

29 《新理解矩阵1》：矩阵是什么？

Oct By 苏剑林 | 2012-10-29 | 4000位读者

前边我承诺过会写一些关于自己对矩阵的理解。其实孟岩在《理解矩阵》这三篇文章中，已经用一种很直观的方法告诉了我们有关矩阵以及线性代数的一些性质和思想。而我对矩阵的理解，大多数也是来源于他的文章。当然，为了更好地理解线性代数，我还阅读了很多相关书籍，以求得到一种符合直觉的理解方式。孟岩的blog已经很久没有更新了，在此谨引用他的标题，来叙述我对矩阵的理解。

当然，我不打算追求那些空间、算子那些高抽象性的问题，我只是想发表一下自己对线性代数中一些常用工具的看法，比如说矩阵、行列式等。同时，文章命名为“理解矩阵”，也就是说这不是矩阵入门教程，而是与已经有一定的线性代数基础的读者一起探讨关于矩阵的其他理解方式，仅此而已。我估计基本上学过线性代数的读者都能够读懂这篇文章。

首先，我们不禁要追溯一个本源问题：矩阵是什么？

我们不妨回忆一下，矩阵是怎么产生的。矩阵可以看成是一个个向量的有序组合，这说明矩阵可以类比向量；但是向量又是怎么产生的？向量则是一个个数字的有序组合，这又把我们的研究方向指向了“数字是什么”这个问题上。比如，数字1是什么？它可以代表1米，可以代表1千克，也可以代表1分钟、1摄氏度甚至1个苹果。它为什么有这么多的表示意义？答案很简单，因为在本质上，它什么都不是，它就是数字1，一个记号，一个抽象的概念。正因为它抽象，它才可以被赋予各种各样直观的意义！回到矩阵本身，我们才会明白，矩阵的作用如此之大，就是因为书本上那个很枯燥的定义——矩阵就是m行n列的一个数表！它把矩阵抽象出来，让它得到了“进化”。它是一个更一般化的概念：一个向量可以看作一个矩阵，甚至一个数都可以看成一个矩阵，等等。

代数方面的理解

当然，上述说法是含糊的，我们还是需要确切知道它究竟有什么用？这可以从代数和几何的角度来分析，因为做到数形结合才是最完美的。首先我们知道数学最基本的元素就是数字，严格来说是自然数，如0, 1, 2, ...；有了数字，我们就可以做到很多东西。但是数字是单一的，而我们很多时候都要批量处理一些类似的运算，比如同时要计算 $1+2$, $1+3$, $2+3$, $4+5$ 这四个算式。不论是从记录还是从研究的角度来说，分开研究它们都是比较繁琐的。于是一种“批量”的记号产生了，我们记为 $(1,1,2,4)+(2,3,3,5)$ ，用两个不同记号记录它们，比如

$\vec{A} = (1, 1, 2, 4)$, $\vec{B} = (2, 3, 3, 5)$ ，我们就可以将它记为 $\vec{A} + \vec{B}$ 。这样不论在研究还是记录方面都能够给我们方便。于是一个我们称之为“向量”的东西产生了，也就是说，从代数的角度来讲，向量是为了研究批量运算而产生的。但是向量并没有解决所有的批量运算的问题。比如3元一次方程组

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

单单用向量我们还是没有办法很好地研究这一类问题。于是我们就要想法子创造出一些新的记号，由于左边的系数的具有一定的排列顺序和统一的形式，我们不妨把它们单独写出来

$$[a_{21}, a_{22}, a_{23}]$$

$$[a_{31}, a_{32}, a_{33}]$$

并用一个简单的符号 A 来表示它，然后把未知数和右边的常数都分别写成向量形式 $x = [x_1, x_2, x_3]^T$ 和 $b = [b_1, b_2, b_3]^T$ （多加了上标T表示列向量）。我们期待上面的方程组可以写成一个简单的形式

$$Ax = b$$

由此我们可以定义一个3阶方阵乘以一个3维列向量的乘法了，这是一种**纯粹的定义**，是为了方便我们记录和研究定义。在此基础上，我们就可以研究更多的东西，比如矩阵乘矩阵会得到什么？

同样，这里要研究的矩阵都是指n阶方阵这个最核心的东西，我们要先把核心问题研究透彻，不然一开始就考虑所有的繁杂的情况，容易让我们陷入迷惘中而不知所措。

在研究一般乘法之前，我们先来了解一下关于运算定律问题。我们知道在实数中，加法满足结合律和交换律，乘法满足结合律、交换律和分配律。哪些定律可以迁移到矩阵乘法中的呢？交换律是无法先验的，它是个定义问题，我可以定义它成立也可以定义不成立，但是为了运算的方便，我们还是希望它满足更多地运算定律，所以我们先来考虑结合律，希望它能够满足这一定律。也就是说

$$(AB)x = A(Bx)$$

其中 Bx 已经是我们所熟知的运算（由定义而来），它将得到一个列向量，所以我们可以轻易算出 $A(Bx)$ ，从直观上来讲， AB 应该也是一个n阶方阵，我们可以先把它设出来，然后与列向量 x 进行运算，最后把两边的结果一一对应起来，就得到了 AB 这个n阶方阵中各个元素的表达式。我们最终可以发现，它就是我们书本上定义的表达式。

以2阶方阵为例，令 A 为

$$\begin{bmatrix} a & b \end{bmatrix}$$

$$\begin{bmatrix} c & d \end{bmatrix}$$

B 为

$$\begin{bmatrix} e & f \end{bmatrix}$$

$$\begin{bmatrix} g & h \end{bmatrix}$$

AB 为

$$\begin{bmatrix} p & q \end{bmatrix}$$

$$\begin{bmatrix} r & s \end{bmatrix}$$

并令 $x = [x, y]^T$ ，那么 $(AB)x$ 就等于

$$\begin{bmatrix} px + qy & rx + sy \end{bmatrix}^T$$

而 $Bx = \begin{bmatrix} ex + fy & gx + hy \end{bmatrix}^T$ ，那么 $A(Bx) = \begin{bmatrix} aex + afy + bgx + bhy & cex + cfy + dgx + dhy \end{bmatrix}^T$

那么根据各个元素的对应，就得到 $p=ae+bg, q=af+bh, r=ce+dg, s=cf+dh$ 。这就完成了2阶方阵乘法的定义。

现在我们就可以从代数的角度来讲，**矩阵是为了简化批量线性运算的一个“终极武器”**！这就是矩阵的一个比较直观和有用的代数意义。

如果根据我们这个定义去考虑交换律，我们会发现矩阵一般不符合交换律。这不能不说是一个遗憾。但是没关系，它服从结合律这一个事实，已经赋予了这个工具极大的力量。比如线性方程组 $Ax = y$ ，我们有 $By = B(Ax) = (AB)x$ ，如果我们想办法找到一个矩阵 B ，使得 $AB = I$ ，那么就很棒了，因为我只要用矩阵 B 作用于向量 y 就可以得到方程组的解了，事实上这样的矩阵 B 是存在的，这就是逆矩阵。要是没有结合律，这一切都免谈！

由于这是实数基本运算（线性运算）的“批量版”，那么我们就可以很自然地把实数的一些公式延伸为矩阵版（只要不是涉及到交换律就行）。比如，在实数中，我们有公式

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots \approx 1 + x$$

那么我们求矩阵的逆阵时，也有类似的公式

$$(I - A)^{-1} = I + A + A^2 + A^3 + \dots \approx I + A$$

其中 I 是单位矩阵， A 是一个“比较小”的矩阵。至于“比较小”怎么定义，现在还说不清楚，可以认为是矩阵的行列式值比较小。类似的，根据

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x$$

也能够相应地给出 $(I + \frac{1}{2}A)^2 \approx I + A$ ，这是求矩阵“平方根”的一个近似公式。

得益于我们定义的矩阵乘法，批量的运算可以直接用单个量的运算公式进行，不用我们煞费苦心、绞尽脑汁地构思新的公式。这就是矩阵的强大所在！它在解决很多线性问题时有着奇迹般的美妙，最简单的例子莫过于线性方程组 $Ax = y$ 的解为 $x = A^{-1}y$ ，解答方程组的时候就好像求解一元方程那样有简单的形式！还有一些关于指数的定义等等，以后在应用时会把它介绍的。它们都好像非常精美的“艺术品”！

下一回，我们将从几何角度来理解矩阵。当然，这里边的绝大多数内容在孟岩的文章里头都已经提到了，我只是重提旧论而已，希望读者不会厌烦。

转载到请包括本文地址：<https://spaces.ac.cn/archives/1765>

31 《新理解矩阵2》：矩阵是什么？

Oct By 苏剑林 | 2012-10-31 | 1822位读者

上一篇文章中我从纯代数运算的角度来讲述了我对矩阵的一个理解，可以看到，我们赋予了矩阵相应的运算法则，它就在代数、分析等领域显示出了巨大作用。但是纯粹的代数是不足够的，要想更加完美，最好是找到相应的几何对象能够与之对应，只有这样，我们才能够直观地理解它，以达到得心应手的效果。

几何理解

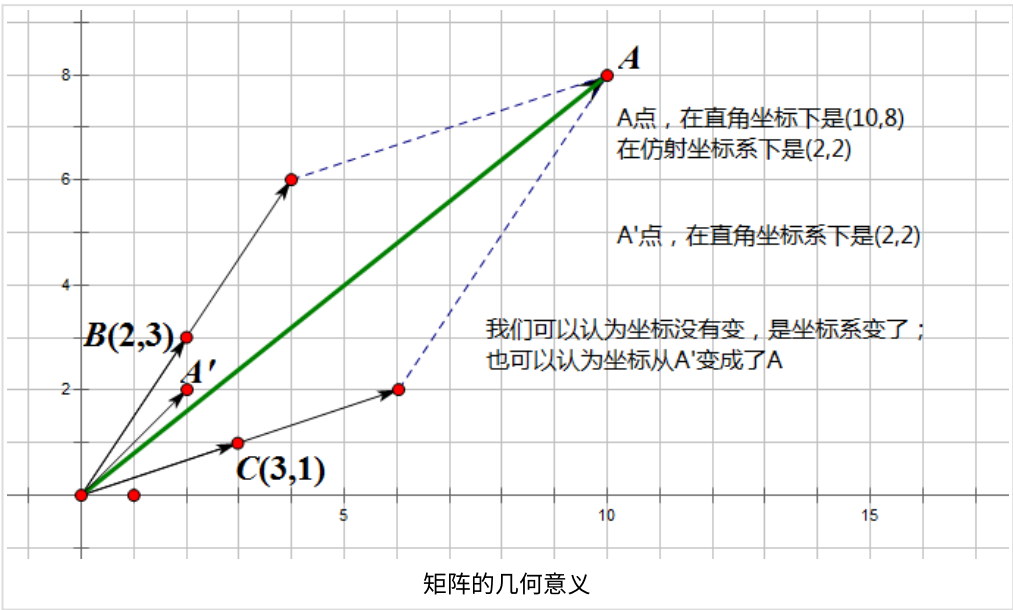
我假设读者已经看过孟岩的《理解矩阵》三篇文章，所以更多的细节我就不重复了。我们知道，矩阵A

$[a_{21}, a_{22}]$

事实上由两个向量 $[a_{11}, a_{21}]^T$ 和 $[a_{12}, a_{22}]^T$ （这里的向量都是列向量）组成，它描述了一个平面（仿射）坐标系。换句话说，这两个向量其实是这个坐标系的两个基，而运算 $y = Ax$ 则是告诉我们，在A这个坐标系下的x向量，在I坐标系下是怎样的。这里的I坐标系就是我们最常用的直角坐标系，也就是说，任何向量（包括矩阵里边的向量），只要它前面没有矩阵作用于它，那么它都是在直角坐标系下度量出来的。

（事实上，单位矩阵I是默认的直角坐标系，这一说法并非总是成立的，但是我们现在寻求直观的理解方式，我们就用最简单的东西来实行。）

太多的文字未必能够把问题说清楚，我们需要一张图来解释一下：



图上所用的矩阵A是

$[3,2]$

$[1,3]$

这构成了一个仿射坐标系，在这个坐标系下，有一个向量 $x = [2, 2]^T$ ，它在直角坐标系下测得的坐标为 $[10, 8]^T$ ，现在不难发现，直接用矩阵乘法来计算，有

$Ax = [3 \cdot 2 + 2 \cdot 2, 1 \cdot 2 + 3 \cdot 2]^T = [10, 8]^T$

正是我们所期待的！

为什么会有这样的特点？其实这源于我们对矩阵乘法的定义，反过来，如果我们用这样的几何方式来定义矩阵乘法，那么我们将得到在书本上了解到的矩阵乘法计算公式。更高阶的矩阵也可以作同样的类比。推导过程只是一道很简单的练习题，读者不妨自己动笔尝试一下？

现在我们又回到孟岩文章上的说法了，对于矩阵作用于一个向量（对应的一个点），我们既可以看作点没有变，只不过是坐标系从直角坐标系变换为仿射坐标系而已；另一方面，我们也可以看做矩阵把直角坐标系的一个A'点“运动”（变换）到了A点。这两种说法都行，正如孟岩所说的“运动是相对的”。更正确地讲，两种说法都要同时被提及，才算是最好的理解。矩阵是一个点到另外一个点的变换，变换的方式就是坐标系的变换。

当然，上面只讨论了矩阵乘以向量的乘法，那么矩阵乘以矩阵呢？比如 AB ，我们就可以看作是矩阵 B 给出了一个坐标系，但是这个坐标系的各个分量是在 A 坐标系下测量得到的，而 A 是在直角坐标系下测量得到的，所以要把 B 的各个分量（列向量）与矩阵 A 作乘法后，才得到了这个仿射坐标系在直角坐标系下的“像”。这很直接地导致了矩阵乘以矩阵的计算公式，也很显然地回答了“为什么 n 阶方阵只有与 n 阶方阵相乘才有意义”，因为两者要在同一空间中测量，才能够完整而唯一地把测量值确定下来。正如，在 $n+1$ 维的空间中讨论 n 个 n 维向量是没有意义的，因为在 $n+1$ 维空间中的观测者看来，它们只不过是一个“面”，多出的一个维度可以随意变化；在 n 维空间中讨论 $n+1$ 维向量就更没有意义了，因为维度根本就不够用。

有了这个直观的几何意义，很多问题看起来几乎都是显然的了，比如那些行列式问题，还有相似矩阵等等，这将在下回谈到。

张量介绍

我们已经大概了解到，数字的有序组合产生了向量，向量的有序组合产生了矩阵。这样两个新构造出来的对象，作用一个比一个大。那么有人会联想到：矩阵的有序组合，就可以产生一个“立方阵”，它的功能会不会更加强大？更一般的， n 维立方阵呢？这种联想是有道理的，数学上也有这样的研究对象，它就是张量。

最通俗的说法， n 阶张量就是一个 n 维立方阵，所以0阶张量就对应一个数，向量、矩阵分别对应1阶和2阶张量，我们所说的三维立方阵，就是3阶张量啦。当然，张量属于很高深的数学理论，它的性质和作用不可能这么简单就说清楚了。回想当年，爱因斯坦就是用张量分析作为工具，建立起他那伟大的广义相对论的。如果有机会的话，我们一定会重新造访它。

接下来，我们还是回到矩阵问题，谈谈矩阵的行列式。

转载到请包括本文地址：<https://spaces.ac.cn/archives/1768>

4 《新理解矩阵3》：行列式的点滴

Nov By 苏剑林 | 2012-11-04 | 1508位读者

本文的最新版本位于：

<http://kexue.fm/archives/2208/>

亲爱的读者朋友们，科学空间版的理解矩阵已经来到了BoJone认为是最激动人心的部分了，那就是关于行列式的叙述。这部分内容没有在孟岩的文章中被谈及到，是我自己结合了一些书籍和网络资源而得出的一些看法。其中最主要的书籍是《数学桥》，而追本溯源，促进我研究这方面的内容的是matrix67的那篇《教材应该怎么写》。本文包含了相当多的直观理解内容，在我看来，这部分内容也许不是正统的观点，但是至少在某种程度上能够促进我们对线性代数的理解。

大多数线性代数引入行列式的方式都是通过讲解线性方程组的，这种方式能够让学生很快地掌握它的计算，以及给出了一个最实际的应用（就是解方程组啦）。但是这很容易让读者走进一个误区，让他们认为线性代数就是研究解方程组的。这样并不能让读者真正理解到它的本质，而只有当我们对它有了一个直观熟练的感觉，我们才能很好地运用它。

行列式的出现其实是为了判断一个矩阵是否可逆的，它通过某些方式构造出一个“相对简单”的函数来达到这个目的，这个函数就是矩阵的行列式。让我们来反思一下，矩阵可逆意味着什么呢？之前已经提到过，矩阵是从一个点到另外一个点的变换，那么逆矩阵很显然就是为了把它变换回来。我们还说过，“运动是相对的”，点的变换又可以用坐标系的变换来实现。但是，按照我们的直觉，不同的坐标系除了有那些运算上的复杂度不同（比如一般的仿射坐标系计算点积比直角坐标系复杂）之外，不应该有其他的不同了，用物理的语言说，就是一切坐标系都是平权的。那么给出一个坐标系，可以自然地变换到另外一个坐标系，也可以自然地将它变换回来。既然矩阵是这种坐标系的一个描述，那么矩阵不可逆的唯一可能性就是：

这个n阶矩阵的n个列向量根本就构不成一个n维空间的坐标系。

也就是说，它没有资格成为一个坐标系！

（用专业术语说，这n个向量是线性相关的。）

那么我们就想办法为某个n阶矩阵A构造一个函数 $|A|$ ，当 $|A|$ 为0时，就意味着这个矩阵没有资格成为一个坐标系，也就是说它不可逆。

先设 $A = [\vec{c}_1, \vec{c}_2, \dots, \vec{c}_n]$ ，其中 \vec{v}_i 是各个列向量。我们想想，哪些情况下矩阵A显然没有资格成为一个坐标系？

- 1、显然，能否构成一个坐标系，与每一个向量 \vec{v}_i 都息息相关，那么 $|A|$ 应该是每一个向量 \vec{v}_i 的函数才能满足要求。
- 2、简单起见，我们函数最好是线性的，也就是说 $A = [\vec{c}_1, \dots, \vec{v}_i, \dots, \vec{c}_n]$ $A' = [\vec{c}_1, \dots, \alpha \vec{v}_i, \dots, \vec{c}_n]$ 那么 $|A'| = \alpha |A|$
- 3、构成n维坐标系需要n个独立的向量，如果矩阵A的n个向量中有两个相同的，那么它自然无法构成坐标系，也就是说此时 $|A| = 0$ ；
- 4、为了确定适当的缩放因子，也为了满足我们的直观感受，我们规定 $|I| = 1$ 。

让人意外的是，就凭以上三个性质，就可以确定唯一一个函数 $|A|$ 了，那个函数就是我们课本给出的行列式计算公式！当然，推导过程有点复杂，在这里不给出了，有兴趣的朋友可以用二阶和三阶矩阵演算一下，然后类比就行了。

让我们再回想一下课本是怎么介绍行列式的：首先是线性方程组给出行列式的计算公式，然后再证明行列式有各种各样的性质，比如“把其中一行乘上5，行列式的值也乘上5”等等。这与上面的说法（这其实也是标准的线性代数讲述方式）相比，未免有点本末倒置的味道。因为行列式本身就是由那些性质来定义的，我们又何必要用行列式的计算公式来反推那些什么性质呢？除了锻炼计算能力，我真的看不出有什么作用。

上面的内容还有一个非常直观的几何理解方式：

三维空间中三个向量能够构成一个仿射坐标系的充要条件是什么？很简单，这三个向量必须不共面！因此，推广到 n 维空间，很显然有：

n 个向量可以构成一个坐标系就意味着这些向量不共面！

当然， n 维空间中的面不是指二维平面，而是 $n-1$ 维的东西，它与 n 维体积相对。

还可以换一种说法

n 个向量可以构成一个坐标系就意味着这些向量构成的一个 n 维平行立体（联想我们的平行六面体）体积不为0！

注意注意，这里涉及到了体积，为0与否关系到矩阵可逆性，这与行列式同出一辙。这个 n 维体积是否和行列式有关呢？我们不妨分析一下矩阵 A 所构成的 n 维立体的体积 V 一些性质：

- 1、很明显，每一个向量 \vec{v}_i 也是每一条边长，那么 V 应该是每一个向量 \vec{v}_i 的函数；
- 2、当某个向量变为原来的 α 倍，那么体积也应该变成原来的 $|\alpha|$ 倍，加负号是因为向量是可以反向延伸的，但体积一定是正数；
- 3、有两个向量相同时，意味着这些向量“共面”，这时体积自然为0；
- 4、单位的 n 维立方体体积自然是1。

出人意料，这些性质几乎与行列式一一对应，除了第2点那个 α 要加绝对值之外。那么，这些性质最终迫使这个 n 维立体的体积为 $V = \text{ABS}(|A|)$ ，为了不引起混乱，我用了ABS表示绝对值。

这真的是一个非常棒的事实，行列式有如此明显的几何意义，它可以用来求体积（二阶矩阵的时候就是平行四边形的面积）！这就不难解释我们一开始在解析几何课程上学到的平行六面体体积公式为什么是这样子的了。它完全就可以撇开什么点乘、叉乘而独立得到！还有很多例子，读者在学习到相关的知识时必然会感受到。但另一方面，这也反映了我们教材的糟糕：要是用我们讲述行列式的方式，我们要花多大功夫才能证明行列式就是体积？

浅陋观点，望各位不吝指出错误。

转载到请包括本文地址：<https://spaces.ac.cn/archives/1770>

11 《新理解矩阵4》：相似矩阵的那些事儿

Nov By 苏剑林 | 2012-11-11 | 3192位读者

这篇文章估计是这个系列最后一篇了，也许以后会继续谈到线性代数，但是将会独立开来讲述。本文主要讲的是相似矩阵的一些事情，本文的观点很是粗糙，自己感觉都有点模糊，因此请读者细细阅读。在孟岩的文章里头，它对矩阵及其相似有了一个非常精彩的描述：

“矩阵是线性空间中的线性变换的一个描述。在一个线性空间中，只要我们选定一组基，那么对于任何一个线性变换，都能够用一个确定的矩阵来加以描述。”

同样的，对于一个线性变换，只要你选定一组基，那么就可以找到一个矩阵来描述这个线性变换。换一组基，就得到一个不同的矩阵。所有这些矩阵都是这同一个线性变换的描述，但又都不是线性变换本身。

上述所有这些同一个线性变换的描述的矩阵互为相似矩阵。孟岩还提到那个相似矩阵的公式可以用一种非常直观的方式来证明，可是就没有后文了。我没有跟他联系过，但是我一直在寻求这方面的直观理解。在翻阅了许多书籍之后，终于有了一个自己比较满意的答案。也许读者会感到意外的是，促使我得到这个理解的，不是数学著作，而是一本偏向物理的数《群论与量子力学的对称性》。

Part 1

首先来一个比较物理的理解：矩阵A描述了向量x到向量y的一个运动，即 $y = Ax$ ；但是，这仅仅是在直角坐标系下测量的，在一个新的坐标系P之下，假设测量结果为 $y' = Bx'$ 。

根据我们在前边给出的矩阵几何理解，在P坐标系下测量的 x' ，在直角坐标系测量为 x ，可以表示成 $Px' = x$ ；同理有 $Py' = y$ 。代入就得到： $Py' = APx'$ ，可以稍稍改成 $Py' = P(P^{-1}AP)x'$ ，换句话说，在P坐标系下，从 x' 到 y' 的运动用矩阵 $B = P^{-1}AP$ 表示，这就是A的一个相似矩阵！所以说，一族相似矩阵，只不过是同一个线性变换在不同坐标系下的一个测量结果而已。

Part 2

其实，相似矩阵还有一个相对直观的几何立体模型。我们知道一个矩阵A由n个列向量组成，它实际上给出了n维空间的一个n维平行方体（类比二维的平行四边形和三维的平行六面体）。而矩阵I实际上给出了一个n维单位方体。假设他们两个存在某种对应关系。

而矩阵A在新坐标系P下的测量结果为 $P^{-1}A$ ，即 $A = P(P^{-1}A)$ ；而I在P的测量结果为 $I = P(P^{-1})$ ，也就是说，在新坐标系下， P^{-1} 与 $P^{-1}A$ 具有对应关系。那么新坐标系下的单位方体对应什么呢？那就是

$$\begin{aligned} P^{-1} &\rightarrow P^{-1}P = I \\ P^{-1}A &\rightarrow P^{-1}AP \end{aligned}$$

也就是说新坐标系下的单位方体对应着相似矩阵所描述的n维方体！

这压根儿就是配对原则嘛！

这就不难理解为什么相似矩阵的行列式值都相同了。行列式的几何意义就是体积，虽然矩阵A代表的立方体经过坐标变换后体积变了，但是单位方体的体积实则也变啦，也就是说，新坐标系下一切标度都变化了，但是从“数格子”的角度来说，格子数目是没有变化的，所以体积也就没有变化了。

伟大的矩阵

在物理学，几乎每一个领域都广泛地用到了矩阵，但是，与矩阵联系最紧密的学科当数量子力学。很多人都知道，量子力学有三种等价表达形式，一种是薛定谔的波动方程（就是我现在学习的），一种是海森堡的矩阵力学，最后一种是天才的费曼的路径积分。话说当年海森堡在构思量子力学时，线性代数这门课程已经发展得很丰富了，但他自己并没有学习到。不过他自己却“发明”了一个自称为“能量表格”的东西，用来作为描述他构思的工具。最后当他把论文提交给导师玻恩时，玻恩毫不客气地跟他说：“你这个新的能量表格，就是数学家早已研究过的矩阵。”呵呵，让人惊讶，矩阵力学的创始人居然不知矩阵为何物。后来海森堡补习了矩阵的知识，并和导师合作发表了矩阵力学的成果。

最近我看量子力学和狭义相对论的内容，发现两者的描述方式其实在很大程度上已经得到了统一，大家都是先讲一下基础知识，然后讲一下线性代数、群论等知识。最后都基本上归结为用矩阵和群论知识来分析了。我想这也是为了物理学统一描述的需要吧。让我觉得一点意外的是，这种综合的抽象模式，反倒让我感觉容易上手了。也许正是因为我是个数学爱好者吧。

最后总结一下我的这几篇《新理解矩阵》

这几篇文章很粗糙、放肆，很不成熟，甚至某些观点不一定正确，因为直观理解会给人一种以偏概全的感觉，忽略掉了抽象的巨大作用。但是我想只有在有了直观认识之后，才可以更熟练地运用它；更加全面的认识，也在这种直观的效果下慢慢感悟，慢慢积累起来的，我想数学史上线性代数知识的发展历程也是相似的，既然如此，我们为什么不按照历史的发展方式来学习它呢？

转载到请包括本文地址：<https://spaces.ac.cn/archives/1777>

25 《新理解矩阵5》：体积=行列式

Dec By 苏剑林 | 2013-12-25 | 1902位读者

在文章《新理解矩阵3》：行列式的点滴中，笔者首次谈及到了行列式的几何意义，它代表了 n 维的“平行多面体”的“体积”。然而，这篇文章写于我初学矩阵之时，有些论述并不严谨，甚至有些错误。最近笔者在写期末论文的时候，研究了超复数的相关内容，而行列式的几何意义在我的超复数研究中具有重要作用，因此把行列式的几何意义重新研究了一翻，修正了部分错误，故发此文，与大家分享。

一个 n 阶矩阵 A 可以看成是 n 个 n 维列向量 x_1, x_2, \dots, x_n 的集合

$$A = (x_1, x_2, \dots, x_n)$$

从代数的角度来看，这构成了一个矩阵；从几何的角度来看，这 n 个向量可以建立一个平行 n 维体。比如：平行四边形就是“平行二维体”，平行六面体就是“平行三维体”，高阶的只需要相应类比，不需要真正想象出高维空间的立体是什么样。

让我们考虑矩阵 A 的行列式 $\det A$ ，我们知道 $\det A$ 有如下性质：

行列式性质

- 1、行列式是 x_1, x_2, \dots, x_n 的一个函数，即 $\det A = f(x_1, x_2, \dots, x_n)$ ；
- 2、（线性1）行列式的某一列乘上常数 α ，则行列式的值也乘上 α ，即 $f(x_1, \dots, \alpha x_i, \dots, x_n) = \alpha f(x_1, \dots, x_i, \dots, x_n)$ ；
- 3、（线性2）将行列式的某一列写成两列之和，那么行列式也相应地成为两个行列式之和，即 $f(x_1, \dots, x_i, \dots, x_n) = f(x_1, \dots, y_i, \dots, x_n) + f(x_1, \dots, z_i, \dots, x_n)$ 其中 $x_i = y_i + z_i$ ，性质二和三表明 f 是关于每个向量的线性函数；
- 4、（反对称）只要有两列相同，那么行列式值为0，即 $f(\dots, x, \dots, x, \dots) = 0$ ；
- 5、（归一）单位矩阵的行列式为1，即 $f(I) = 1$ 。

一个惊人的事实是，行列式可以由上面五条性质唯一确定！即由上面五条性质就可以唯一确定一个函数 f ，这个函数就是矩阵的行列式。

从几何的角度来看，用这 n 个向量，可以生成 n 维空间的一个平行 n 维体。让我们来考虑这个平行 n 维体的体积 V 。只在第一卦限讨论，那么体积具有下面的性质（只在第一卦限讨论，限保证了所有的向量和因子都是正数。）

体积性质

- 1、体积是这 n 个向量的一个函数 $V(x_1, x_2, \dots, x_n)$ ；
- 2、将某个向量乘以 α ，也就是把它的长度变为原来的 α 倍，那么体积也增大 α 倍，即 $V(x_1, \dots, \alpha x_i, \dots, x_n) = \alpha V(x_1, \dots, x_i, \dots, x_n)$ ；

- 3、体积是可加的，即 $V(x_1, \dots, x_i, \dots, x_n) = V(x_1, \dots, y_i, \dots, x_n) + V(x_1, \dots, z_i, \dots, x_n)$ 其中 $x_i = y_i + z_i$ ；这点需要稍加验证，但它的确是正确的。
- 4、只要有二个向量重合，那么体积自然为0，即 $V(\dots, x, \dots, x, \dots) = 0$ ；比如在三维空间中的一个立体，有两条边重合，那么说明这个立体已经压缩为一个面了，面的体积自然为0。
- 5、由单位矩阵 I 构成的平行 n 维体是一个 n 维的单位立方体，它的体积自然是1，即 $V(I) = 1$ 。

比较行列式和体积的性质，可以发现它们是完全相同的，所以在第一卦限中的平行 n 维体的体积就是对应矩阵的行列式！如果将其放到所有卦限中，那只不过是体积概念的推广（允许为负数）。因此，我们不妨这样定义：**体积就是行列式**。

事实上，负体积的引入具有重要意义，它是现在的“外微分”的基础之一。外微分一个典型的用处是它可以把高斯积分公式、斯托克斯积分公式等统一起来。它使微分的理论和形式更完整统一。

转载到请包括本文地址：<https://spaces.ac.cn/archives/2208>

15 《新理解矩阵6》：为什么只有方阵有行列式？

Jul By 苏剑林 | 2014-07-15 | 2631位读者

学过线性代数的朋友都知道，方阵和非方阵的一个明显不同是，对于方阵我们可以计算它的行列式，如果不是方阵的话，就没有行列式这个概念了。在追求统一和谐的数学系统中，为什么非方阵却没有行列式？也许对于这个问题最恰当的回答是——因为不够美。对于非方阵，其实也可以类似地定义它的行列式，定义出来的东西，跟方阵的行列式具有同样的性质，比如某行乘上一个常数，行列式值也就乘以一个常数，等等；而且还可以把其几何意义保留下来。但是，非方阵的行列式是不够美的，因为对于一个一般的整数元素的方阵，我们的行列式是一个整数；而对于一个一般的整数元素的非方阵，却导致了一个无理数的行列式值。另外，一个也比较重要的原因是，单单是方阵的行列式也够用了。综合以上两个理由，非方阵的行列式就被舍弃不用了。

非方阵的行列式不够漂亮

n 阶方阵的行列式是每个向量的线性函数，它代表着向量之间的线性相关性；从几何上来讲，它就是向量组成的平行 n 维体的（有向）体积。我们当然期望非方阵的行列式也保留这些性质，因为只有这样，方阵行列式的那些运算性质才得以保留，比如上面说的，行列式的一行乘上一个常数，行列式值也乘上一个常数。我们考虑 $m \times n$ 的矩阵，其中 $m < n$ ，我们将它看成是 m 个 n 维向量的组合。最简单的，我们先考虑 1×2 矩阵的行列式，也就是二维向量 (a, b) 的行列式。

我们已经知道， 2×2 矩阵的行列式的绝对值，就是这两个向量所围成的平行四边形的面积。类似地，二维向量 (a, b) 的行列式的绝对值，就等于该向量自身的长度了，也就是 $\det(a, b) = \sqrt{a^2 + b^2}$ ，对于有理的 a, b ，大多数情况都会出现无理的行列式值，这是不协调的。因为它是线性的函数，有理数的线性运算导致无理数，却是不大舒服的。

类似的，可以考虑 2×3 矩阵

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

的行列式。按照几何意义，它的行列式的模就是这两个向量所围成的平行四边形的面积。我们可以利用叉积来计算这个面积，但为了更一般化，我们利用正交化的方法来计算它。

以向量 $\mathbf{x}_1 = (a, b, c)$ 为出发点，选取

$$\mathbf{e}_1 = \frac{(a, b, c)}{\sqrt{a^2 + b^2 + c^2}}$$

那么 $\mathbf{x}_1 = |\mathbf{x}_1| \mathbf{e}_1$ 。对 $\mathbf{x}_2 = (d, e, f)$ 正交化，得到

$$\mathbf{x}_2 - \langle \mathbf{e}_1, \mathbf{x}_2 \rangle \mathbf{e}_1$$

取

$$\mathbf{e}_2 = \frac{\mathbf{x}_2 - \langle \mathbf{e}_1, \mathbf{x}_2 \rangle \mathbf{e}_1}{|\mathbf{x}_2 - \langle \mathbf{e}_1, \mathbf{x}_2 \rangle \mathbf{e}_1|}$$

得到

$$x_2 = |x_2 - \langle e_1, x_2 \rangle e_1| e_2 + \langle e_1, x_2 \rangle e_1$$

因此，行列式的模，也就是两个向量所形成的平行四边形的面积等于

$$|\det A| = |x_1| \cdot |x_2 - \langle e_1, x_2 \rangle e_1|$$

最后的表达式出现了两个模，表明两个根号，随便挑个具体例子就可以验证，即使 A 的元素全部是整数，也得不到有理数。这体现了非方阵行列式的不美之处。

方阵的行列式就够了

由于非方阵的行列式不够美，那么我们干脆弃之不用了。可是，这样会不会产生什么“副作用”？也就是说，会不会有哪些地方，非用到非方阵的行列式不可？事实上，至少就我目前的认识来说，答案是没有。

比如说判断 m 个 n 维行向量的线性相关性，我们是这样做的：第一种方法，利用初等变换，看变换后的矩阵的秩是否为 m ；如果嫌第一种方法步骤太多，第二种方法相对来说干脆一点，就是检验删除任意 $n - m$ 列后，剩下的方阵行列式是否为0，如果存在一个不为0，就说明线性无关了。所以说，方阵的行列式够用了。

如果要求面积、体积怎么办？事实上，用的就是类似 $|x_1| \cdot |x_2 - \langle e_1, x_2 \rangle e_1|$ 的公式，这公式不难理解，也不难记忆。由于它带有不可化简的根号，因此，也就没有简单的计算公式了。

最后，一个比较核心的限制了非方阵行列式使用的原因是，非方阵的行列式应用不多。方阵的行列式可以用来求各种因子，比如重积分的坐标变换中的雅可比行列式，等等，这些由于方阵的可逆性，这些行列式都是有直接的应用意义的。相比之下，非方阵没有可逆的说法，所以非方阵的行列式应用也就很窄了。

转载到请包括本文地址：<https://spaces.ac.cn/archives/2757>