**矩阵的内涵**

如果不熟悉线性代数的概念，要去学习自然科学，现在看来就和文盲差不多。然而“按照现行的国际标准，线性代数是通过公理化来表述的，它是第二代数学模型，这就带来了教学上的困难。

\* 矩阵究竟是什么东西？向量可以被认为是具有n个相互独立的性质（维度）的对象的表示，矩阵又是什么呢？我们如果认为矩阵是一组列（行）向量组成的新的复合向量的展开式，那么为什么这种展开式具有如此广泛的应用？特别是，为什么偏偏二维的展开式如此有用？如果矩阵中每一个元素又是一个向量，那么我们再展开一次，变成三维的立方阵，是不是更有用？

\* 矩阵的乘法规则究竟为什么这样规定？为什么这样一种怪异的乘法规则却能够在实践中发挥如此巨大的功效？很多看上去似乎是完全不相关的问题，最后竟然都归结到矩阵的乘法，这难道不是很奇妙的事情？难道在矩阵乘法那看上去莫名其妙的规则下面，包含着世界的某些本质规律？如果是的话，这些本质规律是什么？

\* 行列式究竟是一个什么东西？为什么会有如此怪异的计算规则？行列式与其对应方阵本质上是什么关系？为什么只有方阵才有对应的行列式，而一般矩阵就没有（不要觉得这个问题很蠢，如果必要，针对m x n矩阵定义行列式不是做不到的，之所以不做，是因为没有这个必要，但是为什么没有这个必要）？而且，行列式的计算规则，看上去跟矩阵的任何计算规则都没有直观的联系，为什么又在很多方面决定了矩阵的性质？难道这一切仅是巧合？

\* 矩阵为什么可以分块计算？分块计算这件事情看上去是那么随意，为什么竟是可行的？

\* 对于矩阵转置运算AT，有(AB)T = BTAT，对于矩阵求逆运算A-1，有(AB)-1 = B-1A-1。两个看上去完全没有什么关系的运算，为什么有着类似的性质？这仅仅是巧合吗？

\* 为什么说P-1AP得到的矩阵与A矩阵“相似”？这里的“相似”是什么意思？

\* 特征值和特征向量的本质是什么？它们定义就让人很惊讶，因为Ax =λx，一个诺大的矩阵的效应，竟然不过相当于一个小小的数λ，确实有点奇妙。但何至于用“特征”甚至“本征”来界定？它们刻划的究竟是什么？

今天先谈谈对线形空间和矩阵的几个核心概念的理解。首先说说空间(space)，这个概念是现代数学的命根子之一，从拓扑空间开始，一步步往上加定义，可以形成很多空间。线形空间其实还是比较初级的，如果在里面定义了范数，就成了赋范线性空间。赋范线性空间满足完备性，就成了巴那赫空间；赋范线性空间中定义角度，就有了内积空间，内积空间再满足完备性，就得到希尔伯特空间。

总之，空间有很多种。你要是去看某种空间的数学定义，大致都是“存在一个集合，在这个集合上定义某某概念，然后满足某些性质”，就可以被称为空间。这未免有点奇怪，为什么要用“空间”来称呼一些这样的集合呢？大家将会看到，其实这是很有道理的。

我们一般人最熟悉的空间，毫无疑问就是我们生活在其中的（按照牛顿的绝对时空观）的三维空间，从数学上说，这是一个三维的欧几里德空间，我们先不管那么多，先看看我们熟悉的这样一个空间有些什么最基本的特点。仔细想想我们就会知道，这个三维的空间：1. 由很多（实际上是无穷多个）位置点组成；2. 这些点之间存在相对的关系；3. 可以在空间中定义长度、角度；4. 这个空间可以容纳运动，这里我们所说的运动是从一个点到另一个点的移动（变换），而不是微积分意义上的“连续”性的运动，

事实上，不管是什么空间，都必须容纳和支持在其中发生的符合规则的运动（变换）。你会发现，在某种空间中往往会存在一种相对应的变换，比如拓扑空间中有拓扑变换，线性空间中有线性变换，仿射空间中有仿射变换，其实这些变换都只不过是对应空间中允许的运动形式而已。

因此只要知道，“空间”是容纳运动的一个对象集合，而变换则规定了对应空间的运动。

下面我们来看看线性空间。线性空间的定义任何一本书上都有，但是既然我们承认线性空间是个空间，那么有两个最基本的问题必须首先得到解决，那就是：

1. 空间是一个对象集合，线性空间也是空间，所以也是一个对象集合。那么线性空间是什么样的对象的集合？或者说，线性空间中的对象有什么共同点吗？

2. 线性空间中的运动如何表述的？也就是，线性变换是如何表示的？

我们先来回答第一个问题，回答这个问题的时候其实是不用拐弯抹角的，可以直截了当的给出答案。线性空间中的任何一个对象，通过选取基和坐标的办法，都可以表达为向量的形式。通常的向量空间我就不说了，举两个不那么平凡的例子：

L1. 最高次项不大于n次的多项式的全体构成一个线性空间，也就是说，这个线性空间中的每一个对象是一个多项式。如果我们以x0, x1, ..., xn为基，那么任何一个这样的多项式都可以表达为一组n+1维向量，其中的每一个分量ai其实就是多项式中x(i-1)项的系数。值得说明的是，基的选取有多种办法，只要所选取的那一组基线性无关就可以。这要用到后面提到的概念了，所以这里先不说，提一下而已。

L2. 闭区间[a, b]上的n阶连续可微函数的全体，构成一个线性空间。也就是说，这个线性空间的每一个对象是一个连续函数。对于其中任何一个连续函数，根据魏尔斯特拉斯定理，一定可以找到最高次项不大于n的多项式函数，使之与该连续函数的差为0，也就是说，完全相等。这样就把问题归结为L1了。后面就不用再重复了。

所以说，向量是很厉害的，只要你找到合适的基，用向量可以表示线性空间里任何一个对象。这里头大有文章，因为向量表面上只是一列数，但是其实由于它的有序性，所以除了这些数本身携带的信息之外，还可以在每个数的对应位置上携带信息。为什么在程序设计中数组最简单，却又威力无穷呢？根本原因就在于此。这是另一个问题了，这里就不说了。

下面来回答第二个问题，这个问题的回答会涉及到线性代数的一个最根本的问题。

线性空间中的运动，被称为线性变换。也就是说，你从线性空间中的一个点运动到任意的另外一个点，都可以通过一个线性变化来完成。那么，线性变换如何表示呢？很有意思，在线性空间中，当你选定一组基之后，不仅可以用一个向量来描述空间中的任何一个对象，而且可以用矩阵来描述该空间中的任何一个运动（变换）。而使某个对象发生对应运动的方法，就是用代表那个运动的矩阵，乘以代表那个对象的向量。

简而言之，在线性空间中选定基之后，向量刻画对象，矩阵刻画对象的运动，用矩阵与向量的乘法施加运动。

是的，矩阵的本质是运动的描述。如果以后有人问你矩阵是什么，那么你就可以响亮地告诉他，矩阵的本质是运动的描述

可是多么有意思啊，向量本身不是也可以看成是n x 1矩阵吗？这实在是很奇妙，一个空间中的对象和运动竟然可以用相类同的方式表示。能说这是巧合吗？如果是巧合的话，那可真是幸运的巧合！可以说，线性代数中大多数奇妙的性质，均与这个巧合有直接的关系。

“矩阵是运动的描述”，到现在为止，好像大家都还没什么意见。但是我相信早晚会有数学系出身的网友来拍板转。因为运动这个概念，在数学和物理里是跟微积分联系在一起的。我们学习微积分的时候，总会有人照本宣科地告诉你，初等数学是研究常量的数学，是研究静态的数学，高等数学是变量的数学，是研究运动的数学。大家口口相传，差不多人人都知道这句话。但是真知道这句话说的是什么意思的人，好像也不多。简而言之，在我们人类的经验里，运动是一个连续过程，从A点到B点，就算走得最快的光，也是需要一个时间来逐点地经过AB之间的路径，这就带来了连续性的概念。而连续这个事情，如果不定义极限的概念，根本就解释不了。古希腊人的数学非常强，但就是缺乏极限观念，所以解释不了运动，被芝诺的那些著名悖论（飞箭不动、飞毛腿阿喀琉斯跑不过乌龟等四个悖论）搞得死去活来。因为这篇文章不是讲微积分的，所以我就不多说了。有兴趣的读者可以去看看齐民友教授写的《重温微积分》。我就是读了这本书开头的部分，才明白“高等数学是研究运动的数学”这句话的道理。

不过在我这个《理解矩阵》的文章里，“运动”的概念不是微积分中的连续性的运动，而是瞬间发生的变化。比如这个时刻在A点，经过一个“运动”，一下子就“跃迁”到了B点，其中不需要经过A点与B点之间的任何一个点。这样的“运动”，或者说“跃迁”，是违反我们日常的经验的。不过了解一点量子物理常识的人，就会立刻指出，量子（例如电子）在不同的能量级轨道上跳跃，就是瞬间发生的，具有这样一种跃迁行为。所以说，自然界中并不是没有这种运动现象，只不过宏观上我们观察不到。但是不管怎么说，“运动”这个词用在这里，还是容易产生歧义的，说得更确切些，应该是“跃迁”。因此这句话可以改成：“矩阵是线性空间里跃迁的描述”。

可是这样说又太物理，也就是说太具体，而不够数学，也就是说不够抽象。因此我们最后换用一个正牌的数学术语——变换，来描述这个事情。这样一说，大家就应该明白了，所谓变换，其实就是空间里从一个点（元素/对象）到另一个点（元素/对象）的跃迁。比如说，拓扑变换，就是在拓扑空间里从一个点到另一个点的跃迁。再比如说，仿射变换，就是在仿射空间里从一个点到另一个点的跃迁。附带说一下，这个仿射空间跟向量空间是亲兄弟。做计算机图形学的朋友都知道，尽管描述一个三维对象只需要三维向量，但所有的计算机图形学变换矩阵都是4 x 4的。说其原因，很多书上都写着“为了使用中方便”，这在我看来简直就是企图蒙混过关。真正的原因，是因为在计算机图形学里应用的图形变换，实际上是在仿射空间而不是向量空间中进行的。想想看，在向量空间里相一个向量平行移动以后仍是相同的那个向量，而现实世界等长的两个平行线段当然不能被认为同一个东西，所以计算机图形学的生存空间实际上是仿射空间。而仿射变换的矩阵表示根本就是4 x 4的。又扯远了，有兴趣的读者可以去看《计算机图形学——几何工具算法详解》。

一旦我们理解了“变换”这个概念，矩阵的定义就变成：

“矩阵是线性空间里的变换的描述。”

到这里为止，我们终于得到了一个看上去比较数学的定义。不过还要多说几句。教材上一般是这么说的，在一个线性空间V里的一个线性变换T，当选定一组基之后，就可以表示为矩阵。因此我们还要说清楚到底什么是线性变换，什么是基，什么叫选定一组基。线性变换的定义是很简单的，设有一种变换T，使得对于线性空间V中间任何两个不相同的对象x和y，以及任意实数a和b，有：

T(ax + by) = aT(x) + bT(y)，

那么就称T为线性变换。

定义都是这么写的，但是光看定义还得不到直觉的理解。线性变换究竟是一种什么样的变换？我们刚才说了，变换是从空间的一个点跃迁到另一个点，而线性变换，就是从一个线性空间V的某一个点跃迁到另一个线性空间W的另一个点的运动。这句话里蕴含着一层意思，就是说一个点不仅可以变换到同一个线性空间中的另一个点，而且可以变换到另一个线性空间中的另一个点去。不管你怎么变，只要变换前后都是线性空间中的对象，这个变换就一定是线性变换，也就一定可以用一个非奇异矩阵来描述。而你用一个非奇异矩阵去描述的一个变换，一定是一个线性变换。有的人可能要问，这里为什么要强调非奇异矩阵？所谓非奇异，只对方阵有意义，那么非方阵的情况怎么样？这个说起来就会比较冗长了，最后要把线性变换作为一种映射，并且讨论其映射性质，以及线性变换的核与像等概念才能彻底讲清楚。我觉得这个不算是重点，如果确实有时间的话，以后写一点。以下我们只探讨最常用、最有用的一种变换，就是在同一个线性空间之内的线性变换。也就是说，下面所说的矩阵，不作说明的话，就是方阵，而且是非奇异方阵。学习一门学问，最重要的是把握主干内容，迅速建立对于这门学问的整体概念，不必一开始就考虑所有的细枝末节和特殊情况，自乱阵脚。

接着往下说，什么是基呢？这个问题在后面还要大讲一番，这里只要把基看成是线性空间里的坐标系就可以了。注意是坐标系，不是坐标值，这两者可是一个“对立矛盾统一体”。这样一来，“选定一组基”就是说在线性空间里选定一个坐标系。就这意思。

好，最后我们把矩阵的定义完善如下：

“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。”

理解这句话的关键，在于把“线性变换”与“线性变换的一个描述”区别开。一个是那个对象，一个是对那个对象的表述。就好像我们熟悉的面向对象编程中，一个对象可以有多个引用，每个引用可以叫不同的名字，但都是指的同一个对象。如果还不形象，那就干脆来个很俗的类比。

比如有一头猪，你打算给它拍照片，只要你给照相机选定了一个镜头位置，那么就可以给这头猪拍一张照片。这个照片可以看成是这头猪的一个描述，但只是一个片面的的描述，因为换一个镜头位置给这头猪拍照，能得到一张不同的照片，也是这头猪的另一个片面的描述。所有这样照出来的照片都是这同一头猪的描述，但是又都不是这头猪本身。

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。

但是这样的话，问题就来了如果你给我两张猪的照片，我怎么知道这两张照片上的是同一头猪呢？同样的，你给我两个矩阵，我怎么知道这两个矩阵是描述的同一个线性变换呢？如果是同一个线性变换的不同的矩阵描述，那就是本家兄弟了，见面不认识，岂不成了笑话。

好在，我们可以找到同一个线性变换的矩阵兄弟们的一个性质，那就是：

若矩阵A与B是同一个线性变换的两个不同的描述（之所以会不同，是因为选定了不同的基，也就是选定了不同的坐标系），则一定能找到一个非奇异矩阵P，使得A、B之间满足这样的关系：

A = P-1BP

线性代数稍微熟一点的读者一下就看出来，这就是相似矩阵的定义。没错，所谓相似矩阵，就是同一个线性变换的不同的描述矩阵。按照这个定义，同一头猪的不同角度的照片也可以成为相似照片。俗了一点，不过能让人明白。

而在上面式子里那个矩阵P，其实就是A矩阵所基于的基与B矩阵所基于的基这两组基之间的一个变换关系。关于这个结论，可以用一种非常直觉的方法来证明（而不是一般教科书上那种形式上的证明），如果有时间的话，我以后在blog里补充这个证明。

这个发现太重要了。原来一族相似矩阵都是同一个线性变换的描述啊！难怪这么重要！工科研究生课程中有矩阵论、矩阵分析等课程，其中讲了各种各样的相似变换，比如什么相似标准型，对角化之类的内容，都要求变换以后得到的那个矩阵与先前的那个矩阵式相似的，为什么这么要求？因为只有这样要求，才能保证变换前后的两个矩阵是描述同一个线性变换的。当然，同一个线性变换的不同矩阵描述，从实际运算性质来看并不是不分好环的。有些描述矩阵就比其他的矩阵性质好得多。这很容易理解，同一头猪的照片也有美丑之分嘛。所以矩阵的相似变换可以把一个比较丑的矩阵变成一个比较美的矩阵，而保证这两个矩阵都是描述了同一个线性变换。

这样一来，矩阵作为线性变换描述的一面，基本上说清楚了。但是，事情没有那么简单，或者说，线性代数还有比这更奇妙的性质，那就是，矩阵不仅可以作为线性变换的描述，而且可以作为一组基的描述。而作为变换的矩阵，不但可以把线性空间中的一个点给变换到另一个点去，而且也能够把线性空间中的一个坐标系（基）表换到另一个坐标系（基）去。而且，变换点与变换坐标系，具有异曲同工的效果。线性代数里最有趣的奥妙，就蕴含在其中。理解了这些内容，线性代数里很多定理和规则会变得更加清晰、直觉。

首先来总结一下前面两部分的一些主要结论：

1. 首先有空间，空间可以容纳对象运动的。一种空间对应一类对象。

2. 有一种空间叫线性空间，线性空间是容纳向量对象运动的。

3. 运动是瞬时的，因此也被称为变换。

4. 矩阵是线性空间中运动（变换）的描述。

5. 矩阵与向量相乘，就是实施运动（变换）的过程。

6. 同一个变换，在不同的坐标系下表现为不同的矩阵，但是它们的本质是一样的，所以本征值相同。

下面让我们把视力集中到一点以改变我们以往看待矩阵的方式。我们知道，线性空间里的基本对象是向量，而向量是这么表示的：

 [a1, a2, a3, ..., an]  
 矩阵呢？矩阵是这么表示的：

 a11, a12, a13, ..., a1n

a21, a22, a23, ..., a2n

...

an1, an2, an3, ..., ann

不用太聪明，我们就能看出来，矩阵是一组向量组成的。特别的，n维线性空间里的方阵是由n个n维向量组成的。我们在这里只讨论这个n阶的、非奇异的方阵，如果一组向量是彼此线性无关的话，那么它们就可以成为度量这个线性空间的一组基，从而事实上成为一个坐标系体系，其中每一个向量都躺在一根坐标轴上，并且成为那根坐标轴上的基本度量单位（长度1）。现在到了关键的一步。看上去矩阵就是由一组向量组成的，而且如果矩阵非奇异的话（我说了，只考虑这种情况），那么组成这个矩阵的那一组向量也就是线性无关的了，也就可以成为度量线性空间的一个坐标系。结论：矩阵描述了一个坐标系。之所以矩阵又是运动，又是坐标系，那是因为——“运动等价于坐标系变换”。对不起，这话其实不准确，我只是想让你印象深刻。准确的说法是：“对象的变换等价于坐标系的变换”。或者：“固定坐标系下一个对象的变换等价于固定对象所处的坐标系变换。” 说白了就是： “运动是相对的。”

让我们想想，达成同一个变换的结果，比如把点(1, 1)变到点(2, 3)去，你可以有两种做法。第一，坐标系不动，点动，把(1, 1)点挪到(2, 3)去。第二，点不动，变坐标系，让x轴的度量（单位向量）变成原来的1/2，让y轴的度量（单位向量）变成原先的1/3，这样点还是那个点，可是点的坐标就变成(2, 3)了。方式不同，结果一样。从第一个方式来看，那就是我在《理解矩阵》1/2中说的，把矩阵看成是运动描述，矩阵与向量相乘就是使向量（点）运动的过程。在这个方式下，

Ma = b的意思是：

“向量a经过矩阵M所描述的变换，变成了向量b。”而从第二个方式来看，矩阵M描述了一个坐标系，姑且也称之为M。那么：

Ma = b的意思是：

“有一个向量，它在坐标系M的度量下得到的度量结果向量为a，那么它在坐标系I的度量下，这个向量的度量结果是b。”

这里的I是指单位矩阵，就是主对角线是1，其他为零的矩阵。而这两个方式本质上是等价的。我希望你务必理解这一点，因为这是本篇的关键。正因为是关键，所以我得再解释一下。在M为坐标系的意义下，如果把M放在一个向量a的前面，形成Ma的样式，我们可以认为这是对向量a的一个环境声明。它相当于是说： “注意了！这里有一个向量，它在坐标系M中度量，得到的度量结果可以表达为a。可是它在别的坐标系里度量的话，就会得到不同的结果。为了明确，我把M放在前面，让你明白，这是该向量在坐标系M中度量的结果。” 那么我们再看孤零零的向量b：

  b  多看几遍，你没看出来吗？它其实不是b，它是：

 Ib

也就是说：“在单位坐标系，也就是我们通常说的直角坐标系I中，有一个向量，度量的结果是b。”

而 Ma = Ib的意思就是说：

“在M坐标系里量出来的向量a，跟在I坐标系里量出来的向量b，其实根本就是一个向量啊！”这哪里是什么乘法计算，根本就是身份识别嘛。从这个意义上我们重新理解一下向量。向量这个东西客观存在，但是要把它表示出来，就要把它放在一个坐标系中去度量它，然后把度量的结果（向量在各个坐标轴上的投影值）按一定顺序列在一起，就成了我们平时所见的向量表示形式。你选择的坐标系（基）不同，得出来的向量的表示就不同。向量还是那个向量，选择的坐标系不同，其表示方式就不同。因此，按道理来说，每写出一个向量的表示，都应该声明一下这个表示是在哪个坐标系中度量出来的。表示的方式，就是 Ma，也就是说，有一个向量，在M矩阵表示的坐标系中度量出来的结果为a。我们平时说一个向量是[2 3 5 7]T，隐含着是说，这个向量在 I 坐标系中的度量结果是[2 3 5 7]T，因此，这个形式反而是一种简化了的特殊情况。

注意到，M矩阵表示出来的那个坐标系，由一组基组成，而那组基也是由向量组成的，同样存在这组向量是在哪个坐标系下度量而成的问题。也就是说，表述一个矩阵的一般方法，也应该要指明其所处的基准坐标系。所谓M，其实是 IM，也就是说，M中那组基的度量是在 I 坐标系中得出的。从这个视角来看，M×N也不是什么矩阵乘法了，而是声明了一个在M坐标系中量出的另一个坐标系N，其中M本身是在I坐标系中度量出来的。

回过头来说变换的问题。我刚才说，“固定坐标系下一个对象的变换等价于固定对象所处的坐标系变换”，那个“固定对象”我们找到了，就是那个向量。但是坐标系的变换呢？我怎么没看见？

请看：

Ma = Ib

我现在要变M为I，怎么变？对了，再前面乘以个M-1，也就是M的逆矩阵。换句话说，你不是有一个坐标系M吗，现在我让它乘以个M-1，变成I，这样一来的话，原来M坐标系中的a在I中一量，就得到b了。

我建议你此时此刻拿起纸笔，画画图，求得对这件事情的理解。比如，你画一个坐标系，x轴上的衡量单位是2，y轴上的衡量单位是3，在这样一个坐标系里，坐标为(1，1)的那一点，实际上就是笛卡尔坐标系里的点(2, 3)。而让它原形毕露的办法，就是把原来那个坐标系:  
2 0  
0 3  
的x方向度量缩小为原来的1/2，而y方向度量缩小为原来的1/3，这样一来坐标系就变成单位坐标系I了。保持点不变，那个向量现在就变成了(2, 3)了。

怎么能够让“x方向度量缩小为原来的1/2，而y方向度量缩小为原来的1/3”呢？就是让原坐标系：  
2 0  
0 3  
被矩阵：  
1/2 0  
0 1/3  
左乘。而这个矩阵就是原矩阵的逆矩阵。

下面我们得出一个重要的结论：

“对坐标系施加变换的方法，就是让表示那个坐标系的矩阵与表示那个变化的矩阵相乘。”

再一次的，矩阵的乘法变成了运动的施加。只不过，被施加运动的不再是向量，而是另一个坐标系。

如果你觉得你还搞得清楚，请再想一下刚才已经提到的结论，矩阵MxN，一方面表明坐标系N在运动M下的变换结果，另一方面，把M当成N的前缀，当成N的环境描述，那么就是说，在M坐标系度量下，有另一个坐标系N。这个坐标系N如果放在I坐标系中度量，其结果为坐标系MxN。

在这里，我实际上已经回答了一般人在学习线性代数是最困惑的一个问题，那就是为什么矩阵的乘法要规定成这样。简单地说，是因为：

1. 从变换的观点看，对坐标系N施加M变换，就是把组成坐标系N的每一个向量施加M变换。

2. 从坐标系的观点看，在M坐标系中表现为N的另一个坐标系，这也归结为，对N坐标系基的每一个向量，把它在I坐标系中的坐标找出来，然后汇成一个新的矩阵。

3. 至于矩阵乘以向量为什么要那样规定，那是因为一个在M中度量为a的向量，如果想要恢复在I中的真像，就必须分别与M中的每一个向量进行内积运算。我把这个结论的推导留给感兴趣的朋友吧。应该说，其实到了这一步，已经很容易了。

综合以上1/2/3，矩阵的乘法就得那么规定，一切有根有据，绝不是哪个神经病胡思乱想出来的。

我已经无法说得更多了。矩阵又是坐标系，又是变换。到底是坐标系，还是变换，已经说不清楚了，运动与实体在这里统一了，物质与意识的界限已经消失了，一切归于无法言说，无法定义了。道可道，非常道，名可名，非常名。矩阵是在是不可道之道，不可名之名的东西。到了这个时候，我们不得不承认，我们伟大的线性代数课本上说的矩阵定义，是无比正确的：

“矩阵就是由m行n列数放在一起组成的数学对象。”

好了，这基本上就是我想说的全部了。还留下一个行列式的问题。矩阵M的行列式实际上是组成M的各个向量按照平行四边形法则搭成一个n维立方体的体积。对于这一点，我只能感叹于其精妙，却无法揭开其中奥秘了。也许我掌握的数学工具不够，我希望有人能够给我们大家讲解其中的道理了。

我不知道是否讲得足够清楚了，反正这一部分需要您花些功夫去推敲。

此外，请大家不必等待这个系列的后续部分。以我的工作情况而言，近期内很难保证继续投入脑力到这个领域中，尽管我仍然对此兴致浓厚。不过如果还有（四）的话，可能是一些站在应用层面的考虑，比如对计算机图形学相关算法的理解。但是我不承诺这些讨论近期内会出现了。

“分”的反义字是“和”，是我们熟悉的字。比如：2+3=5，从左往右运算，我们叫求和。那么“分”呢，既然是反义字，就把上面的等式反过来：5=2+3。

把一个对象表示成两个以至更多的对象的和，这个过程叫分析。

通常来说，分析对象应当与被分析对象一致。是数就都是数，是函数就都是函数，是向量就都是向量，是矩阵就都是矩阵。

求和是数学里最基本的运算，减、乘、除是从求和中衍生出来的。而更高级的幂、指、对、三角、微积分等，也是一层一层建立起来的， 最根本的还是这个求和。求和最简单，最容易计算，性质也最简单。所以成了分析的基本出发点。

分析的妙处在于，通过分析可以将较复杂的对象划分为较简单的对象。比如2和3就比5简单。单独研究2的性质，再单独研究3的性质，再通过简单的求和，就可以把握5的性质。把复杂的东西划分成若干简单对象的和，对各简单对象搞各个击破，再加起来，复杂的东西也就被掌握了。

分析是西方思想中一个根本性的东西。西方人认为，事物总是有因果的，看到了结果，要分析原因。所谓分析原因，就是找出一堆因素，说明这堆因素合起来导致了结果。西方人认为，事物总是可以分析的。看到了整体，就要把那些合成这个整体的局部一一分析出来。 现代科学很大一部分就是这么回事。

大学数学里，有很多内容就是在讲分析。数学里的分析还要把含义拓展，就是把一个数学对象合理地表示成若干更简单对象与实数系数之积的和。但微积分和线性代数各有侧重。微积分研究的是无穷项求和。无穷项之和与有穷项之和是本质不同的。但是无穷项之和是无法运算的，至少不实际。所以要想办法通过一种办法用有穷项之和来近似的代替，这就是逼近。逼近成立的条件是收敛，就是说，只有从一个收敛的无穷项的开头截出一部分来求和，才能被认为是逼近。华人数学家项武义说，微积分就逼近这一板斧，但是无往而不利。

微积分主要研究函数，连续函数的因变量y会由于自变量x的变化而变化。这种变化也是要分析的。当x从 x0变成x1时，y是怎样从y0变到y1 的？按照上面的说法，“y的变化（y1-y0)”这一个数学对象，要用一系列比较简单的“变化”相加来表示。数学家找到了一个收敛的“变化”对象的序列，排在头一位的是一个线性的变化量，它的系数就是导数，它本身就是微分dy。数学家又发现，当x的变化量无穷小时，从这个无穷的、收敛的“变化”对象序列中，只要截出第一项，也就是微分dy，就无论如何可以精确描述y的变化了。曾在一本书上见过这样的说法，泰勒公式是数学分析的顶峰。不知道是不是有道理。我自己觉得是这么回事。有了泰勒公式，我们可以任意精确地算一个函数在某一点上的值。毕竟只是实数求和嘛。

但是为了表示泰勒公式，我们却用了一个挺复杂的连加代数式。代数式不能象实数那样简单加起来得到一个对象，它只能表示成和的形式。这是我们意识到，在这个连加式中各对象存在某些特别的不同，使它们没法简单地加到一起。 因此我们有必要讨论，把一些性质不同的东西加到一起所形成的这个对象有什么性质。 这就是向量。

微积分研究如何把一个对象分解为无穷项同质对象之和，线性代数研究“有限项异质对象之和”这个新对象的性质。一方面，上面说过，微积分到最后还是要化无穷为有穷，化精确为逼近；另一方面，异质对象经过某种处理可以转化为同质对象。比如不同次的幂函数是异质对象，但是一旦代入具体数值则都可以转化为实数，变成了同质对象。因此线性代数研究的问题对微积分很重要。故我认为大学里应先讲线性代数，后讲微积分。

我们的微积分教学，将重点过分倾注在微分和积分的运算上了，其实实践中更为重要的是我们称为“级数”的那部分内容。即研究如何将一个量表达为一个数项级数，如何将一个函数表达为一个函数项级数。

线性代数把异质对象之和（向量）作为研究的基础，研究这些新定义的对象加起来又可以表示什么。其结论是，有限数量的向量连加起来，有可能具有这样的能力，即同维的全部向量都可以表示成这些向量的和。这样的一组具有充分表现能力的向量，是线性无关的向量，组成了一个向量空间，而它们自己构成了这个向量空间里的一组基。

回到分析的概念上，一个向量总可以表示为若干个同阶向量之和，这就是向量的分析。但是并不是所有的这些分析都具有相同的价值。在某种运算中，某种特别的分析能够提供特别优越的性，从而大大简化运算。比如在大多数情况下，将一个向量表示成一组单位正交基向量的和，就能够在计算中获得特别的便利。面对某个问题，寻找一个最优越的分析形式，把要研究的对象合理地表示成具有特殊性质的基对象与实数系数之积的和，这是分析的重要步骤，也是成功的关键。在这种表示式中，系数称为坐标。

经典的方法都是以找到一组性质优良的基为开端的，例如：

傅立叶分析以正交函数系为基，因此具有优良性质，自1904年以来取代幂函数系，成为分析主流。

在曲线和曲面拟合中，正交多项式集构成了最佳基函数。 拉格朗日插值多项式具有一个特别的性质，即在本结点上为1，在其他结点上为0。

有限元中的形函数类似拉氏插值多项式。

结构动力学中的主振型迭加法，也是以相互正交的主振型为基，对多质点体系位移进行分析的。

举两个例子：说到采样，大家的第一反应肯定是一个词“2倍”（采样定理）。学得比较扎实的，可能还会把为什么是2倍解释清楚。但我对采样的理解是：采样实际上是在进行正交分解，采样值不过是在一组正交基下分解的系数。如果原信号属于该组正交基所张成的线性子空间，那么该信号就能无失真的恢复（满足采样定理）。学过信号处理的朋友，你知道这组正交基是什么吗？:)第二个例子是关于为什么傅里叶变换在线性系统理论中如此重要？答案可能五花八门，但我认为我的理解是比较深入的：原因是傅里叶基是所有线性时不变算子的特征向量（和本文联系起来了）。这句话解释起来比较费工夫，但是傅里叶变换能和特征向量联系起来，大家一定感觉很有趣吧。

特征向量确实有很明确的几何意义，矩阵(既然讨论特征向量的问题，当然是方阵，这里不讨论广义特征向量的概念，就是一般的特征向量)乘以一个向量的结果仍是同维数的一个向量，因此，矩阵乘法对应了一个变换，把一个向量变成同维数的另一个向量，那么变换的效果是什么呢？这当然与方阵的构造有密切关系，比如可以取适当的二维方阵，使得这个变换的效果就是将平面上的二维向量逆时针旋转30度，这时我们可以问一个问题，有没有向量在这个变换下不改变方向呢？可以想一下，除了零向量，没有其他向量可以在平面上旋转30度而不改变方向的，所以这个变换对应的矩阵(或者说这个变换自身)没有特征向量(注意：特征向量不能是零向量)，所以一个变换的特征向量是这样一种向量，它经过这种特定的变换后保持方向不变，只是进行长度上的伸缩而已(再想想特征向量的原始定义Ax=cx,你就恍然大悟了,看到了吗？cx是方阵A对向量x进行变换后的结果，但显然cx和x的方向相同)，而且x是特征向量的话，ax也是特征向量(a是标量且不为零)，所以所谓的特征向量不是一个向量而是一个向量族，另外，特征值只不过反映了特征向量在变换时的伸缩倍数而已，对一个变换而言，特征向量指明的方向才是很重要的，特征值不是那么重要，虽然我们求这两个量时先求出特征值，但特征向量才是更本质的东西！

比如平面上的一个变换，把一个向量关于横轴做镜像对称变换，即保持一个向量的横坐标不变，但纵坐标取相反数，把这个变换表示为矩阵就是[1 0;0 -1]，其中分号表示换行，显然[1 0;0 -1]\*[a b]'=[a -b]',其中上标'表示取转置，这正是我们想要的效果，那么现在可以猜一下了，这个矩阵的特征向量是什么?想想什么向量在这个变换下保持方向不变，显然，横轴上的向量在这个变换下保持方向不变(记住这个变换是镜像对称变换，那镜子表面上(横轴上)的向量当然不会变化)，所以可以直接猜测其特征向量是 [a 0]'(a不为0)，还有其他的吗？有，那就是纵轴上的向量，这时经过变换后，其方向反向，但仍在同一条轴上，所以也被认为是方向没有变化，所以[0 b]'(b不为0)也是其特征向量，去求求矩阵[1 0;0 -1]的特征向量就知道对不对了!