

教你充分理解矩阵

线性代数课程，无论你是否从行列式入手还是直接从矩阵入手，从一开始就充斥着莫名其妙。比如说，在全国一般工科院系教学中应用最广泛的同济线性代数教材（现在到了第四版），一上来就介绍逆序数这个“前无古人，后无来者”的古怪概念，然后用逆序数给出行列式的一个极不直观的定义，接着是一些简直犯傻的行列式性质和习题——把这行乘一个系数加到另一行上，再把那一列减过来，折腾得那叫一个热闹，可就是压根看不出这个东西有嘛用。大多数像我一样资质平庸的学生到这里就有点犯晕：连这是个什么东西都模模糊糊的，就开始钻火圈表演了，这未免太“无厘头”了吧！于是开始有人逃课，更多的人开始抄作业。这下就中招了，因为其后的发展可以用一句峰回路转来形容，紧跟着这个无厘头的行列式的，是一个同样无厘头但是伟大的无以复加的家伙的出场——矩阵来了！多年之后，我才明白，当老师犯傻似地用中括号把一堆傻了吧叽的数据括起来，并且不紧不慢地说：“这个东西叫做矩阵”的时候，我的数学生涯掀开了何等悲壮辛酸、惨绝人寰的一幕！自那以后，在几乎所有跟“学问”二字稍微沾点边的东西里，矩阵这个家伙从不缺席。对于我这个没能一次搞定线性代数的笨蛋来说，矩阵老大的不请自来每每搞得我灰头土脸，头破血流。长期以来，我在阅读中一见矩阵，就如同阿 Q 见到了假洋鬼子，揉揉额角就绕道走。

事实上，我并不是特例。一般工科学生初学线性代数，通常都会感到困难。这种情形在国内外皆然。瑞典数学家 Lars Garding 在其名著 *Encounter with Mathematics* 中说：“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。”事实上，当我们开始学习线性代数的时候，不知不觉就进入了“第二代数学模型”的范畴当中，这意味着数学的表述方式和抽象性有了一次全面的进化，对于从小一直在“第一代数学模型”，即以实用为导向的、具体的数学模型中学习的我们来说，在没有并明确告知的情况下进行如此剧烈的 **paradigm shift**，不感到困难才是奇怪的。

大部分工科学生，往往是在学习了一些后继课程，如数值分析、数学规划、矩阵论之后，才逐渐能够理解和熟练运用线性代数。即便如此，不少人即使能够很熟练地以线性代数为工具进行科研和应用工作，但对于很多这门课程的初学者提出的、看上去是很基础的问题却并不清楚。比如说：

* 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 $m \times n$ 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 A^T ，有 $(AB)^T = B^T A^T$ ，对于矩阵求逆运算 A^{-1} ，有 $(AB)^{-1} = B^{-1} A^{-1}$ 。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 $P^{-1}AP$ 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 $Ax = \lambda x$ ，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？

这样的一类问题，经常让使用线性代数已经很多年的人都感到为难。就好像大人面对小孩子的刨根问底，最后总会迫不得已地说“就这样吧，到此为止”一样，面对这样的问题，很多老手们最后也只能用：“就是这么规定的，你接受并且记住就好”来搪塞。然而，这样的问题如果不能获得回答，线性代数对于我们来说就是一个粗暴的、不讲道理的、莫名其妙的规则集合，我们会感到，自己并不是在学习一门学问，而是被不由分说地“抛到”一个强制的世界中，只是在考试的皮鞭挥舞之下被迫赶路，全然无法领略其中的美妙、和谐与统一。直到多年以后，我们已经发觉这门学问如此的有用，却仍然会非常迷惑：怎么这么凑巧？

我认为，这是我们的线性代数教学中直觉性丧失的后果。上述这些涉及到“如何能”、“怎么会”的问题，仅仅通过纯粹的数学证明来回答，是不能令提问者满意的。比如，如果你通过一般的证明方法论证了矩阵分块运算确实可行，那么这并不能够让提问者的疑惑得到解决。他们真正的困惑是：矩阵分块运算为什么竟然是可行的？究竟只是凑巧，还是说这是由矩阵这种对象的某种本质所必然决定的？如果是后者，那么矩阵的这些本质是什么？只要对上述那些问题稍加考虑，我们就会发现，所有这些问题都不是单纯依靠数学证明所能够解决的。像我们的教科书那样，凡事用数学证明，最后培养出来的学生，只能熟练地使用工具，却欠缺真正意义上的理解。

自从 1930 年代法国布尔巴基学派兴起以来，数学的公理化、系统性描述已经获得巨大的成功，这使得我们接受的数学教育在严谨性上大大提高。然而数学公理化的一个备受争议的副作用，就是一般数学教育中直觉性的丧失。数学家们似乎认为直觉性与抽象性是矛盾的，因此毫不犹豫地牺牲掉前者。然而包括我本人在内的很多人都对此表示怀疑，我们不认为直觉性与抽象性一定相互矛盾，特别是在数学教育中和数学教材中，帮助学生建立直觉，有助于它们理解那些抽象的概念，进而理解数学的本质。反之，如果一味注重形式上的严格性，学生就好像被迫进行钻火圈表演的小白鼠一样，变成枯燥的规则的女奴隶。

对于线性代数的类似上述所提到的一些直觉性的问题，两年多来我断断续续地反复思考了四、五次，为此阅读了好几本国内外线性代数、数值分析、代数和数学通论性书籍，其中像前苏联的名著《数学：它的内容、方法和意义》、龚昇教授的《线性代数五讲》、前面提到的 *Encounter with Mathematics*（《数学概观》）以及 Thomas A. Garrity 的《数学拾遗》都给我很大的启发。不过即使如此，我对这个主题的认识也经历了好几次自我否定。比如以前思考的一些结论曾经写在自己的 blog 里，但是现在看来，这些结论基本上都是错误的。因此打算把自己现在的有关理解比较完整地记录下来，一方面是因为我觉得现在的理解比较成熟了，可以拿出来与别人探讨，向别人请教。另一方面，如果以后再有进一步的认知，把现在的理解给推翻了，那现在写的这个 snapshot 也是很有意义的。

因为打算写得比较多，所以会分几次慢慢写。也不知道是不是有时间慢慢写完整，会不会中断，写着看吧。

今天先谈谈对线性空间和矩阵的几个核心概念的理解。这些东西大部分是凭着自己的理解写出来的，基本上不抄书，可能有错误的地方，希望能够被指出。但我希望做到直觉，也就是说能把数学背后说的实质问题说出来。

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，

上面的这些性质中，最最关键的是第**4**条。第**1**、**2**条只能说是空间的基础，不算是空间特有的性质，凡是讨论数学问题，都得有一个集合，大多数还得在这个集合上定义一些结构（关系），并不是说有了这些就算是空间。而第**3**条太特殊，其他的空间不需要具备，更不是关键的性质。只有第**4**条是空间的本质，也就是说，容纳运动是空间的本质特征。

认识到了这些，我们就可以把我们关于三维空间的认知扩展到其他空间。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相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 x_0, x_1, \dots, x_n 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 a_i 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 L_1 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 $n \times 1$ 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。

接着理解矩阵。

上一篇里说“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。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 4×4 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 4×4 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。下面我们只探讨最常用、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：

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$$

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坏的。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。