 A 的 n 个特征值，对于不重根的特征值，可直接求出其对应的特征向量；对于 r 重根的特征值 λ_i，求方程（$A-\lambda_iE$）$x=0$ 的基础解系，得到 r 个线性无关特征向量；将重根的特征值对应的 r 个特征向量中的 r-1 个正交化，根据前面讨论的结论，正交化得到的向量仍是 A 的对应于特征值 λ_i 特征向量，最终可以得到 A 的 n 个两两正交的特征向量；将 n 个两两正交的特征向量单位化，可以得到一个正交阵 P，由 $P^{-1}AP= P^{T}AP=\Lambda$ 即可将 A 对角化。

风萧萧兮乱翻书

§6 算不上小结的总结

根据我们前面讨论的情况，我们曾经说过，从向量组的角度，矩阵可以看做是列向量组或者行向量组；从向量空间的角度，可逆矩阵可以看做是向量空间的一组基，矩阵也可以看做是向量组的坐标；从线性变换的角度，矩阵是线性变换，矩阵的乘法 AB 就是对 B 进行 A 线性变换。当然，从线性变换的角度进一步讨论了特征向量、特征值及在不同的基下的同一线性变换——相似矩阵。那矩阵到底是啥？……变形金刚？

拿线性变换来说吧，我们说矩阵乘法 AB=C 可以看做是对 B 进行 A 线性变换。从这个角度来说，矩阵 A 是一种线性变换，矩阵 B 又是线性变换的对象，而矩阵 C 是这种线性变换的结果。此外，由 AE=A 可解释为对 E 进行线性变换 A 得到的结果是 A。这似乎有点太绕了吧。

说实话，我也绕不清楚了。还是来类比一下吧。数的乘法当然可以看做是一种倍乘变换，$3\times p$ 就是对 p 进行 3 倍乘变换，说通俗点就是把输入变量 p 乘以 3；同时 3 也可以成为倍乘变换的对象，4×3 就是对 3 进行 4 倍乘变换；3 当然也是对 1 进行 3 倍乘变换即 3×1 的结果。然后问个同样的问题，“数”究竟是什么？

这算什么类比！“数”不就是……？！？也是啊，“数”究竟是什么？

维基百科上说——“数是一个用作计算、表示数量或用作量度的抽象概念，是比较同质或同属性事物的等级的简单符号记录形式（或称度量）”，去掉相关的修饰定语就是“数是简单符号记录形式”。我不知道这算不算“数”的标准定义。可这算什么定义啊？跟“矩阵是一个数表”的定义一样“无厘头”嘛！

其实，数本身并没有倍乘变换的涵义，有了数的乘法才能将 $3\times p$ 看做是对 p 进行 3 倍乘变换。数可以是数量，可以是排序，可以是坐标，甚至可以是电话号码。同样，有了向量的定义后，才能将矩阵看成是向量组；在定义了矩阵乘法后，才有线性变换的相关涵义；定义了向量空间后，才能将可逆矩阵看做一组基，也才有坐标的概念。所以，也许可以说，针对不同的计算，在不同的应用中可以赋予矩阵不同的涵义，从不同的角度可以不同的理解矩阵的“实质”。而矩阵本身和数一样，只是一个貌似简单，却极端抽象的对象。

是不是等于什么也没说？好吧，除此之外，我没什么好说的了！

结 语

由于本人的水平实在有限，本书到此就将结束了。在结束前，我还想和大家聊聊自己的一些感慨。

数学（当然其他科学也类似）就像一片浩瀚的海洋，里面有着无穷的宝藏，许多的前辈大师曾经在上面驰骋、冲浪、扬帆，他们用他们的努力寻找、挖掘出了许多宝物，并以他们的名字来命名这些宝物，而他们也因此成为数学史上一个个熠熠生辉的星斗。

我们的老师习惯于教给我们，大师们发现的这些宝物有什么性质、什么特点、什么功效，并教我们运用它们来解答一个个难题。这让我们更加体会到这些宝物是多么的神奇，而发现它的人是多么的了不起，同时又是多么的遥不可及。于是，我们小心翼翼地站在海边，遥望着看起来神秘、深邃的大海，遗憾着我们还不能娴熟掌握“牛顿分水杖”、“高斯避水珠”、“拉普拉斯水上漂”这些宝物的性质、特点。于是……我们在海边等待并景仰着新的大师们找回更多的宝物。

可是，我们的老师却很少告诉我们，大师们是如何发现这些宝贝的。观看网易公开课中一些国际名校的教学视频给我最大的感触就是，这些老师更乐意告诉学生，某位前辈是如何走过一片浅滩找到了一串珍珠，某位大师是如何绕过一片礁石发现了一个洞穴，甚至于有位精英只是在沙滩上搬开了一块石头却发现了下面精美的贝壳。从这个意义上来说，我觉得这些老师更应该称为——“导师”。而在导师引导我们重走大师们发现宝藏的历程时，我们才发现大师们用的也不都是“莱布尼兹探测器”、“拉格朗日定位仪”等尖端工具，他们用的工具或许是我们都会用的手杖、绳梯或撬棍。

有一个令我印象深刻的例子是，麻省理工学院的老教授 William Gilbert Strang 讲解完“高斯—约当消元”的时候，对学生们说“如果让你们来做的话，你们也能整出这么一套东西。这方法之所以以高斯的名字命名，只不过是因为他生的比你们早！”顿了顿，他还不忘再补充了一句“当然，他死的也比你们早。”在幽默风趣的背后，他在向他的学生们传递一个思想——看看，这没什么高深莫测。是的，你们也可以！！我们在熏陶学生膜拜大师的时候，人家在引导学生超越大师。我们在遥望大海，踟蹰不前，感慨万千的时候，人家在引领学生走向大海，沿着大师用手杖、绳梯或撬棍找寻宝物的足迹，更加深刻的掌握宝物的来历

风萧萧兮剑翻书

和使用。在这个过程中，导师更加鼓励学生们，哪怕还不会扬帆、甚至还不会游泳，也可以走近大海，哪怕试着在沙滩上掀开一块块石头，即使什么也没发现，其实也并非一无所获。

我们似乎从小就对许多名人儿时的聪明故事耳熟能详：临危不乱、砸缸救人的司马光；利用镜子的反光原理，让医生在明亮的反光下，为妈妈成功进行手术的爱迪生；九岁就能心算出“$1+2+3+\dots+100$”的高斯……这些故事潜移默化地告诉我们，大师们是如何的聪明而不可望其项背。但是，引导我们由 $1+100=101$，$2+99=101$，……从而得出 $1+2+3+\dots+100=50\times101=5050$ 的做法，显然比直接教给我们等差数列的求和公式更加能让我们豁然开朗，虽然从数学上来说后者显然比前者更具有普遍意义。我们也会发现心算“$1+2+3+\dots+100$”所要使用的工具只是撬棍，而找到 $1+100=101$，$2+99=101$……这个“支点”的方法才更让我们受益无穷。

这或许是这么多天来，我以一个数学草根“无知者无畏”的执着，写下前面这洋洋洒洒数十页的最大动力所在。我就是想试图用我们已有的手杖、绳梯和撬棍，以最简单的知识背景，诠释线性代数中曾经让我感觉高深、迷茫的东西。当然我只是说——“试图”！我前面写的东西可能有很多是站不住脚，甚至是错误的，但我还是把我所思所想写出来了。对于其中存在的错误，在欢迎专家、网友们指正的同时，也希望能够以最简单的方式告诉我们正确的理解。

最后，感谢网易、新浪等提供公开课的网站，这实在是一件功德无量的大好事；感谢《矩阵的理解》的作者孟岩，《线性代数的几何意义》的作者任广千、胡翠芳以及很多不知名的网友，他们的精彩文章以及很多网友对线性代数的独到论述使我受益匪浅。此外，还有北京航空航天大学的李尚志老师，他的讲课视频“浅入深出”，让人豁然得解，也让我们知道国内并不是没有优秀的老师的。遗憾的是网上李尚志老师的线性代数讲课视频不全，而且是针对数学专业的学生上的，“浅入”确实启发很大，“深出”像我辈草根就跟不上了。衷心希望国内的优秀大学也能将其优秀课程的真实、完全的教学视频以公开课的形式给广大网友共享学习。比较网易公开课的中外名校讲课视频就能发现，其中国外名校的讲课视频大部分是给学生上课的完整真实视频，一般都是二、三十集以上，甚至学生的

行为艺术，课堂文化都真实记录在内；而国内名校的视频则更像是一种特意的讲座，一般都是几集。你在大学有哪门真正的课程是上几节课的吗？

好了，如果你觉得我写的这些东西还有点帮助，欢迎推荐、复制、转发，但请注明出处和作者，也好满足下本人小小的虚荣心。如果你感念本人的辛苦，还愿意往我这个可怜草根的支付宝（**fengshou94@vip.sina.com**）资助个十块八块的，本人更将不胜感激。知识是无价的，支持却可以有价！呵呵。

风萧萧兮乱翻书