题　　目：《神奇的矩阵》(修改版 1.12)

学　　校：哈尔滨工程大学

姓　　名：黎文科

联系方式：QQ 群：53937814

联系方式：190356321@qq.com

Contents

前言

这里只是整理了孟岩老师的《理解矩阵》和任广千、胡翠芳老师的《线性代数的几何意义》，在改正一些错误的基础上，前一部分基本忠实原文。但是本文并不是一次简单的复制粘贴，我在最后一部分加入了自己的一些感悟和理解，分别对矩阵的乘法，等价，相似、对角化等做出了讨论。文中重要的内容处采用楷体加粗，以示区分。可以说，本文是对原来文章的一次整理和升华。这里只是出于一种对数学的爱好！有兴趣的读者建议阅读原文，希望不会引起读者朋友的拍板砖。

本文的大部分内容取材于David C.Lay的《线性代数及其应用》、孟岩老师的《理解矩阵》、任广千、胡翠芳老师的《线性代数的几何意义》、维基百科、范崇金、王锋老师的《线性代数》(也就是我们学校的教材，我感觉这是一本不错的教材，讲解很透彻)。由于线性代数大家都学过，没有秘密可言。数学的好经验应该大家共享，我们自己也是这么学来的。作者愿意公开本文的电子文档。

由于本人水平有限，错误在所难免，欢迎读者对本文提出批评建议。相信每一次的思考，不管对错，都能对你的理解做出贡献。希望这篇拙作能起到抛砖引玉的作用。

谨以此文献给我的母校哈尔滨工程大学，作为一份建校六十周年的纪念！

——作者

2013年9月于哈尔滨

绪论

学过线性代数的人都感觉到，线性代数带来的困惑实在太多了：对于线性代数，无论你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，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。

要想提高自己的专业水平，你肯定深感数学能力的重要。随便打开一篇专著或论文，满纸的微分方程、矩阵扑面而来。竭力迎头而上，每每被打得灰头土脸、晕头转向。我天生就不是搞数学的？我的智力有问题吗？

太失望了，太伤自尊了。转头看看周围，莫不雷同。大多的工程师们靠经验来工作，经验靠时间或试验来积累。数学应用的层次最多就是高中水平。也有硕士博士级的牛人，但也少见把数学工具在工作中应用的得心应手、手到擒来的。

数学工具在科技实践中缺失的严重，导致我们的科技创新能力的严重缺失。普遍现象，绝对的。返回来想一想，我的智力应该没问题，重点大学都毕业了，能有多严重的问题？所有的工程师们、大学毕业生们的智力也没问题。问题是大家没把数学学好，没有真正掌握它。为啥没有在四年的大学阶段学好《线代》呢？要知道，学生是通过高考百里挑一录取的，智力应是足够正常的。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。

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，变成枯燥的规则的奴隶。

”我们在小学和中学的学习阶段，老师常常也讲一些抽象概念所对应的几何意义，为何到了大学我们的大脑就一下子高度抽象起来了？把形象仍得远远的，象瘟疫一样躲着他？目的是训练抽象思维？最终实际结果呢？不可否认，大学毕业后大家确实是抽象了，抽象得只会夸夸其谈讲理论不会干具体活了。既然你具体的活计不会干那干脆就专搞抽象的理论去嘛，结果也搞不了，为啥？只会做做过的抽象的数学题不会发明创造，没学会真正的抽象，真是越抽象越糊涂。

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。

我觉得，抽象和形象是相辅相成，缺一不可的。由形象而抽象，再由抽象到形象，人的知识结构螺旋架才能旋转而上，达到越来越高的知识峰巅。

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？

* 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？它又有什么神通广大的应用呢？

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？

线性代数有什么用？这是每一个圈养在象牙塔里，在灌输式教学模式下的“被学习”的学生刚刚开始思考时的第一个问题。我稍微仔细的整理了一下学习线代的理由，竟然也罗列了不少，不知道能不能说服你：

1、如果你想顺利地拿到学位，线性代数的学分对你有帮助；

2、如果你想继续深造，考研，必须学好线代。因为它是必考的数学科目，也是研究生科目《矩阵论》、《泛函分析》的基础。例如，泛函分析的起点就是无穷多个未知量的无穷多线性方程组理论。

3、如果你想提高自己的科研能力，不被现代科技发展潮流所抛弃，也必须学好，因为瑞典的 L.戈丁说过，没有掌握线代的人简直就是文盲。他在自己的数学名著《数学概观》中说：要是没有线性代数，任何数学和初等教程都讲不下去。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，其根源来自于欧几里得几何、解析几何以及线性方程组理论。…，如果不熟悉线性代数的概念，像线性性质、向量、线性空间、矩阵等等，要去学习自然科学，现在看来就和文盲差不多，甚至可能学习社会科学也是如此。

4、如果毕业后想找个好工作，也必须学好线代：

- 想搞数学，当个数学家（我去，这个还需要列出来，谁不知道线代是数学）。恭喜你，你的职业未来将是最光明的。如果到美国打工的话你可以找到最好的职业。

- 想搞电子工程，好，电路分析、线性信号系统分析、数字滤波器分析设计等需要线代，因为线代就是研究线性网络的主要工具；进行 IC 集成电路设计时，对付数百万个集体管的仿真软件就需要依赖线性方程组的方法；想搞光电及射频工程，好，电磁场、光波导分析都是向量场的分析，比如光调制器分析研制需要张量矩阵，手机信号处理等等也离不开矩阵运算。

- 想搞软件工程，好，3D 游戏的数学基础就是以图形的矩阵运算为基础；当然，如果你只想玩 3D 游戏可以不必掌握线代；想搞图像处理，大量的图像数据处理更离不开矩阵这个强大的工具，《阿凡达》中大量的后期电脑制作没有线代的数学工具简直难以想象。

- 想搞经济研究。好，知道列昂惕夫（Wassily Leontief）吗？哈佛大学教授，1949 年用计算机计算出了由美国统计局的 25 万条经济数据所组成的 42 个未知数的 42 个方程的方程组，他打开了研究经济数学模型的新时代的大门。这些模型通常都是线性的，也就是说，它们是用线性方程组来描述的，被称为列昂惕夫“投入-产出”模型。列昂惕夫因此获得了 1973 年的诺贝尔经济学奖。

- 相当领导，好，要会运筹学，运筹学的一个重要议题是线性规划。许多重要的管理决策是在线性规划模型的基础上做出的。线性规划的知识就是线代的知识啊。比如，航空运输业就使用线性规划来调度航班，监视飞行及机场的维护运作等；又如，你作为一个大商场的老板，线性规划可以帮助你合理的安排各种商品的进货，以达到最大利润。

- 对于其他工程领域，没有用不上线代的地方。如搞建筑工程，那么奥运场馆鸟巢的受力分析需要线代的工具；石油勘探，勘探设备获得的大量数据所满足的几千个方程组需要你的线代知识来解决；飞行器设计，就要研究飞机表面的气流的过程包含反复求解大型的线性方程组，在这个求解的过程中，有两个矩阵运算的技巧：对稀疏矩阵进行分块处理和进行 LU 分解； 作餐饮业，对于构造一份有营养的减肥食谱也需要解线性方程组；知道有限元方法吗？这个工程分析中十分有效的有限元方法，其基础就是求解线性方程组。知道马尔科夫链吗？这个“链子”神通广大，在许多学科如生物学、商业、化学、工程学及物理学等领域中被用来做数学模型，实际上马尔科夫链是由一个随机变量矩阵所决定的一个概率向量序列，看看，矩阵、向量又出现了。

- 另外，矩阵的特征值和特征向量可以用在研究物理、化学领域的微分方程、连续的或离散的动力系统中，比如结构动力学、刚体动力学、振动力学等，而且不论是机械振动还是振荡电路，只要有振动的地方就有求矩阵的特征值和特征向量的问题。甚至数学生态学家用以在预测原始森林遭到何种程度的砍伐会造成猫头鹰的种群灭亡；大名鼎鼎的最小二乘算法广泛应用在各个工程领域里被用来把实验中得到的大量测量数据来拟合到一个理想的直

线或曲线上，最小二乘拟合算法实质就是超定线性方程组的求解；计算机人脸识别中也应用到矩阵的特征值和特征向量。

 二次型常常出现在线性代数在工程（标准设计及优化）和信号处理（输出的噪声功率）的应用中，他们也常常出现在物理学（例如势能和动能）、微分几何（例如曲面的法曲率）、经济学（例如效用函数）和统计学（例如置信椭圆体）中，某些这类应用实例的数学背景很容易转化为对对称矩阵的研究。

嘿嘿（脸红），说实在的，我也没有足够经验讲清楚线代在各个工程领域中的应用，只能大概人云亦云地讲述以上线代的一些基本应用。因为你如果要真正的讲清楚线代的一个应用，就必须充分了解所要应用的领域内的知识，最好有实际的工程应用的经验在里面；况且线性代数在各个工程领域中的应用真是太多了，要知道当今成为一个工程通才只是一个传说。

总结一下，线性代数的应用领域几乎可以涵盖所有的工程技术领域。如果想知道更详细的应用材料，建议看一下《线性代数及应用》，这是美国 David C. Lay 教授写的迄今最现代的流行教材。国内的教材可以看看《线性代数实践及 MATLAB 入门》，这是西电科大陈怀琛教授写的最实用的新教材。

线性代数的内容及发展简史

线性代数是高等代数的一大分支。通过前面的章节的介绍，我们知道，在研究关联着多个因素的量所引起的问题，则需要考察多元函数。如果所研究的关联性是线性的，那么称这个问题为线性问题。一次方程就是研究线性问题的方程，被称为线性方程，讨论线性方程及线性运算的代数就叫做线性代数。历史上线性代数的第一个问题是关于解线性方程组的问题，而线性方程组理论的发展又促成了作为工具的行列式理论和矩阵论的创立与发展，这些内容已成为我们线性代数教材的主要部分。作为代表“线性”的最基本的概念--向量的概念，从数学的观点来看不过是有序三元数组的一个集合，然而它以力或速度作为直接的物理意义，并且数学上用它能立刻写出物理上所说的事情。向量用于梯度，散度，旋度就更有说服力。同样，行列式和矩阵如导数一样（虽然 dy/dx 在数学上不过是一个符号，表示包括△y/△x 的极限的式子，但导数本身是一个强有力的概念，能使我们直接而创造性地想象物理上发生的事情）。因此，虽然表面上看，行列式和矩阵不过是一种语言或速记，但它的大多数生动的概念能对新的思想领域提供钥匙。然而已经证明这两个概念是数学物理上高度有用的工具。

用一个表格总结一下线性代数的发展历史上做出重要贡献的数学家，如下：

	人物	理论贡献	力量指数
1	せき たかかず（关孝和）	最早提出行列式概念	★★★
2	A-T.Vandermonde（范德蒙）	首次把行列式理论与线性方程组求解相分离	★★★
3	Augustin Louis Cauchy（柯西）	1815 年启用行列式名词，1841 年提出特征方程概念	★★★★
4	James Joseph Sylvester（西尔维斯特）	1850 年启用矩阵名词，1852 年发现惯性定律	★★★
5	Arthur Cayley（凯莱）	1855 年引入定义矩阵乘法等运算	★★★★★
6	Henry John Stephen Smith（史密斯）	引进了方程组的增广矩阵和非增广矩阵的概念	★★★
7	Carl Gustav Jacob Jacobi（雅可比）	重新发现并证明惯性定律	★★★
8	Hermann Günther Graßmann（格拉斯曼）	1844 至 1862 年间创建高维线性空间理论	★★★★★
9	Karl Theodor Wilhelm Weierstraß（魏尔斯特拉斯）	1868 年完成二次型理论	★★★

1 线性代数的根基–空间

好了，言归正传，下面让我们来理解一下矩阵的神奇！我们先谈谈对线形空间和矩阵的几个核心概念的理解。

首先，线性代数里面的线性主要的意思就是线性空间里的线性变换。线性变换或线性映射是把中学的线性函数概念进行了重新定义，强调了函数的变量之间的变换的意义。线性函数的概念在初等数学和高等数学中含义不尽相同(高等数学常常把初等数学的关键概念进行推广或进一步抽象化，初等数学的概念就变成了高等数学概念的一个特例)。

在中学的初等数学里，我们知道，函数 f(x)=kx+b（k 和 b 是不变量），称为一元线性函数，因为在平面直角坐标系中这个函数的图形就是一条直线，就是变量（包括自变量和因变量）之间的关系描述为一条直线，所以把这种函数形象地称为“线性”函数；如果 b=0 ，这个函数的外观就变成 f(x)=kx 的形式了，这是一条过原点的直线。显然，过原点的直线是最简单的线性函数。

在大学的代数里面，为了线性函数的进一步推广（如推广至双线性函数、多线性函数、线性空间、线性泛函…）的远大未来，我们忍痛割“尾”，把一元线性函数 f (x)= kx + b 的 b 割舍掉，成了 f(x)=kx 的形式。呵呵，简单点说，只有过原点的最简单的直线 f (x)= kx 才被称为一元线性函数。

为什么？只因为不过原点的直线不满足我们对线性函数的比例性的要求。

线性函数表现为直线，这只是几何意义。那么所谓“线性”的代数意义是什么呢？实际上，最基本的意义只有两条：**可加性和比例性**。用数学的表达来说就是：**对加法和数乘封闭。**

然后说说空间(space)，这个概念是现代数学的命根子之一。对于空间的的理解需要更抽象一些，**简单的说，能装东西的就是空间。**比如计算机内有存储单元，那么就有内存空间；我们上课有课表，那么就有课表空间；有一个能装载梦境的东西，我们可以叫它盗梦空间。对于数学来说，数学家定义的空间里装载的当然是能运算的东西。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，如果空间里装载所有类型的函数，就叫泛函空间。

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或关系，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？

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 $\mathbf{x_0, x_1, ..., x_n}$ 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 $\mathbf{a_i}$ 其实就是多项式中 $\mathbf{x_{(i-1)}}$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。换句话说，给你一个空间，你就能建立一个坐标系，来描述这个空间中的对象！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。

上面说了，如果有一个线性空间，就能建立一个坐标系，选取一组基和坐标来描述这个空间里的对象（也就是向量）。那么这组基怎么选取呢？我们知道，线性空间里的基本对象是向量，所以，基一定也是向量。但有一些要求：对基的要求是数量要够，还要线性无关。常常一组基表示成如下形式

$$\left[\vec{\alpha}_1,\ \vec{\alpha}_2,\ \vec{\alpha}_3,\ ...,\ \vec{\alpha}_n\right]$$

可以看到这是一个矩阵而且是非奇异的（也称满秩），对于非奇异方阵，它也描述了一个 n 维空间。举个例子吧，比如在平面内的任意向量，我们可以用两个不共线（线性无关）的向量描述，这就是我们在高中熟知的平面向量基本定理，这两个不共线的向量就构成描述这个空间的 2×2 矩阵。对于空间的任意向量，我们可以用三个不共线（线性无关）的三维向量描述，这三个向量就构成描述这个空间的 3×3 矩阵。对于高维空间的任意向量，自然要 n 个线性无关的向量构成，于是就构成一个 n×n 的矩阵！**这些线性无关的向量就叫做基！**

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（**单位是 1，但是长度不一定是 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，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：**由一组线性无关的向量放在一起构成的矩阵描述了一个坐标系。**

2 矩阵描述运动

先回顾一下上节的内容：我们先确立了一种叫线性空间的东西，然后我们建立了坐标系：选取了一组基，于是空间里的对象就有了坐标。世界是物质的，物质是运动的，接下来肯定要研究一下线性空间里面的运动是怎么实现的，也就是来回答第二个问题，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。**

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。

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V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。（本人看后觉得有点问题.注：线性变换说的是从自己空间到自己空间的变换）

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。

用数学的表达式，我们写成：

$\vec{\beta} = M\vec{\alpha}$ **的意思是说矩阵 M 描述了向量 $\vec{\alpha}$ 到 $\vec{\beta}$ 的变换（运动）。而矩阵对一个向量的作用无非是把它伸缩或者旋转。**

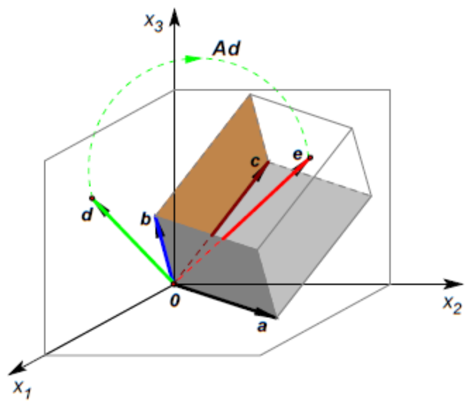
$B = MA$ **的意思是说矩阵 M 描述了一组向量 {$\vec{\alpha}$} 到 {$\vec{\beta}$} 的变换（运动）。而矩阵对这一个组向量的作用无非是把它们伸缩或者旋转。**

让我们看几个图像：

<table>
<tr><th>顺序</th><th>图形解释</th><th>对应的几何图形</th></tr>
<tr><td>1</td><td>

$$\mathbf{A}_1\cdot\mathbf{d}_1=\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}+d_2\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}+d_3\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}=\begin{pmatrix}e_1\\e_2\\e_3\end{pmatrix}$$

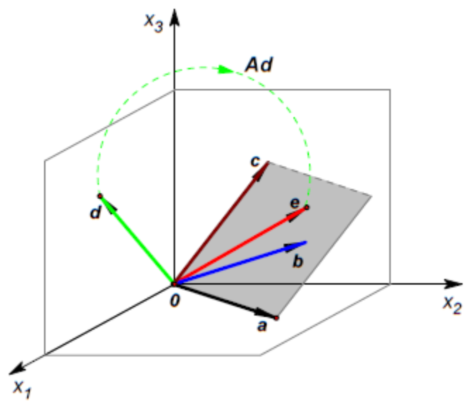
3×3 矩阵 **A** 把一个三维向量 **d** 映射到一个三维向量 **e** 。</td><td>



</td></tr>
<tr><td>2</td><td>

$$\mathbf{A}_2\cdot\mathbf{d}_2=\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1\begin{pmatrix}a_1\\a_2\end{pmatrix}+d_2\begin{pmatrix}b_1\\b_2\end{pmatrix}+d_3\begin{pmatrix}c_1\\c_2\end{pmatrix}=\begin{pmatrix}e_1\\e_2\end{pmatrix}$$

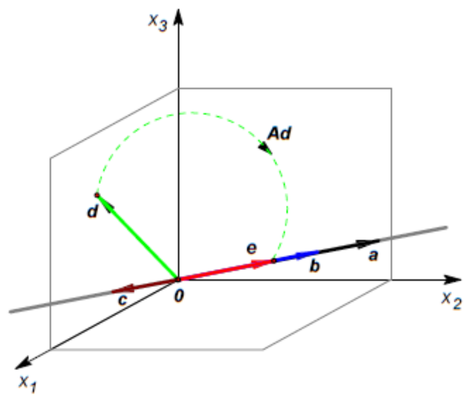
2×3 矩阵 **A** 把一个三维向量 **d** 映射到一个二维向量 **e** 。</td><td>



</td></tr>
<tr><td>3</td><td>

$$\mathbf{A}_3\cdot\mathbf{d}_3=\begin{bmatrix}a_1&b_1&c_1\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1a_1+d_2b_1+d_3c_1=e$$

1×3 矩阵 **A** 把一个三维向量 **d** 映射到一个一维向量 **e** （或实数 e ）。</td><td>



</td></tr>
</table>

到这里你可能有点晕了，矩阵到底是坐标系还是运动的描述？其实，应该这样理解：坐标系是由线性无关的向量放在一起构成的，只是表示成矩阵的形式而已，而我们将常用的是把矩阵看成运动（线性代数里常称为变换）的描述，用乘法来施加，即**矩阵本身描述了一个坐标系，矩阵与矩阵的乘法描述了一个运动。换句话说：如果矩阵仅仅自己出现，那么他描述了一个坐标系，如果他和另一个矩阵或向量同时出现，而且做乘法运算，那么它表示运动！**。

3 矩阵与方程组

接下来，我们从矩阵与方程组的角度理解。当解一个方程组的时候，我们知道，方程组的解只与系数有关。于是，将系数提取出来，放在一起，就构成了矩阵，而矩阵的行变换就相当于解方程时的加减消元过程（也叫高斯消元法）。这是直观的，也是很好理解的。

下面我们从方程组解的几何意义来解释方程组 **Ax=b** 的意义。先把结论给出来：

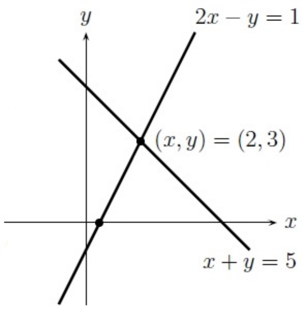
1、方程组有解还可以理解成空间几何图形有公共交点或者交线（高维度时还可能出现“交面”等情况）

2、方程组有解就说明 **b** 这个向量能用 A 的列向量线性表示，也可以说 **b** 这个向量在 **A** 的列向量所构成的空间中！

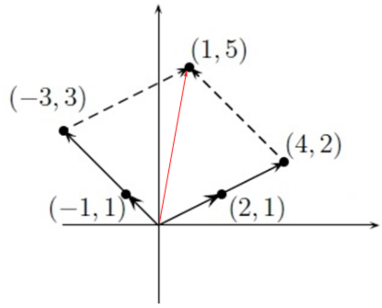
对于第一种解释我们非常熟悉，我就不做过多解释了，相信你能明白。

对于第二种解释，我们可以将 n×m 的矩阵 **A** 写成这样：$A=[\overrightarrow{\alpha_1},\overrightarrow{\alpha_2}\cdots\overrightarrow{\alpha_n}]$,于是，方程组就变成 $x_1\overrightarrow{\alpha_1}+x_2\overrightarrow{\alpha_2}\cdots x_n\overrightarrow{\alpha_n}=\vec{b}$。这样就很清楚了：如果一个方程组有解，就说明 **b** 这个向量能用 A 的列向量线性表示，也可以说 **b** 这个向量在 **A** 的列向量所构成的空间中！

举一个例子，对于方程组 $\begin{cases}2x-y=1\\ x+y=5\end{cases}$ 的几何意义如下



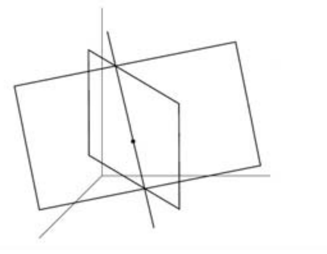
第一种解释



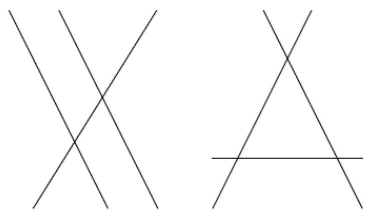
第二种解释

那么，无穷解和无解怎么理解呢？

对于第一种解释，无穷解代表有公共相交部分，但是最终形成的不是一个点，而是线或者面或者更高维的东西，而无解代表着没有公共相交的部分。如下图所示

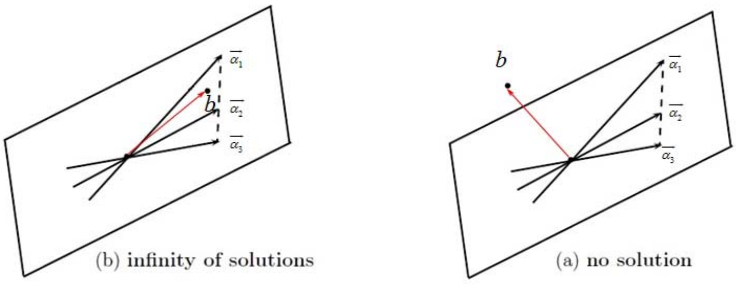


无穷解



无解

对于第二种解释，无穷解表示矩阵 **A** 中的列向量线性相关，所以导致表示方式不唯一。而无解则表示矩阵 **A** 中的列向量和 **b** 不在同一个“次元”里，或者说不在一个空间里。如果 **A** 的列向量和 **b** 在同一个空间里，那么就需要 **A** 的维数和**[A,b]**的维数相同。我们知道，矩阵的秩表示向量构成的空间的维数，于是我们知道 **r(A)**应该等于 **r([A,b])**，这仿佛是理所应当的事情。怎么理解无穷多解呢？比如在下图中，我们选取 $\overrightarrow{\alpha_1},\overrightarrow{\alpha_2}$ 做基向量，那么 **b** 可以用 $\overrightarrow{\alpha_1},\overrightarrow{\alpha_2}$ 唯一表示，这是我们让 $x_3=0$ 就有一组特解，然后我们发现一些组合可以使 $x_{01}\overrightarrow{\alpha_1}+x_{02}\overrightarrow{\alpha_2}+x_{03}\overrightarrow{\alpha_3}$ 变成零向量，这就是齐次解。为了更好地理解这个问题，我们看下图：



如果有很多方程组，其实可以写成 **AX=B** 的形式。因为它是单个方程组合而来的，因此之前得到的一些东西就可以拿过来直接应用了。比如，矩阵乘法 **AX=B** 中能不能解出 **X** 来呢？当满足 **r(A)=r([A,B])**的时候，虽然 A 不一定可逆，但一定能解出 **X** 来！

$$\begin{cases} Ax_1 = \beta_1 \\ Ax_2 = \beta_2 \\ \cdots \\ Ax_n = \beta_n \end{cases} \rightarrow AX = B$$

4 坐标系与坐标

经过以上的讨论，我们基本对于矩阵是变换的描述有了清晰的认识。接下来，让我们从空间的角度更深入的理解一下矩阵。之前用变换的角度理解矩阵，我们的矩阵可以不是满秩（非奇异）的方阵。下面，我们用空间的角度理解矩阵，在后面的叙述中，我们讨论的矩阵都是非奇异的！

顺便说一句：线性无关和不共线说的是一回事，只不过前者是代数描述，后者是几何描述。而维度的理解就是：描述一个现象需要几个各自独立的变量，这个现象就是几维的！

线性空间 V 的基不是唯一的，只要 n 个 n 维向量线性无关，就可以用来描述这个空间。那么线性空间中的两组基之间有什么联系呢？同一个向量在不同基下有不同的坐标，他们之间有什么联系呢？

设 $\vec{\alpha}_1,\vec{\alpha}_2,\vec{\alpha}_3\cdots\vec{\alpha}_n$ 和 $\vec{\beta}_1,\vec{\beta}_2,\vec{\beta}_3\cdots\vec{\beta}_n$ 是线性空间 V 中的两组基，$\vec{\beta}_1,\vec{\beta}_2,\vec{\beta}_3\cdots\vec{\beta}_n$ 可以用 $\vec{\alpha}_1,\vec{\alpha}_2,\vec{\alpha}_3\cdots\vec{\alpha}_n$ 线性表示出来

$$\begin{aligned} &\vec{\beta}_1 = m_{11}\vec{\alpha}_1 + m_{21}\vec{\alpha}_2 + \cdots m_{n1}\vec{\alpha}_n \\ &\vec{\beta}_2 = m_{12}\vec{\alpha}_1 + m_{22}\vec{\alpha}_2 + \cdots m_{n2}\vec{\alpha}_n \\ &\cdots \\ &\vec{\beta}_n = m_{1n}\vec{\alpha}_1 + m_{2n}\vec{\alpha}_2 + \cdots m_{nn}\vec{\alpha}_n \end{aligned}$$

写成矩阵的形式：

$$(\vec{\beta}_1,\vec{\beta}_2,\vec{\beta}_3\cdots\vec{\beta}_n) = (\vec{\alpha}_1,\vec{\alpha}_2,\vec{\alpha}_3\cdots\vec{\alpha}_n)\begin{bmatrix} m_{11}+m_{12}+\cdots m_{1n} \\ m_{21}+m_{22}+\cdots m_{2n} \\ \cdots \\ m_{n1}+m_{n2}+\cdots m_{nn} \end{bmatrix}$$

这样，我们就找到了旧的基 $\{\vec{\alpha}\}$ 到新的基 $\{\vec{\beta}\}$ 的过度，[M]称为过度矩阵。

这里 M 是可逆的。可逆的意思是可以变过来也可以变回去，不可逆就是只能朝着一个方向变换。从条件的角度来说：可逆是充要条件，不可逆是充分条件。我们可以看出来：**右乘一个方阵就是将一组基换到另一组基！（请注意不加说明的时候，我们讨论的是列向量）**

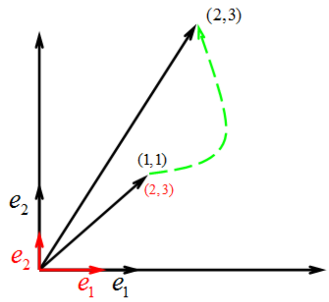
又一次矩阵乘法表示了运动，只不过这次不是左乘，不是变换对象，而是右乘，是变换坐标。理解方式不同，但是结果是一样的。于是我们得出这样的认识“对象的变换等价于坐

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。”而这个相对体现在左乘和右乘上。

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。



对于上图，黑色的代表变换的作用效果，矩阵把（1,1）变换到（2,3）。红色的代表变动坐标系后的效果，变动坐标系后原来的（1,1）点在新的基下坐标就变成了（2,3）了。

综合以上的讨论，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！对于坐标系的变换这里只是介绍了一个简单的例子，矩阵对坐标系施加的变换还有很多需要解释的内容，更多的内容在《神奇的矩阵第二季中》介绍。

上面的讨论我们得到了线性空间两组基的联系，这是矩阵右乘，下面我们讨论两组基对应坐标的联系。

接着上面的例子：若基$\{\vec{\alpha}\}$对应的坐标是$\{x\}$ ，基$\{\vec{\beta}\}$对应的坐标是$\{y\}$ 则有

$$Y = M^{-1}X$$

其中 M 就是我们刚刚说的过度矩阵，我们可以看出**左乘一个方阵就是将一组基下的坐标变换到另一组基下的坐标！**

同时结合刚才的讨论，我们注意到：对于 X 和 Y 中的单个向量，有 **By=Ax** ，而对于矩阵有 **BY=AX**。我们很自然的对于这个等式的理解是这样的：

“有一个向量，它在坐标系 A 的度量下得到的度量结果向量为 **x**，那么它在坐标系 B（单位阵）的度量下，这个向量的度量结果是 **y**。”

在 A 为坐标系的意义下，如果把 A 放在一个向量 **x** 的前面，形成 A**x** 的样式，我们可以认为**这是对向量 x 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 A 中度量，得到的度量结果可以表达为 **x**。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 A 放在前面，让你明白，这是该向量在坐标系 A 中度量的结果。”

同样，我们可以理解理解 B 和 **y** 相乘的关系。而对于等式 **By=Ax** 的意思就是说：

“在 A 坐标系里量出来的向量 **x**，跟在 B 坐标系里量出来的向量 **y**，其实根本就是一个向量啊！”

同理对于 BY=AX 的理解就是向量组{**x**}和{**y**}。我们将得到惊人的结论：**这哪里是什么左乘法计算，根本就是身份识别嘛！**

我希望你务必理解这一点，因为这是本篇的关键。

矩阵乘法不满足交换律，从这里我们也可以做出一个解释，因为他们的的意义不同：一个是变换基，一个是变换坐标，当然不满足交换律，更不满足消去率！

好了，到此，我们基本呢说清楚了矩阵最核心的概念，下面，我们针对一些比较特殊的现象，来更深入理解一下线性代数和矩阵！

在接下来的讨论中，我们将从应用的角度理解矩阵乘法、矩阵等价、矩阵的秩、矩阵相似、特征值、行列式的几何意义等问题，如果你是一口气看到这里的，建议你休息一下，然后我们再进行进一步的讨论。

矩阵连乘表示变换的叠加。为了说明问题，我们从两个特殊的矩阵入手，旋转矩阵和对角矩阵，然后我们讨论一般矩阵的连乘问题。

5 特殊的矩阵

旋转矩阵乘幂的几何及物理解释：大家已经知道，一个矩阵乘以一个向量，一般将会对向量的几何图形进行旋转和伸缩变化。如果仅仅进行伸缩，就是特征值和特征向量的问题，我们在后面会深入讨论它的意义。如果仅仅是旋转，就是我们下面讨论的问题：在教科书中，我们常见的一个例子就是旋转矩阵，旋转矩阵只对向量进行旋转变化而没有伸缩变化。例如二阶旋转矩阵 A：

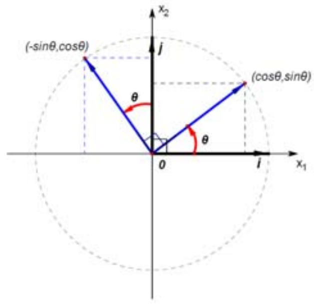
$$A=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$$

显然，二阶的所有的某一特定角度的旋转矩阵都分布在单位圆上。对 θ 取几个不同的弧度，就会得到几个旋转矩阵：

θ	0	$\pi/4$	$\pi/2$	$3\pi/4$
$A=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}\sqrt{2} & -\sqrt{2}\\ \sqrt{2} & \sqrt{2}\end{pmatrix}$	$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$	$\frac{1}{2}\begin{pmatrix}-\sqrt{2} & -\sqrt{2}\\ \sqrt{2} & -\sqrt{2}\end{pmatrix}$

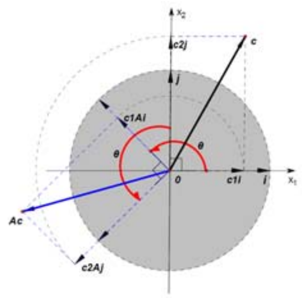
比如单位矩阵 $A_1=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$，实际上就是对一个向量 **x** 旋转的角度是 0，也就是 **Ax=x**；而矩阵 $A_2=\frac{1}{2}\begin{pmatrix}\sqrt{2} & -\sqrt{2}\\ \sqrt{2} & \sqrt{2}\end{pmatrix}$就是对一个向量 **x** 旋转的角度是 $\frac{\pi}{4}$

是不是这样的呢？首先看一下旋转矩阵 **A** 对单位向量 **i 和 j** 的作用效果。



对于任意向量 **x**，我们知道，可以分解为单位向量的线性表示

$$\vec{x}=\begin{pmatrix}c_1\\ c_2\end{pmatrix}=c_1\begin{pmatrix}1\\ 0\end{pmatrix}+c_2\begin{pmatrix}0\\ 1\end{pmatrix}=c_1 i+c_2 j$$



由上表格和图形知道，当 θ=0 时，旋转矩阵的行向量就是 **x** 和 **y** 轴的单位向量和 **i**、**j**，随着 θ 角度的增大，单位向量在单位圆上同时进行顺时针等角度旋转。

如果我们从列向量的角度看旋转矩阵，又会得到什么图形呢？显然，从上图看出列向量在随着 θ 的角度的增加而逆时针旋转，刚好与行向量同时反向旋转！这正是正交矩阵的特征。也就是说，旋转矩阵就是正交矩阵。

从而得出旋转矩阵的逆矩阵是它的转置矩阵:

$$MM^{-1} = MM^{T} = E$$

一个矩阵是旋转矩阵，当且仅当它是正交矩阵并且它的行列式是单位一。

旋转矩阵是正交矩阵，如果它的列向量形成 R^n 的一个正交基，就是说在任何两个列向量之间的标量积是零(正交性)而每个列向量的大小是单位一(单位向量)。

正交矩阵的行列式是 ±1；如果行列式是 −1，则它包含了一个反射而不是真旋转矩阵。

对于旋转矩阵乘幂，我们可以知道，就是一直旋转，乘了几次就是旋转了几次。当好多个旋转矩阵（可以不同）连乘时，我们就能理解了，这是把一个向量沿着多个方向旋转的叠加啊！

对于旋转矩阵我就说这么多，感兴趣的读者可以看一下网上的介绍。

下面我们来对角矩阵乘幂的几何及物理解释，对于矩阵

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \cdots & & & \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

相信这个不用我解释大家都知道：这个矩阵的作用是对向量进行放大（缩小），即不断的放大（缩小）数值而不改变向量的方向，这里我就不给出例子而直接给出结论。

对于对角矩阵乘幂，我们可以知道，就是一直放大（缩小），乘了几次就是放大（缩小）了几次。当好多个对角矩阵（可以不同）连乘时，我们就能理解了，这是把一个向量放大（缩小）叠加啊！

对于一般矩阵乘幂的几何及物理解释，我们有了以上的基础，讨论起来就方便多了：如果把两个矩阵看作等同的，那么**AB**的结果是把两个线性变换进行了叠加或复合。

对于矩阵乘法，主要是考察一个矩阵对另一个矩阵所起的变换作用。其作用的矩阵看作是动作矩阵，被作用的矩阵可以看作是由行或列向量构成的几何图形。这个理解我们最开始就讨论过，也就是矩阵描述一个变换的问题。

同样，如果一连串的矩阵相乘，就是多次变换的叠加么。而矩阵左乘无非是把一个向量或一组向量（即另一个矩阵）进行伸缩或旋转。**乘积的效果就是多个伸缩和旋转的叠加！**

如在**S=ABCDEF中**，会把所有的矩阵线性变化的作用力传递并积累下去，最终得到一个和作用力**S**。工业上的例子就是机器人的手臂，机械臂上的每个关节就是一个矩阵（比如可以是一个旋转矩阵），机械臂末端的位置或动作是所有关节运动的综合效果。这个综合效果可以用旋转矩阵的乘法得到。机械臂末端的位置或动作是所有关节运动的综合效果。这个综合效果可以用旋转矩阵的乘法得到。

注意，我们讨论的都是左乘矩阵，这和我们在最开始讨论矩阵是描述运动（变换）的那部分分析是一致的，只不过这里我们主要说的是多次的变换，即好多个矩阵相乘而不仅仅限于两个矩阵的乘法。同时，我们这里从两个特殊的矩阵（旋转矩阵和对角矩阵）乘法入手，然后讨论一般现象。

为什么把这两个矩阵单独拿出来讨论呢？是因为他们有着良好的性质：**旋转矩阵把一个矩阵乘向量的作用变成对向量的旋转，对角矩阵乘向量的作用变成对向量的伸缩。**后面我们讨论在讨论相似的时候会再次见到他们熟悉的身影，那时他们强大的作用就会凸现出来！

至此，我相信你对矩阵乘法的理解已经相当充分了，在接下来的讨论中，我们就开始转向其他问题，你可以在此做一个短暂休息了，我们还有一段路要走，但是离终点已经不远了。

6什么是等价与相似

什么是矩阵等价呢？从定义角度就是：存在可逆矩阵 P 和 Q 满足 $PAQ = B$ 则我们说 A 和 B 是等价的。

为了说明矩阵等价的意义，我们先来说明 M×N 矩阵的几何意义：假设 dim=max{M, N}，从在开始的讨论中我们知道矩阵在空间中，那么现在有一个 M×N 的矩阵 A，它就在维度为 dim 的空间中。你可以这样理解：这就像在一个三维空间，一条线，一个面一样。我们也可以把 M×N 的矩阵 A 叫做维度为 dim 的空间在子空间。好，接受了这个之后，让我们来讨论矩阵等价的意义。我们知道，对于可逆变换的意义就是变换过去还可以变换回来，无论怎么变换，只要是可逆的，变换是不改变矩阵维度的，于是我们得到等价的矩阵 A 和 B 的维度相同。用数学来描述就是他们的秩 r(A)相同。那么，两个同型矩阵维度相同说明什么呢？从空间的角度来说，**就是他们是同一种类型的子空间！你可以这样理解：如果两个矩阵等价，一个矩阵能表示一个线，另一个矩阵也表示线；一个矩阵能表示一个面，另一个矩阵也表示面……**。我们在给矩阵分类的时候就可以根据它的类型（秩）进行划分。怎么理解 P 和 Q 的作用呢？P 和 Q 可以看成是可逆变换， 变换的作用就是为了把元素之间相互干扰的矩阵化成能一眼看出维数的简单矩阵，看看 M×N 的矩阵所代表的子空间是什么维数的。当然，如果将来你读了研究生，学习了奇异值分解，对矩阵 P 和 Q 还会有更深的理解，奇异值分解在工程中有很重要要的应用，一个简单的应用例子可以看吴军的《数学之美》。维数代表什么？在控制中，维数反映系统的可观测和可控性的量度。也不用太高深，就是看看这个空间多大，这个空间里面着装的变量够不够。希望你注意，这里的变换和接下来的相似变换是不同的，因为 P 和 Q 对应行初等变换和列初等变换的叠加，这里只能反映空间的维数。然而，在一些实际工程中，这些信息对我们了解一个系统已经足够用了。有没有更好的变换呢？有，就是我们接下来介绍的矩阵相似。

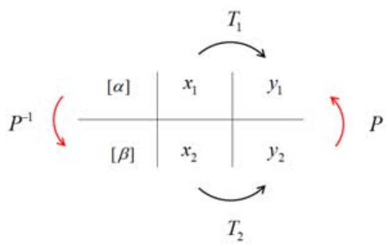
什么是矩阵相似呢？从定义角度就是：存在可逆矩阵 P 和 Q 满足 $B = P^{-1}AP$ 则我们说 A 和 B 是相似的。

首先，让我们来回顾一下之前得出的重要结论：对于同一个线性空间，可以用两组不同的基$[\alpha]$和基$[\beta]$来描述，他们之间的过渡关系是这样的：$[\beta] = [\alpha]P$，而对应坐标之间的过渡关系是这样的：$x_2 = P^{-1}x_1$。其中 P 是可逆矩阵，可逆的意义是我们能变换过去也要能变换回来，这一点很重要。

我们知道，对于一个线性变换，只要你选定一组基，那么就可以用一个矩阵 T_1 来描述这个线性变换。换一组基，就得到另一个不同的矩阵 T_2（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）。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。

具体来说，有一个线性变换 $x_1 \to y_1$，我们选择基$[\alpha]$来描述，对应矩阵是 T_1；同样的道理，我们选择基$[\beta]$来描述，就是 $x_2 \to y_2$，对应矩阵是 T_2；我们知道基$[\alpha]$和基$[\beta]$是有联系的，那么他们之间的变换 T_1 和 T_2 有没有联系呢？

现在我们做这样一件事：把 x_1 变到 x_2，即 $x_2 = P^{-1}x_1$。然后乘以变换矩阵 T_2 得到 y_2。然后再进行逆一次逆变换，将 y_2 变成 y_1，即 $Py_2 = y_1$。我们发现 $T_1x_1 = y_1$ 竟然和 $PT_2P^{-1}x_1 = y_1$ 的效果是一样的！即 $T_2 = P^{-1}T_1P$ 具体的请看下图：



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基的描述矩阵。这就是相似变换的几何意义。**

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。就像信号处理（积分变换）中将信号（函数）进行拉氏变换，在复数域处理完了之后又进行拉式反变换，回到实数域一样。信号处理中是主要是为了将复杂的卷积运算变成乘法运算。其实这样的变换还有好多，有兴趣可以看积分变换的教材。

为什么这样做呢？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。至于什么样的矩阵是"美"的,什么样的是"丑"的，我们说对角阵是美的。在线性代数中，我们会看到，如果把复杂的矩阵变换成对角矩阵，作用完了之后再变换回来，这对求解矩阵的 n 次幂很有用处！而学了矩阵论之后你会发现，矩阵的 n 次幂是工程中非常常见的运算。这里顺便说一句，将矩阵对角化在控制工程和机械振动领域具有将复杂方程解耦的妙用！**总而言之，相似变换是为了简化计算！**

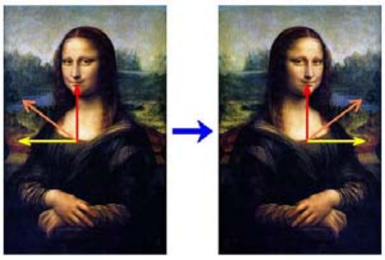
7 矩阵的相似对角化

接下来，我们就来探讨这个特殊的相似现象——矩阵对角化问题。

通过前面的讨论，我们知道：**矩阵乘法是把任意一个向量变成另一个方向或长度都大多不同的新向量。在这个变换的过程中，向量主要发生旋转、伸缩的变化。**

这时我们可以想一个问题：有没有变换只改变长度，不改变方向呢？如果存在矩阵对某一个向量或某些向量只发生伸缩变换，不对这些向量产生旋转的效果，那么这些向量就称为这个矩阵的特征向量。

下面我们引用维基百科里面的一个例子：当蒙娜丽莎的图像左右翻转时，中间垂直的红色向量方向保持不变。而水平方向上黄色的向量的方向完全反转，因此它们都是左右翻转变换的特征向量。红色向量长度不变，其特征值为 1。黄色向量长度也不变但方向变了，其特征值为-1。橙色向量在翻转后和原来的向量不在同一条直线上，因此不是特征向量。



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 $A\mathbf{x}=\lambda\mathbf{x}$。你就恍然大悟了。看到了吗? $\lambda\mathbf{x}$ 是方阵 A 对向量 $\mathbf{x}$ 进行变换后的结果。但显然 $\lambda\mathbf{x}$ 和 $\mathbf{x}$ 的方向相同)。而且如果 $\mathbf{x}$ 是特征向量的话。$a\mathbf{x}$ 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，但是需要说明的是：这个向量族线性相关。

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$[1\ 0;0\ -1]*[a\ b]^T=[a\ -b]^T$。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

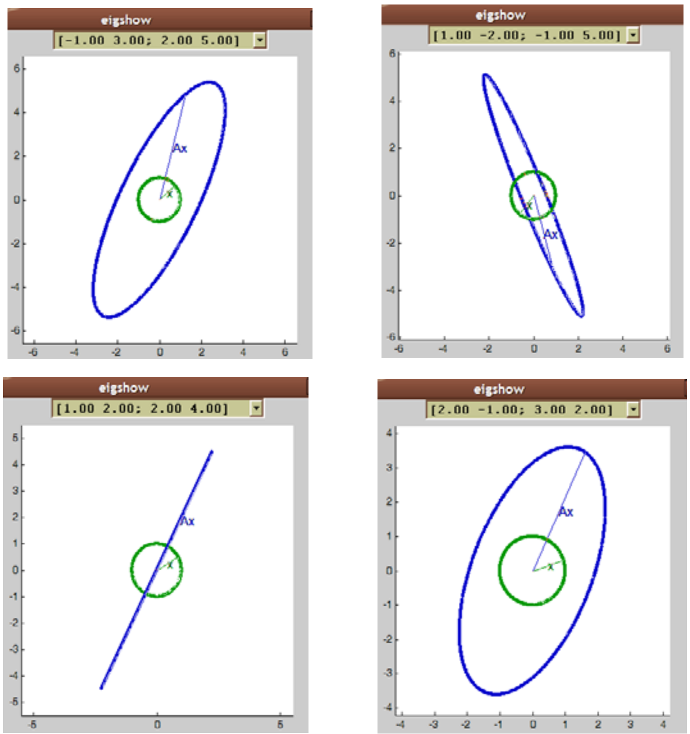
保持方向不变(记住这个变换是镜像对称变换。那镜子表面上(横轴上)的向量当然不会变化)。所以可以直接猜测其特征向量是 $[a\ 0]^T$(a 不为 0)。还有其他的吗?有。那就是纵轴上的向量。这时经过变换后。其方向反向。但仍在同一条轴上。所以也被认为是方向没有变化。

可以想一下，对于前面我们举的旋转矩阵的例子，除了零向量。没有其他向量可以在平面上旋转 30° 而不改变方向的。**所以这个逆时针旋转 30° 的变换（旋转矩阵）对应的矩阵没有实数特征值和实数特征向量，但是有复数特征特征值和特征向量。而且特征值和特征向量是成对出现的。将特征值写成指数形式，它的幅值代表伸长或者缩小的程度，相角代表 Ax 和 x 之间的夹角。**我会在讨论完实特征值之后给出一些例子来说明这个问题。

下面，让我们回到实特征值和实特征向量的范畴来继续讨论。

我们可以用 matlab 中的 eigshow()函数观察一下 **x** 和 **Ax** 之间的关系。在这里，我给出几个矩阵的变换结果，建议读者亲自试验一下，效果会更好。

下图是矩阵 $A_1=\begin{bmatrix}-1 & 3\\ 2 & 5\end{bmatrix}, A_2=\begin{bmatrix}1 & -2\\ -1 & 5\end{bmatrix}, A_3=\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}, A_4=\begin{bmatrix}2 & -1\\ 3 & 2\end{bmatrix}$ 特征向量的几何含义。



试想，对于矩阵 A 这个变换所在的一个 n 维空间，如果存在 n 个线性无关的特征向量，**我们就能用这 n 个特征向量作为基来表示这个空间的任意向量**，这样的计算就变得极其简单了！这样想：**那么一个大矩阵 A 对向量的作用就变成了一个数乘以特征向量了再求个和，复杂的矩阵运算就变成了数乘运算！**

简化计算——这就是我们要对角化的原因。你想啊，让你算一个矩阵乘以一个向量，你肯定说：“唉，不想算这个，计算量太大了”。但是要是有个人说“你就用 λ 乘以这个向量，和用矩阵算是一样的结果”。你肯定觉得这个很神奇！

下面的问题是能不能找到这样的“好”向量作为一组基？矩阵能不能对角化，为什么不能对角化，不能对角化说明了矩阵出了什么问题？对于实数范围内的特征值和特征向量，我们可以从几何角度来解释：我们知道，矩阵乘法是对向量的伸缩或者旋转。**如果矩阵作用了一个任意向量 x， Ax 和 x 之间总存在不是 0 或者 π 的夹角，那么就不能对角化。当然这太绝对了，是一种极端情况。一般情况是：x 在 n 维空间里面随便转啊转，出现了一小部分满足 Ax 和 x 共线的向量（即特征向量）。我们知道，如果这些不相关的向量的数量小于 n，**

那么就不能作为基来描述整个 n 维空间，矩阵 A 也就不能对角化了！

下面的问题是怎么找这样的“好”向量呢？相信大家都明白了，直接令 **Ax**=λ**x** 相等，能算出 n 个线性无关的向量就说明可以找到。算不出来或者算出来数目不够，那么好，我们找不到这样的基！具体的计算我就不再解释了，相信读者们都练习不知道多少遍了，这是所有线性代数考试必考的知识点，你算得肯定非常熟练了。

线性代数教材上还证明：**如果矩阵与一个对角阵相似，那么他们具有相同的特征多项式、特征值、秩、迹。**由于对角阵的这些量是容易求出来的，而这些值在工程中是非常重要而且有意义的。我们会在后面的讲述中介绍一些重要的应用。对角化的矩阵还可以简化 A 的某些计算，比如乘幂运算（乘幂运算在工程中非常有用），这一点在我们讲述对角阵的时候已经初现端倪。基于以上几点，我们说对角阵是美的，可对角化的矩阵也是美的！至于他的应用，可以说非常重要！但是由于专业性太强，就不展开介绍了。

特征向量的几何意义我们已经讨论过了，特征向量们作为一组基，把空间中的变换 A 简单化了。那么和它对应的特征值怎么理解呢？

从变换（运动）的角度，特征向量与特征值描绘了矩阵连乘（或者说持续作用）图景：特征向量在一个矩阵的反复作用下作伸缩运动，伸缩的幅度由特征值确定。特征值大于 1，所有属于此特征值的特征向量身形暴长；特征值大于 0 小于 1，特征向量身形猛缩；特征值小于 0，特征向量缩过了界，反方向到 0 点那边去了。

特征值反映了特征向量在变换时的伸缩倍数。对一个变换而言。特征向量指明变换的方向，而特征值反映的是变换的剧烈程度！结合我们之前在相似里面得到的结论，可以知道，一个矩阵，只要我们找到合适的坐标，它的全部信息就可以用坐标系和特征值表示！

这里蕴含了一个哲理：**我们从不同角度看问题，其难易程度是不一样的！而矩阵对角化为我们提供了一个很好的看问题的角度。**

为什么矩阵的特征值有时会被称为谱呢？其实这和特征值的物理意义有关，如果你学过线性系统理论或者振动分析类的课程，就会知道，一阶线性微分方程组对应的矩阵的特征值就是系统的固有频率。因为涉及研究生的课程，这部分内容会在《神奇的矩阵第二季》中详细介绍。我们知道，凡是叫谱的东西，都是可以分解的，比如光谱。那么这个矩阵可不可以分解呢？答案是肯定的：我们可以将一个具有良好性能的矩阵分解成多个作用的叠加。

在数学中有一个谱定律（Spectral theorem），也叫谱分解或者特征值分解。它的核心内容如下：一个矩阵 A 可表示为如下线性组合

$$A=\lambda_1 P_1+\lambda_2 P_2+\lambda_3 P_3\cdots\cdots$$

其中，$P_1, P_2 \cdots P_n$ 称为 A 的谱族，$\lambda_1, \lambda_2 \cdots \lambda_n$ 等表示特征值。

你也许会问，这个矩阵 $P_1, P_2 \cdots P_n$ 都是什么意思啊？它们和特征值有关系么？为了清晰明白的说明问题，这里做了两个假设：

1、矩阵 A 是可以对角化的。

2、所有特征值都是不同的。(这个不是必须的，只是为了结果比较清晰)

对于可对角化的矩阵 A 我们可以将它表示成这样

$$A=P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}P^{-1}$$

这里将 P 拆成列向量，P^{-1} 拆成行向量：$P=[p_1, p_2\cdots p_n]$，$P^{-1}=[q_1, q_2, \cdots q_n]^T$ 于是，就可以得到矩阵 $P_1=p_1q_1, P_2=p_2q_2\cdots P_n=p_nq_n$ 。

如果特征值相同，只要矩阵 A 可对角化，上面的分解也是可以进行的。把相同特征值的块合并在一起就可以，从它的应用角度来说这样处理是合理的。

从这里我们可以看出，一个变换（矩阵）可由它的所有特征值和谱族完全表示。**特征值和谱族的乘积就代表了它对矩阵 A 的贡献率——说的通俗一点就是能量（power）或者权重(weight)。这样，能量多的、权重大的部分当然重要，问题的主要矛盾就凸显出来了！**当然，也可以这样理解：把特征值看成坐标，谱族看成基。于是，一组基+一组坐标就表示了一个对象——矩阵 A。

一般第一次接触到谱分解这个东西都会有这样的疑问：为什么要分解？这只是一种数学上的游戏么？如果你接触过主成分分析(PCA)的话，你肯定会有感触的。通俗的来解释，主成分分析就是将一个复杂的东西(可以使函数，图像，音频等)分解成若干个成分，然后将其中主要的成分留下，将细枝末节忽略或者删掉，来简化对这个复杂东西的分析。举一个简单易懂而且你也学过的例子就是泰勒展开

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots$$

在一个离 x_0 很近的小区域内，我们发现阶次越高，前面的系数(也成为权系数)越小，而且变小的很“快”，同时基(多项式也可以作为基)也很“快”变小。因此，很多时候我们就为了分析简单，只取一阶，或者二阶近似代替 f(x)。**这样带来的好处就是分析变得简单而不影响 f(x)在展开点附近的主要特性！**如果你学过电工或者信号类的课程，你会知道将一个周期信号分解成傅里叶级数，然后取前几项代替这个周期信号是我们常用的手段，也是这个道理。同样的，我们将矩阵谱分解，然后就知道那些成分对矩阵的影响大，那些成分影响小了。去掉烦人的小量而不影响矩阵的主要特性，正是我们追求的目标。举一个通俗一点的例子，假如你考古发现一个破碎的花瓶，只需要把大块的碎片拼接起来，就能基本还原花瓶的原貌。类似的例子还有矩阵的奇异值分解等。这些会在《神奇的矩阵第二季》中作介绍，敬请期待。

8 矩阵的复数域对角化

(ps 这节考研不考，没兴趣可以略过)实特征值的问题我们已经说得差不多了，让我们来谈一谈复特征值。我们知道，**复数运算天生就带有旋转的结构。**前面我们说过旋转矩阵没有实数特征值和实数特征向量，（这一点你从图像上就能看到）但是有复数特征特征值和特征向量。**貌似旋转矩阵和复数有着某种天生的联系！**他们之间到底是什么关系呢？**如果将复特征值写成指数形式，它的幅值代表伸长或者缩小的程度，相角代表 Ax 和 x 之间的夹角。**

讨论前让我们来看一下这个矩阵 $R = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$，a、b 都不为零我们可以求得它的特征值为 $\lambda = a \pm bi = \sqrt{a^2+b^2}e^{\pm i\theta} = re^{\pm i\theta}$，则矩阵 R 可以写成一个旋转矩阵和对角阵的乘积：

$$R = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

1、对于矩阵

$$A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$$

我们求得它的特征值为 $\lambda = 0.8 \pm 0.6i$，特征向量为

$$\alpha_1=\begin{bmatrix}-2+4i\\5\end{bmatrix}\quad \alpha_2=\begin{bmatrix}-2-4i\\5\end{bmatrix}$$

我们取

$P_1=[\operatorname{Re}\alpha_1\ \ \operatorname{Im}\alpha_1]=\begin{bmatrix}-2&4\\5&0\end{bmatrix}$，计算 $P_1^{-1}AP_1=\begin{bmatrix}0.8&0.6\\-0.6&0.8\end{bmatrix}$。

$P_2=[\operatorname{Re}\alpha_2\ \ \operatorname{Im}\alpha_2]=\begin{bmatrix}-2&-4\\5&0\end{bmatrix}$，计算 $P_2^{-1}AP_2=\begin{bmatrix}0.8&-0.6\\0.6&0.8\end{bmatrix}$。

我们发现 A 相似于一个旋转矩阵，并且这个旋转矩阵中的数值和特征值的数值有着某种关系，我们写成一般的形式就是

$$\lambda=a\pm bi,\ \ P^{-1}AP=R$$

其中：$P=[\operatorname{Re}\alpha\quad \operatorname{Im}\alpha]$，

$$R=\begin{bmatrix}a&-b\\b&a\end{bmatrix}=\begin{bmatrix}r&0\\0&r\end{bmatrix}\begin{bmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{bmatrix}$$

2、我们来看一个三维的矩阵

$$A=\begin{bmatrix}0.8&-0.6&0\\0.6&0.8&0\\0&0&1.07\end{bmatrix}$$

它有特征值 $\lambda_{12}=0.8\pm0.6\quad \lambda_3=1.07$ ，特征向量分别为

$$[\alpha_1\quad \alpha_2\quad \alpha_3]=\begin{bmatrix}-\frac{\sqrt{2}}{2}i&\frac{\sqrt{2}}{2}i&0\\-\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}&0\\0&0&1\end{bmatrix}$$

经过上面的分析，我们知道，在对于任意的一个向量 **x**，在将其分解成 $x=x_1\operatorname{Re}\alpha+x_2\operatorname{Im}\alpha+x_3\alpha_3$，A 对在 $\operatorname{Re}\alpha$ 和 $\operatorname{Im}\alpha$ 所构成的那个平面上的任意向量都进行相同角度的旋转（也能伸缩，这取决于特征值的模是否为 1），而对于在 α_3 上的分量只进行伸缩变化，这不仅仅适用于三维空间试用，对于高维空间也同样满足！

同样的思路，我们来分析一下在复数领域可对角化的几何条件。当特征值是实数的时候，跟我们之前讨论的一样。**x 在 n 维空间里面随便转啊转，出现了一小部分满足 Ax 和 x 共线的向量（即特征向量）这些是实数向量。如果我们还能找到一个平面，A 对这个平面上的向量的作用效果是将他们旋转相同的角度，即使有伸缩，那么对这个平面上所有向量的伸缩是倍数是等大小的，即特征值的模长。这时，我们就能找到一对虚特征向量。同理，也能找到下一个平面对应的一对虚特征向量。如果不共线的实特征向量和虚特征向量的个数足够，那么 A 就能在复数域对角化。**

同时我们也看到，当旋转角度是 0° 或者 180° 的时候，此时虚部变成 0，也就是实数特征值了。所以说复数是实数的推广。

这里我们看到，通过复数乘法运算，我们可以把一个向量旋转，用旋转矩阵作用一个向

量，我们也可以将它旋转，我们得出这样一个结论：**旋转矩阵和复数运算在对向量的作用形式上具有相同的结构！**同时你也看到了，用复数表示旋转比用旋转矩阵表示旋转要简洁的多，这也就是为什么我们要引入虚数 i 的原因！

总结上面的讨论，我们认识到：**对角化的过程，实质上就是找特征向量的过程，选取特征向量作为基来表征矩阵 A**。如果一个矩阵在复数域不能对角化，我们还有办法把它化成比较优美的形式——Jordan 标准型。高等代数理论已经证明：**一个方阵在复数域一定可以化成 Jordan 标准型**。这一点有兴趣的同学可以看一下高等代数后或者矩阵论。复特征值在电路领域以及振动领域将发挥重要的作用，可以说，没有复数，就没有现代的电气化时代！

9 内积与相关

在线性代数中我们还提到内积的概念，对于内积的理解，有更广泛的含义。形式上可以理解成**对应的元素相乘再求和**（对于函数就是对应微分时的两个函数乘积再求无穷和（就是积分啦））。对于平稳过程，**内积可以衡量两个函数的线性相似程度（注意是线性！）**。内积值越大表示两个函数线性相似程度越大，内积为零表示完全不线性相似。而矩阵的转置运算，其实就是为了算内积而出现的。两个函数内积为零则两个函数正交,在三维空间中它们的夹角为 90 度,在三维以上不是这样的。

顺便说一句，内积和向量的夹角和相关有着密切的联系。我们知道，如果两个向量是共线的，那么他们的夹角就是 0° 或者 180°。此时向量夹角余弦值的绝对值是 1。反过来，通过计算向量之间的夹角余弦值，我们就能判断他们之间线性关系的强弱。它的定义是这样的：

$$\cos\theta = \frac{x \cdot y}{\|x\| \cdot \|y\|}$$

因为现实中一般出现都是类似 y=ax+b 的这种关系，这就给上面式子的应用带来了麻烦，（在最开始就说了，我们需要的是 y=ax 的形式）比如这个问题：

假设五个国家的国民生产总值分别是 1、2、3、5、8，又假设这五个国家的贫困比例分别是 11%、12%、13%、15%、18%。则我们现在有两个有序的包含 5 个元素的向量 x、y：$x = (1, 2, 3, 5, 8)$ ，$y = (0.11, 0.12, 0.13, 0.15, 0.18)$ 使用一般的方法来计算 $\cos\theta = 0.9208$ 。而上面的数据实际上是故意选择了一个完美的线性关系：$y = 0.01\,x + 0.10$ 。

怎么办呢？为了解决这个问题，我们把数据居中（x 中数据减去 $E(x) = 3.8$ ，y 中数据减去 $E(y) = 0.138$）后得到：$x = (-2.8, -1.8, -0.8, 1.2, 4.2)$、 $y = (-0.028, -0.018, -0.008, 0.012, 0.042)$，由此得到了预期结果：$\cos\theta = 1$ 。写出公式来就是

$$\cos\theta = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}} = \frac{\mathrm{cov}(x, y)}{\sqrt{Dx} \cdot \sqrt{Dy}}$$

看着熟悉吧，是不是和概率论中的相关系数很像啊，其实他们的原理是一样的！它还有一个名字：皮尔逊相关系数。而真正重要的是上面那个简洁而伟大的表达式，它有着广泛而重要的应用：用在概率论中，我们可以计算随机变量的相关系数，来判断两个随机变量是否相关。用在数字信号处理中，我们也可以计算两组采样值的相关系数来提取信号或者去除白噪声。在气象学中，可以判断两个不同气象要素或同一气象要素在不同时间和空间的相互关联。在经济学科中，可以用相关判断广告投入与月均销售额之间的关系。在医学中，可以用相关判断体内物理化学参数和疾病之间的关系。在遗传学中，可以判断群体中不同性状的育种值之间的关系。大名鼎鼎的最小二乘法也要利用相关来判断能否很好地拟合。用在互联网的文本搜索和文章分类中，我们可以通过计算相关系数（其实只要计算内积就可以）来快速解决，这个问题可以看吴军的《数学之美》这本书，写的挺详细的。相关和内积还有好多应用，这里我就不再赘余了。期待你慢慢去发现。

10 优美的二次型

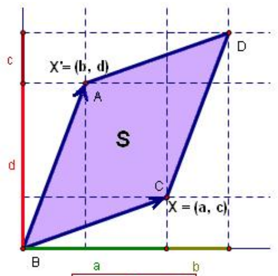
我们知道，**矩阵是由元素和元素的位置这两方面组成的，主对角线元素代表自身和自身的关系，其他位置代表相互关系。**在实际应用中，我们很多情形都是对称的相互关系，比如两个物体之间力的关系，作用力和反作用力相等。这样一个系统的力学关系矩阵就是对称的，比如振动矩阵，电路矩阵等。在数学上，这种关系用二次型来描述。二次型的理论在物理学、几何学、概率论等学科中都已得到了广泛的应用。由于每个二次型都可以经过线性替换变成若干个平方和的形式，对于矩阵来说，就是每个对称矩阵都合同于一个对角矩阵，后者称为一个标准形。二次型的几何意义是二次曲面，比如三维中的椭球和双曲抛物面等。而进行正交变换的意义就是将其在不改变几何形状的基础上（因为标准正交变换实质上是对坐标系只进行旋转变换），将坐标换到好位置，所谓的好位置就是让其几何中心在坐标原点，而且它还关于坐标轴对称。因为**二次型矩阵具有对称性，所以二次型一定可以化成对角矩阵**。因此，可以利用二次型的线性变换化简二次曲面。当然，它的用途还很多，而且很强大。

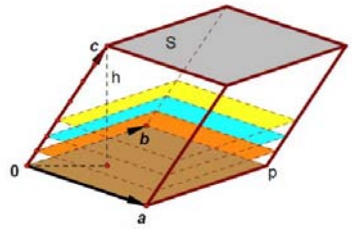
对于二次型的正定问题，它有惯性定理决定。它还有一个很好的应用，就是用在微积分的多元函数求极值问题。我们知道，对于一个多元函数，我们可以用泰勒公式在驻点展开到二次项，在很小的邻域内就能反映函数的极值性质。对于极值，其一阶偏导为零，同时将常数项移到等式左边，只剩下二阶项。如果这个二次多项式是正定的，则函数有极大值，如果是负定的就有极小值。因此，才有微积分里面用二阶导数的关系式来判断一元函数和多元函数的极值判别式！这里还涉及一个合同的概念，其实合同主要是为了二次型而存在的。对于二次型的矩阵表示来说，做一次非退化的线性替换相当于将二次型的矩阵变为一个与其合同的矩阵。二次型的更多介绍会在《神奇的矩阵第二季》中分享。

11 行列式的意义

还留下一个行列式的问题。行列式的几何意义是什么呢？概括说来有两个解释：一个解释是行列式就是行列式中的行或列向量所构成的“平行多面体”的**有向面积或有向体积**；另一个解释是矩阵 A 的行列式 detA 就是线性变换 A 下的**图形面积或体积的伸缩因子**。

我们首先来看行列式表示体积的解释，以二阶行列式和三阶行列式的为例：





$$V = (\vec{a} \times \vec{b} \cdot \vec{c}) = \left|[\vec{a}, \vec{b}, \vec{c}]^T\right|$$

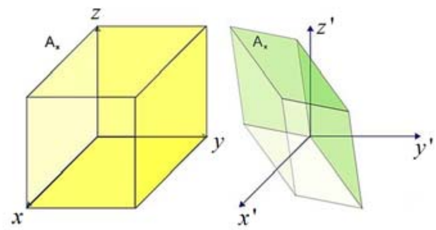
对于高维的，其意义也是类似的。行列式的加减，其实就是体积之间的加减。行列式为零，就表示在这个 n 维空间中这 n 个向量不能构成一个“体”，但是可以构成“点”或者“面”。可是怎么理解这种事情呢？我们不妨换一个角度：对于一个对角阵，它的行列式的几何意义比较好理解：对于二维平面就是长×宽；从三维空间就是长×宽×高；对于高维度的空间也是一样。那么好，如果行列式对应的矩阵可以对角化，也就是说这个矩阵和一个对角阵相似，问题就好理解了，还记得这个公式么

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$$

也就是先将 A 对角化，将坐标之间的耦合关系解除。我们知道相似变换就是把线性空间里面的对象换了一组基来表示而已。这时复杂的几何体就变成一个规则的“空间几何体”，有点类似长方体。然后取行列式的几何意义就清楚了，**它真的是表示空间“几何体”的体积！**

这里还存在一个问题，就是变换坐标的事情，我们知道，同一个空间体积可以用不同的基和坐标表示，我们可以分别算出这两组基表示下几何体的体积，其计算出来的值一般是不一样的。那么，两个体积之间有什么关系呢？

$$\begin{aligned} x' &= a_1x + b_1y + c_1z \\ y' &= a_2x + b_2y + c_2z \\ z' &= a_3x + b_3y + c_3z \end{aligned}$$



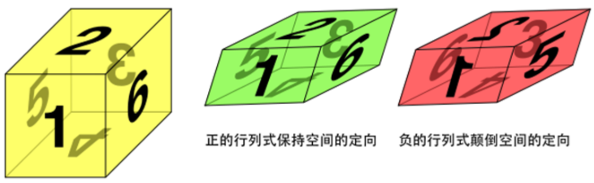
答案是两个体积之间相差的是一个坐标变换矩阵的行列式！无论是线性变换，还是非线性变换（比如微积分中的直角坐标和极坐标之间的变换），两个体积之间相差的都是一个行列式的关系（微积分中相差的是雅克比行列式）。也就是第二种解释的伸缩因子。

这里顺便简单说一下雅克比行列式吧，在微积分中，我们在进行 $(x,y)\to(u,v)$ 的坐标变换时，我们可以得到全微分 dx,dy 和 du,dv 之间的关系，写成矩阵形式就是

$$\begin{cases} dx = \frac{\partial x}{\partial u}du + \frac{\partial x}{\partial v}dv \\ dy = \frac{\partial y}{\partial u}du + \frac{\partial y}{\partial v}dv \end{cases} \quad \to \quad \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \cdot \begin{bmatrix} du \\ dv \end{bmatrix} = J \cdot \begin{bmatrix} du \\ dv \end{bmatrix}$$

于是，可以得到体积微元之间的关系就是 $\mathrm{dxdy} = \det(J)\,\mathrm{dudv}$，其中 $\det(J)$ 就是我们熟悉的雅克比行列式！对于直角坐标和极坐标之间有 $\mathrm{dxdy} = \det(J)\,\mathrm{drd}\theta = r\mathrm{drd}\theta$，对于三维的关系是 $\mathrm{dxdy}dz = \det(J)\,\mathrm{drd}\theta d\varphi = r^2\sin\varphi\,\mathrm{drd}\theta d\varphi$。

以上二维和三维行列式的例子中，行列式被解释为向量形成的图形的面积或体积。面积或体积的定义是恒正的，而行列式是有正有负的，因此需要引入有向面积和有向体积的概念。负的面积或体积在物理学中可能难以理解，但在数学中，它们和有向角的概念类似，都是对空间镜面对称特性的一种刻画。如果行列式表示的是线性变换对体积的影响，那么行列式的正负就表示了空间的定向。



如上图中，左边的黄色骰子（可以看成有单位的有向体积的物体）在经过了线性变换后变成中间绿色的平行六边形，这时行列式为正，两者是同定向的，可以通过旋转和拉伸从一个变成另一个。而骰子和右边的红色平行六边形之间也是通过线性变换得到的，但是无论怎样旋转和拉伸，都无法使一个变成另一个，一定要通过镜面反射才行。这时两者之间的线性变换的行列式是负的。可以看出，线性变换可以分为两类，一类对应着正的行列式，保持空间的定向不变，另一类对应负的行列式，颠倒空间的定向。

到此，线性代数学习过程中的大部分问题都已经解决了，希望这篇文章对你的学习能有所帮助！