Vector Space

FrankZhou-jun*

2019 年 10 月 2 日

1 为啥叫向量空间

向量空间也叫线性空间，不得不得吐槽国内的某些老师，让我感觉上了线性代数之后，是一种萌萌的状态，不了解其本质，被强制接受某种概念，最终过了，但毕业只有一只迷茫困惑，线性代数到底是个什么玩意，怎么用才能对得起交的那些学费。参考网上的一篇帖子[1], 是一篇 CSDN 的博客：

对于提到的这几本书像前苏联的名著《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》，先行几下，以备以后之需。原文说的很好这里直接引用如下：

先谈谈对线形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识扩展到其他的空间。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？

---

*研究方向：信号处理，机械故障诊断，深度学习，强化学习，邮箱:zhoujun14@yeah.net

[1]《关于矩阵最通俗的解释》

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。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。

2 常见的英文单词意思

2.1 pointwise

这里的 wise 是一个后缀，英文解释就是 in a specified manner,direction,or postion, 以一种特定方式，方向，或位置，比如 clockwise，顺时针方向，please turn the key in a clockwise direction.

所以 pointwise 指的是逐点的方式.

维基百科里的解释是：In mathematics, the qualifier pointwise is used to indicate that a certain property is defined by considering each value$\boldsymbol{f}(x)$of some function $\boldsymbol{f}$. An important class of pointwise concepts are the pointwise operations, that is operations defined on functions by applying the operations to function values separately for each point in the domain of definition. Important relations can also be defined pointwise.[2]

$$\begin{aligned}(\boldsymbol{f}+\boldsymbol{g})(x) &= \boldsymbol{f}(x)+\boldsymbol{g}(x)\\ (\boldsymbol{f}\cdot\boldsymbol{g})(x) &= \boldsymbol{f}(x)\cdot\boldsymbol{g}(x)\\ (\lambda\boldsymbol{f})(x) &= \lambda\boldsymbol{f}(x)\end{aligned}\tag{1}$$

虽然上面英文看不懂，但是通过公式可以看到，将函数之间的操作，分开应用到定义域间每一个变量得到函数值上，比如函数 $\boldsymbol{f}+\boldsymbol{g}$ 这种操作，（逐点的）应用到在定义域间每一个函数值 $\boldsymbol{f}$(x),$\boldsymbol{g}$(x) 上，即 $\boldsymbol{f}(x)+\boldsymbol{g}(x)$。

2.2 component wise

Componentwise operations are usually defined on vectors, where vectors are elements of the set$\boldsymbol{K}^n$ for some natural number n and some field $\boldsymbol{K}$ If we denote the i-th component of any vector $\boldsymbol{v}$ as $\boldsymbol{v}_i$, then componentwise addition is $(\boldsymbol{u}+\boldsymbol{v})_i=\boldsymbol{u}_i+\boldsymbol{v}_i$。

Componentwise operations can be defined on matrices. Matrix addition, where $\boldsymbol{A}+\boldsymbol{B}_{ij}=\boldsymbol{A}_{ij}+\boldsymbol{B}_{ij}$ is a componentwise operation while matrix multiplication is not. A tuple can be regarded as a function, and a vector is a tuple. Therefore, any vector $\boldsymbol{v}$ corresponds to the function $\boldsymbol{f}:n\to\boldsymbol{K}$such that $\boldsymbol{f}(i)=\boldsymbol{v}_i$, and any componentwise operation on vectors is the pointwise operation on functions corresponding to those vectors[3].

component-wise 可以理解为逐主元素的方式，因为向量可以看做是由元素组成的集合，对集合里面的元素进行逐元素的操作。

[2]Wikipedia

[3]Wikipidia