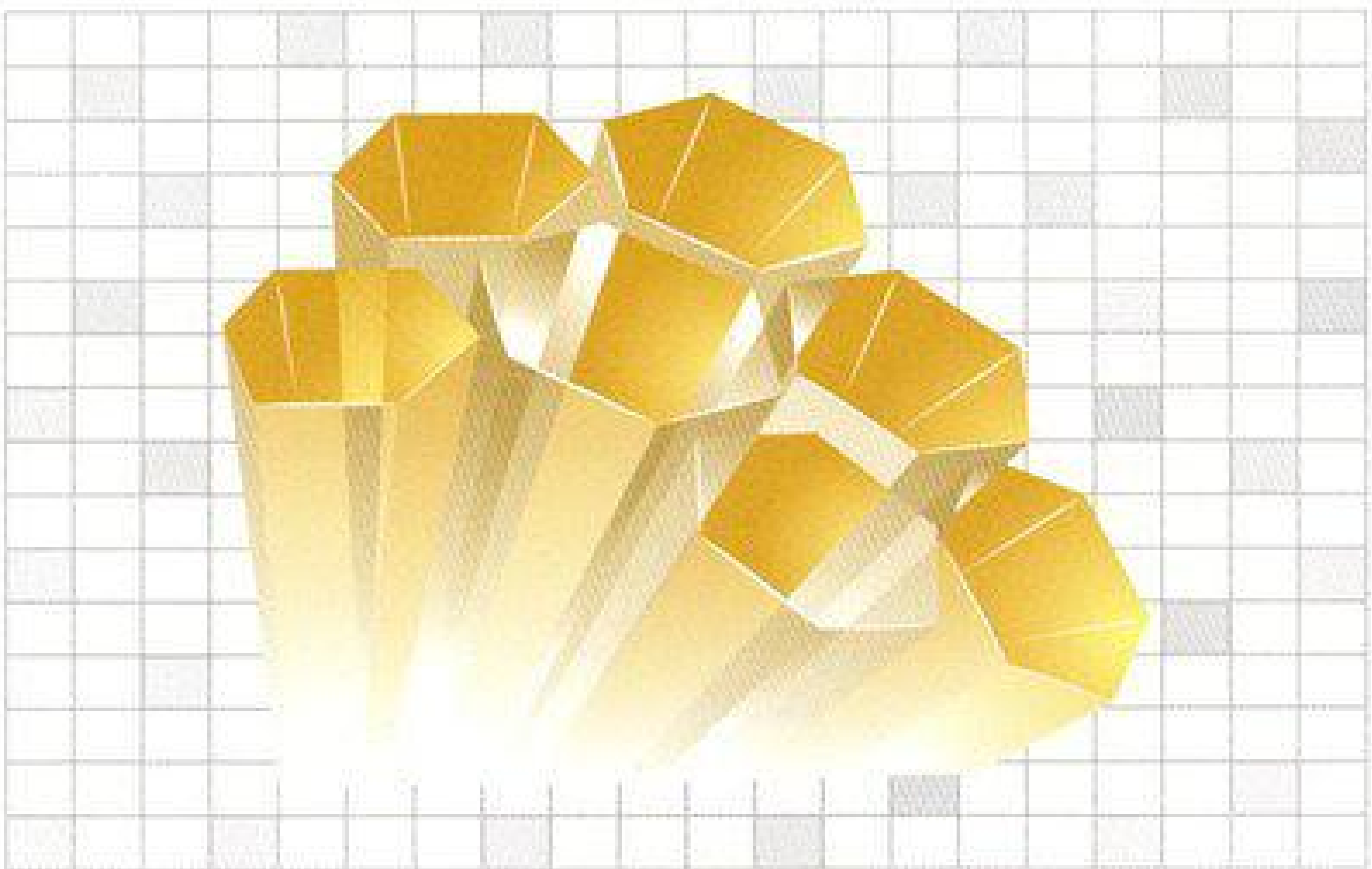
图解线性代数

信息环境下数学课程改革系列教材

线性代数的几何意义

The Geometric Meaning of Linear Algebra

任广千　谢聪　胡翠芳　编著



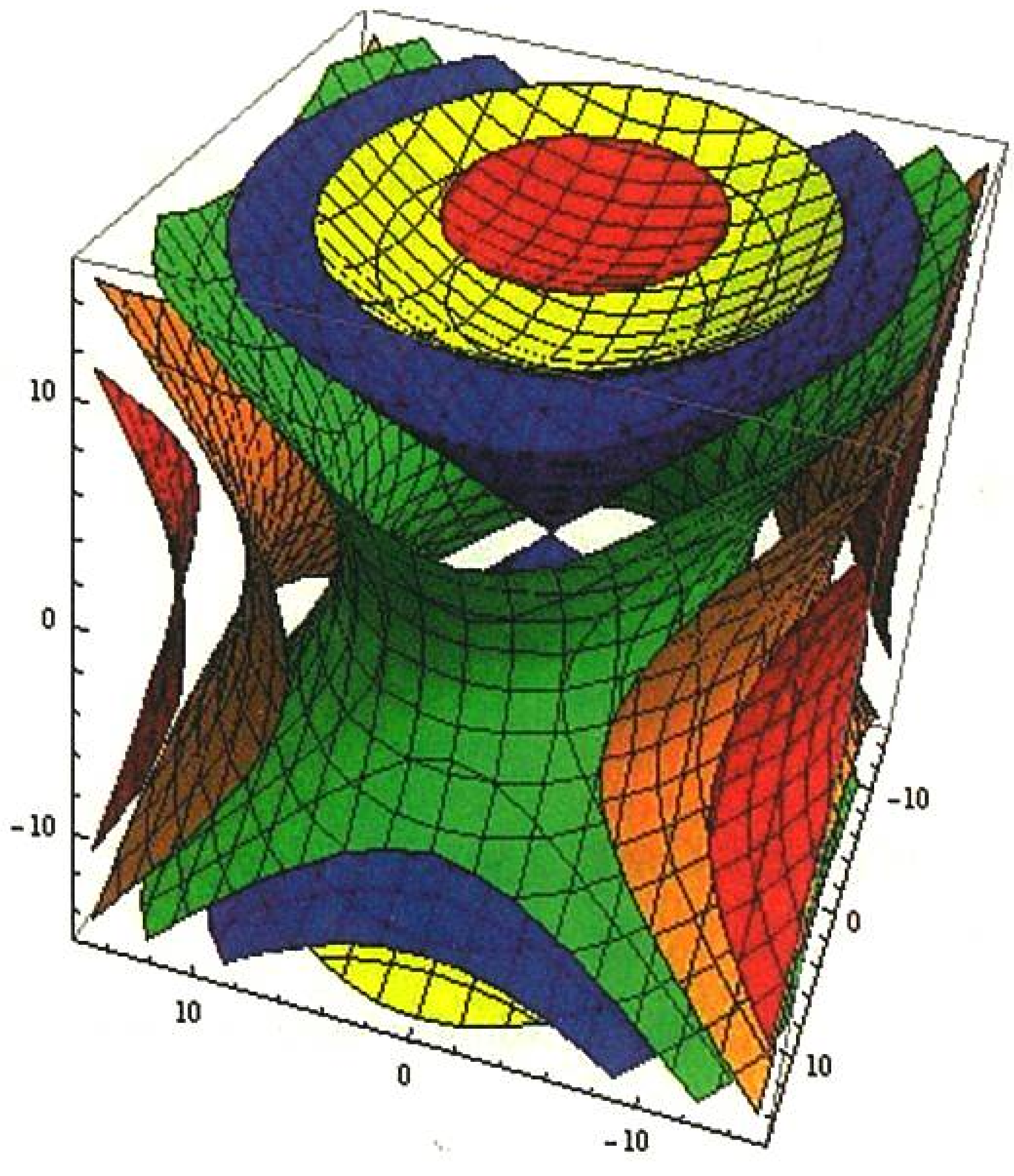
西安电子科技大学出版社
http://www.xduph.com

图解线性代数

信息环境下数学课程改革系列教材

线性代数的几何意义

任广千　谢　聪　胡翠芳　编著



西安电子科技大学出版社

内 容 简 介

本书是为热爱科学、热爱数学的思考者准备的！

本书是为喜欢把握科学脉搏探寻科学真谛的人准备的！

本书是为不愿被动地接受知识而奋力挣扎的学子准备的！

本书使用向量的概念对国内高校工科“线性代数”的课程内容进行了较全面的几何分析。从向量的几何意义开始，分别讲述了向量组、向量空间、行列式、矩阵、线性方程组和二次型的几何意义或几何解释，其中不乏重要概念的物理意义的解释。大量的代数概念及定理的几何意义的解释也可以使它成为您学习线性代数的参考手册。

本书大多为作者的原创，比如叉积的物理意义，克莱姆法则、雅可比矩阵、相似/合同矩阵、转置矩阵/对偶、矩阵乘积的行列式等系列概念的几何意义等，应用方面如使用矩阵分析的方法分析电子振荡器的工作原理等。

本书图文并茂，思路清晰、语言流畅，概念及定理解释得合理、自然，适合有一定线性代数基础的大学生阅读。由于本书是直接根据线性代数课程的要求进行解释的，同时具有通俗性、科普性，因而也适合初学者和自学者使用。

图书在版编目(CIP)数据

线性代数的几何意义/任广千，谢聪，胡翠芳编著. —西安：西安电子科技大学出版社，2015.7

ISBN 978-7-5606-3454-8

Ⅰ. ① 线… Ⅱ. ① 任… ② 谢… ③ 胡… Ⅲ. ① 线性代数—研究 Ⅳ. ① O151.2

中国版本图书馆 CIP 数据核字(2015)第 152412 号

策　　划　毛红兵

责任编辑　王　瑛　毛红兵

出版发行　西安电子科技大学出版社(西安市太白南路 2 号)

电　　话　(029)88242885　88201467　　邮　　编　710071

网　　址　www.xduph.com　　电子邮箱　xdupfxb001@163.com

经　　销　新华书店

印刷单位　深圳市精英印刷有限公司

版　　次　2015 年 7 月第 1 版　2015 年 7 月第 1 次印刷

开　　本　787 毫米×1092 毫米　1/16　印　张　17.875

字　　数　435 千字

定　　价　46.00 元

ISBN 978-7-5606-3454-8/O

XDUP 3746001-1

前 言

1. 为什么要给出线性代数的几何意义

作为一名工作十多年的电子工程师，作者想再提高自己的专业水平时，深感数学能力的重要。随便打开一篇论文或一部专著，满纸的向量、矩阵、微分方程词汇扑面而来。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？

在高校、科研院所或大型企业里工作，可能还有专家、教授级的导师指引着进行理论上的研究，但大部分的大学生一毕业就进入了各类中小企业，在这些企业里的研发工作，基本上都是低级模仿，想弄点理论研究常无从下手，感觉大学是白上了。找点资料看看吧，大多是数学公式满篇，解释抽象模糊、思维之断崖重重。茫然之中仿佛看到专业作者们高深莫测、神龙无尾、高傲自去的背影。

太失望了，太伤自尊了。扭头看看周围的同行，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。数学应用的层次多是高中水平。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，导致我们的企业科技创新能力的严重不足。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。

屌丝的声明:

数学绝顶高手和天才们不在俺说的范围之内，对我等来说，他们是极少数的一小撮的火星人，对他们只能顶礼膜拜，不敢评论。不过拜完之后有点小嘀咕：为何钱学森还讲中国没有大师呢？为什么数学总考一百分的天才也难以成为大师？

为啥没有在四年的大学阶段学好“线性代数”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。思来想去，得到几个原因：教材编的大多不好，老师教的大多乏味，学生大多有些偷懒。学生偷懒原因不只一个，但我觉得主要一个原因是他们学起来困难——因为他们大多不知道这些内容有啥用，概念为啥这么叫，定理为啥那样推，老师为啥像刘谦的魔术一样七推八导就证毕了——郁闷多了导致了无语的偷懒。

太多的为啥了。既然错不在学生那就是老师的问题了？其实老师也很委屈：教学大纲要求在几十个学时学习如此多的内容，不填鸭行吗？在如此短的时间内讲完就不错了，哪里还有时间给你释疑解惑——韩愈定义的传道授业解惑的师道中的解惑被迫取消了，自己悟道吧。殊不知，惑之不解则道难传而业无授也。

嘿，错也不全在老师那里。错在哪里？找来找去，似乎有一个大家都可以责备而少有抗议的地方，就是教材不够好，因为学生靠看教材自习的话也困难重重：工科教材太简单的话看起来不深入，理解很肤浅，似懂非懂；专业数学书确是详细，但还是太抽象，一个新概念接着一个新概念从天而降，也让人发晕。

有人说，线性代数是培养大学生的抽象思维用的，就是让你“发晕”的。还有人说，工科的线性代数课程要求低，只是让你掌握一个数学的计算工具罢了，你会做题通过考试就可以了，干吗思前想后的。

有点欺负人不是？俺学一门学科真心就是要掌握它的真谛嘛，不发晕才是真掌握，真掌握了才算是学会了抽象思维。再说了，会做题也不等于真正掌握了这个数学工具，只有在实际应用中对它建模才能应用，数学建模的过程正是由具体到抽象的过程，但是不发晕才能抽象啊——问题又回来了。

看来还是要真掌握真明白才是正道。真明白就是把自己的疑问一一解决，找书看去！

到附近的大学图书馆看看，哇塞，一行行、一列列的教材琳琅满目、浩如烟海。名字叫“线性代数”的教材足有一千多册。

随便打开一本《线性代数》看看，跟十八年前的教材内容几乎一样：简单、生硬。最多就是把一点解析几何塞进去，水油不溶，感觉不像一本书（线性代数和高等几何的理论框架是不同的，线性代数只讲向量及其向量集合的几何空间图形，而高等解析几何是向量（如法向量）和点集图形混为一谈，目的是坐标点的几何分析。请参考本书 5.14.2 节里面的对偶空间的分析内容），疑问还是没得到解决。再打开一本看看，内容还是那个内容，疑问还是那个疑问……

当浏览到第 500 本的时候（有点夸张哈），终于看到了我那个问题的答案了。长出一口气后我又陷入了郁闷之中。要知道，我至少有十打问题要解决呀，老天。

真是皇天不负有心人！眼前一亮，一本老外所编教材图文并茂的内容一下子照亮了俺那已黯然的心灵之窗（俺不想崇洋）。我兴奋得一阵眩晕。要知道，老外的教材大都是引入了当代科技的典型应用案例的，代表了本学科的国际新潮流，具有超强的问题杀伤力。

（注： 当然，国内也不乏优秀的线代教材，比如陈怀琛的《线性代数实践及 MATLAB 入门》，侧重工程实际和计算工具使用；李尚志的《线性代数》教材，理论讲解全面细腻，试图将几何代数熔于一炉，但仍稍嫌抽象，专业性强。另外，俺发现一个现象，现在的教材多不如十几年前的老教材，如张远达的《 线性代数原理》讲解就很牛。）

大学图书馆里的读者很多，朝气蓬勃的青春学子们在图书馆里做作业。我很羡慕他们这一代：在开放的图书馆里，学生们在书架前可以随意地浏览、挑选适合自己的纸质或电子版读物。要知道，当年我就读的大学图书馆是闭架的，每每借书要查半天小卡片，查完填好借书单交给工作人员大多得到两个结果：要么书被借完了，要么借的书不合适。而且还没有这么多的引进教材供参考，自学的效率大打折扣。

扯来扯去，千言万语汇成一句话：什么样的线性代数学习资料较好，较适合中国学生？我想，本子的物理尺寸要越薄越好，内容要越通俗易懂越好。

书本越薄，大家学习的信心越强：小样，这么点厚度还搞不定你，看，信心先有了。

如果只是容量精简了还不行，因为内容不全面，讲述不到位使得考试的时候受打击，工作中更受打击。如当年我学的《线性代数》课本是同济编的，内容是精简到家，千锤百炼，没一句废话，超薄。死记硬背，看似搞定了，实际是囫囵吞枣，似懂非懂。

如何通俗易懂还不能多说，我一直认为，加上几何意义或者物理意义啥的（再加上点当代的经典应用就更美了），一步到位搞定。

这就是本书的由来，也是本书的目标。

方向有了，具体如何编写呢？模仿一下科学大德牛顿的口气：

从线性代数书籍之浩瀚海洋的沙滩上（还没有更高的能力去远洋），用一双自己的眼睛，寻找到了一个个闪亮的小珍珠，一片片如玉的小彩贝，然后细细地擦拭、打磨，拂去沙尘，使它们重放光彩，用一根几何意义的锦丝，穿就了这本《线性代数的几何意义》的项链，献给热爱思考、痴迷于创造的人们。

呵呵，自不量力，终极目标而已，但意思还是有了。

2. 重要的几何直观意义

在学习中，一旦碰到较抽象难懂的新概念或定理，如何搞定？几个办法：一个是看推导过程，推导可以加强你相信它的信心并连通你原有的知识体系。如果推导把你弄昏了，最好弄懂它的几何意义或物理意义。

几何意义或者讲几何解释会和人们看到的平面和空间中物体几何外观联系起来，几何上说得通，物理上也就说得通，因为几何意义和物理意义本质上是一回事，新概念的几何意义通了，新概念就会和读者大脑中的既有经验或知识网络连通，一下子就“懂了”，满心欢喜的，原来是这么一回事。

有的哥们不太赞同几何意义和物理意义本质上是一回事。仔细想想吧，几何是描述现实空间的，而现实空间又是由物理规律所决定的；几何里蕴含着物理，物理决定着几何结构的存在。而且著名的进化论说过，物竞天择，适者生存。这个道理不光在说生物界，整个宇宙都符合进化论——只有适合物理规律的物体才能生存。直接说吧，n 亿年过去了，不符合物理规律的物质几何空间早就灭亡了。有点武断？提几个活生生的例子吧，杨振宁是物理学家，他搞的规范场是纯粹的物理理论，后来他突然发现，规范场正是微分几何学科里面的微分流形纤维丛上的联络。因此微分几何大师陈省身感慨地说“数学家和物理学家所研究的，只是一头大象的不同部分”。如果你再不信物理和几何是一回事，就想想爱因斯坦，想想他的相对论，相对论把宇宙包括牛顿的世界直接几何化了，用几何直接解释物理现象。

真理总是简单的和直观的，一位先贤说，不管多么复杂高深的数学理论，总有其直观的背景，不管多么繁难深奥的定理，其证明总有一个简单而直观的中心思想。几何图形能以其生动的直观形象给人留下深刻的印象。可以这样说，在数学中再没有别的什么东西，能比几何图形更容易进入人们的脑海了。

从宏观上看，一种数学理论(包括它的主要概念和方法)往往都有其直观的背景，它们或者是从对某些特殊的事例的观察分析中得到的，或者是直接从几何图形中看出的，或者是从已有的结果类比联想引来的，从几何直观上分析问题的能力，首先是指对于一种数学理论能“洞察其直观背景”。对于它是如何被发现的或如何形成的作出合理的解释或猜测。

一句话，一部皇皇巨著的理论特别是抽象的数学理论的核心常常可以从几何意义的角度得到解释。

从微观上看，国外的数学教育家波利亚曾经说:“一个长的证明常常取决于一个中心思想，而这个思想本身却是直观的和简单的”。因此，从几何直观上分析问题的能力，也包括找出证明中的那个关键的简单而直观的思想，也就是像希尔伯特所要求的，能透过概念的严格定义和实际证明中的推演细节，“描绘出证明方法的几何轮廓”。

大师庞加莱和阿达玛关于数学领域的发明创造的观点也认为，数学创造发明的关键在于选择数学观念间的“最佳组合”，从而形成数学上有用的新思想和新概念，而这种选择的基础是“美的直觉”。在这种美的直觉中，也就是在追求某种对称性、和谐性、统一性、简洁性和奇异性当中，以及在某种联想、猜想、假设及非逻辑思维中，几何直观具有头等重要的意义。

事实上，很多数学家都是先利用几何直观猜测到某些结果，然后才补出逻辑上的证明的。这正如我国拓扑学家张素诚先生所说的：对数学中的许多问题来说，“灵感”往往来自几何，表达的简洁靠代数，计算的精确靠分析。

革命年代的数学偶像华罗庚也曾吟过“数缺形时少直观，形少数时难入微；数形结合百般好，割裂分家万事休”。

嘿嘿，看看上面的数学上的历史牛人的观点，几何形象直观的意义何等重要。其实，大家都知道几何意义的重要，我们在小学和中学的学习阶段，老师也常常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？把形象扔得远远的，像瘟疫一样躲着它？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。既然你具体的活儿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。

几何意义和几何图形有点区别，我觉得几何意义应该是直观的、空间的、形象的、感性而可想象的含义，当然重点包括几何图形。如果只是些三角形、圆锥曲线、三维曲面类的几何图形远不能够让我们把握越来越抽象的数学，我们还应有点莫比乌斯带、克莱因瓶、不动点类的东东来激发我们的想象力。强调几何意义就是别忘了我们数学的想象力，具有了数学理论的想象力才会在我们的工程实践中运用她，体会她，发展她。

我觉得，抽象和形象是相辅相成，缺一不可的。套用马列主义辩证法的说法就是由形象而抽象，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巅峰。

3. 如何使用这本书

拼命阐述几何直观在数学学习中的重要意义，但这并不意味着可以否定逻辑推理论证的重要作用。实际上，单纯地依据直观而导致错误的数学例子真是数不胜数。概念或定理的几何直观解释，往往并不等同于原来的概念或定理。运用几何直观可以帮助我们猜想，但猜想并不能代替证明，只有经过一步步严格的逻辑论证以后，才算给出了证明。

形象（或直观）和抽象本来是一切科学的两面。只是近年来过分强调了抽象思维能力的训练而忽视了几何意义的解释。反过来，我们不能只强调几何意义而丢掉了计算和推导。因此建议读者：

- 从几何或物理意义入手，轻松而迅速理解和把握线性代数的基本概念和定理的几何本质，建立对线性代数的感性认识，具备理解复杂及抽象数学的基础。
- 然后，再回到现在学的抽象的线性代数的教材，通过适量的习题训练，短时间内构筑个人的线性代数知识体系的“向量空间”，巩固解决具体问题的动手能力。此时**，具体与抽象一体，理想与现实齐飞。**您，已经成为线性代数的高手和大牛。

我希望这本书对于高中或大学的初学者、学习过线性代数的、考研生、工程师、教师和任何喜欢数学想象的人都有帮助。具有一些线性代数基础的读者可能更喜欢这本书，因为书里一直没有回避公式和必要的推导。我不喜欢把她写成一个没有公式的大众科普读物，像霍金的科普书那样做只能让读者有一个仍然空洞的宇宙感性而不能真正深入数学的内部。

注：本文中，几何意义和几何解释的文字意思没有根本区别，一般对于数学概念的对应的几何图形而言称为几何意义，而对运算、变换的过程可对应几何图形的变化过程称为几何解释。

作者

2015 年 3 月

目 录

第1章 什么是线性代数

这一章的内容主要是想对线性代数的大的概念如线性函数、映射和线性变换以及线性代数的发展和应用作一简要介绍，本章的目的是让读者知道我们所学的线性代数的实质是什么，到底有什么用。

线性代数是代数学乃至整个数学的一个忒重要的学科，顾名思义，它是研究**线性**问题的**代数**理论。那么什么是代数呢？

1.1 “代数”的意义

“代数”的英文是 Algebra，源于阿拉伯语，其本意是“结合在一起”。就是说代数的功能是把许多看似不相关的事物“结合在一起”，也就是进行抽象。抽象的目的不是故弄玄虚，而是为了解决问题的方便，为了提高效率，把许多看似不相关的问题化归为一类问题。

抽象实际上并不神秘和高深，我们从小就学会了抽象:

蹒跚学步的时候，爸妈是这样通过举例子教会孩子数的概念是如何抽象出来的:

这是 1 个苹果，那是 1 个糖块，还有 1 个皮球……慢慢地我们忽略了物质上的差别，明白了“1”这个数量的含义，并及时地应用上了：“妈妈我要 1 个冰激淋！”

幼儿园及小学时老师也是这样教会我们数的加法运算法则是如何抽象出来的：

2 个苹果加上 3 个苹果是 5 个苹果;

2 个糖块加上 3 个糖块是 5 个糖块;

……

2 加上 3 就等于 5，用符号表示就是 2+3=5。

这样我们进一步地忽略了相加物体的大小、长短及原料的差别，只关心数量叠加的运算法则。

初中的时候，老师又进一步地教会了我们数及运算法则的进一步抽象：

$(1+2)^2=1^2+2(1\times2)+2^2$；

$(3+4)^2=3^2+2(3\times4)+4^2$；

……

用字母代替数值，得到完全平方公式：$(a+b)^2=a^2+2ab+b^2$。

好了，抽象又进了一步：不关心具体数值的运算，只关心它们的运算规律。到了这时候，我们开始学习一门叫代数的数学课，代数代数就是用字母代替数进行运算。从某种意义上来说，代数就是把算术推广到比具体的数更抽象的对象（运算规则）上面去。

抽象还可以进一步：高中时开始知道，公式$(a+b)^2=a^2+2ab+b^2$中的字母不仅可以代表数，还可以同时代表方向——也就是可以代表向量；并将其中的乘积a^2、ab解释为向量

内积，公式仍然成立。

画出有向线段来表示公式中的向量（见图 1-1）：$\overrightarrow{OA}=\boldsymbol{a},\overrightarrow{AB}=\boldsymbol{b}$，则 $\overrightarrow{OB}=\boldsymbol{c}=\boldsymbol{a}+\boldsymbol{b}$。

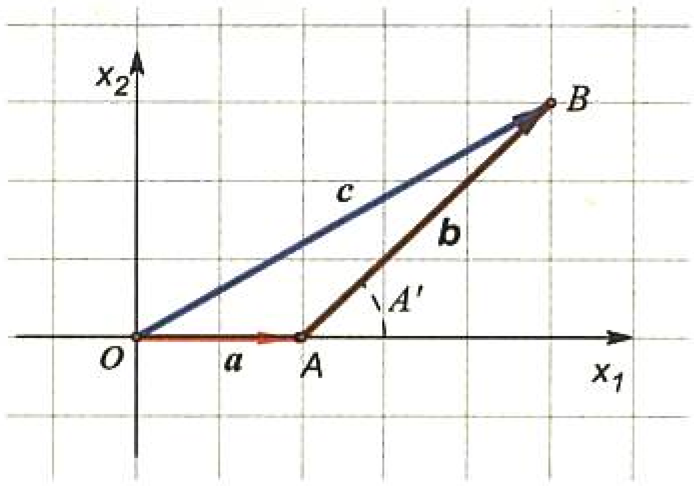


图 1-1　向量平方公式的几何解释

所以，向量的完全平方公式 $(\boldsymbol{a}+\boldsymbol{b})^2=\boldsymbol{a}^2+2\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{b}^2$ 的几何解释就是

$$|OB|^2=|OA|^2+|AB|^2+2|OA||AB|\cos A'$$

这就是余弦定理，当 A' 是直角时就是勾股定理（注：向量内积运算参见第 2 章，其中角 A' 是向量 $\boldsymbol{a}$ 转动到向量 $\boldsymbol{b}$ 时的夹角，是三角形内角 A 的外角：$A'=\pi-A$）。将数与向量混为一谈，立刻从简单的完全平方公式得到勾股定理和余弦定理，数学的抽象威力由此可见一斑啊！

另外，抽象还可以沿着另一个途径进行下去。恒等式 $(a+b)^2=a^2+2ab+b^2$ 中的字母是可变的数，因此是变量。固定一些变量为常量，多项式和方程式便出现了。其中一元方程式就有线性方程、二次方程、三次方程……

$$ax+b=0,\quad ax^2+bx+c=0,\quad ax^3+bx^2+cx+d=0,\quad \cdots$$

有了各种方程，如何求解啊？于是在数学天才们的努力下，这些方程的根的表示渐渐地知道了：

$ax+b=0$ 的根是 $x=-\dfrac{b}{a}$；

$ax^2+bx+c=0$ 的根是 $x=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$；

……

呵，这些解好完美，数学真是无懈可击啊。那 $x^2+1=0$ 的根是什么？

呃……

n 百年后，这个问题最终抽象出来虚数的概念（呵，把数都抽虚了）。

数学在持续完美中：

三元方程式的根被解出来了。

四元方程式的根也被解出来了。

那五元方程式的根如何表示啊？如何像二次方程的根一样可以用系数表示出来的形式？这个问题厉害，解决这个问题伽罗瓦更犀利，他竟然把运算规律也进行了更抽象的分类，一下子竟抽出了群的伟大概念！

由此现代代数学的核心概念群、环、域、映射、线性空间等纷纷现世，标志着代数从局部性研究转向系统结构的整体性分析研究阶段。

比如线性代数中的核心抽象概念是线性空间，即所谓的要满足“加法”和“数乘”等八条公理的元素的集合。也就是说，只要某个集合里的元素满足那么几条公理，元素之间的变化满足这些规律，我们就可以对这个集合进行一系列线性化处理和分析，这个集合就叫线性空间。这个陌生的集合的性质和结构特点我们一下子就全知道了，因为宇宙间的所有的线性空间类的集合的性质都一样，地球人都知道（如果地球人都学了线性代数的话）。多么深刻而美妙的结论！这就是代数的一个抽象特性。

代数学

代数学是对算术的推广，是根据算术法则把一些表示数的字母结合起来。它研究的是抽象实体（如复数、集合、向量、矩阵、群、环、域等）以符号形式进行运算，类似于算术运算的性质和关系。

“代数”这一个词在我国出现较晚，在清代时才传入中国，当时被人们译成“阿尔热巴拉”，直到 1859 年，清代著名的数学家、翻译家李善兰才将它译为“代数学”，一直沿用至今。

那么线性问题又是什么样的问题呢？

“线性”是数学中使用十分广泛的词汇。我们常说的“一次方程”和“一次函数”，原本都是“线性方程(Linear Equation)”和“线性函数(Linear Function)”。在大学里，流行“线性”、“线性代数”、“线性变换”、“线性常微分方程”、“线性偏微分方程”、“线性规划”、“线性算子”、“线性泛函”、“线性控制系统”、“拟线性”、“准线性”等。为什么以向量为基本对象的“线性数学”会流行呢?

从科技实践中来的数学问题无非分为两类：一类线性问题，一类非线性问题。线性问题是研究最久、理论最完善的；而非线性问题则可以在一定基础上转化为线性问题求解。因此遇到一个具体的问题，首先判断是线性还是非线性的；其次若是线性问题如何处理，若是非线性问题如何转化为线性问题。

相对于“非线性数学”来说，线性数学比较简单。 微积分学的基本思想是“以直代曲”，局部地以切线代替曲线。 于是，在某种条件下，微分方程就可以近似地变成“线性代数方程组”。20世纪有了电子计算机，无论未知数的个数n有多大，都可以设法计算。于是，把以n维向量为对象的线性方程组搞清了，许多复杂的数学问题也就有解了(至少是近似的)。

既然线性问题如此重要，到底什么是线性呢？下面我们通过介绍几个主要的概念来逐渐地把握“线性”这个数学的核心概念。

1.2 “线性”的意义

线性代数里面的线性主要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意义。

1.2.1　线性函数的概念

线性函数的概念在初等数学和高等数学中的含义不尽相同（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，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（见图 1-2）；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，即图中直线 $\boldsymbol{m}$。显然，过原点的直线是最简单的线性函数。

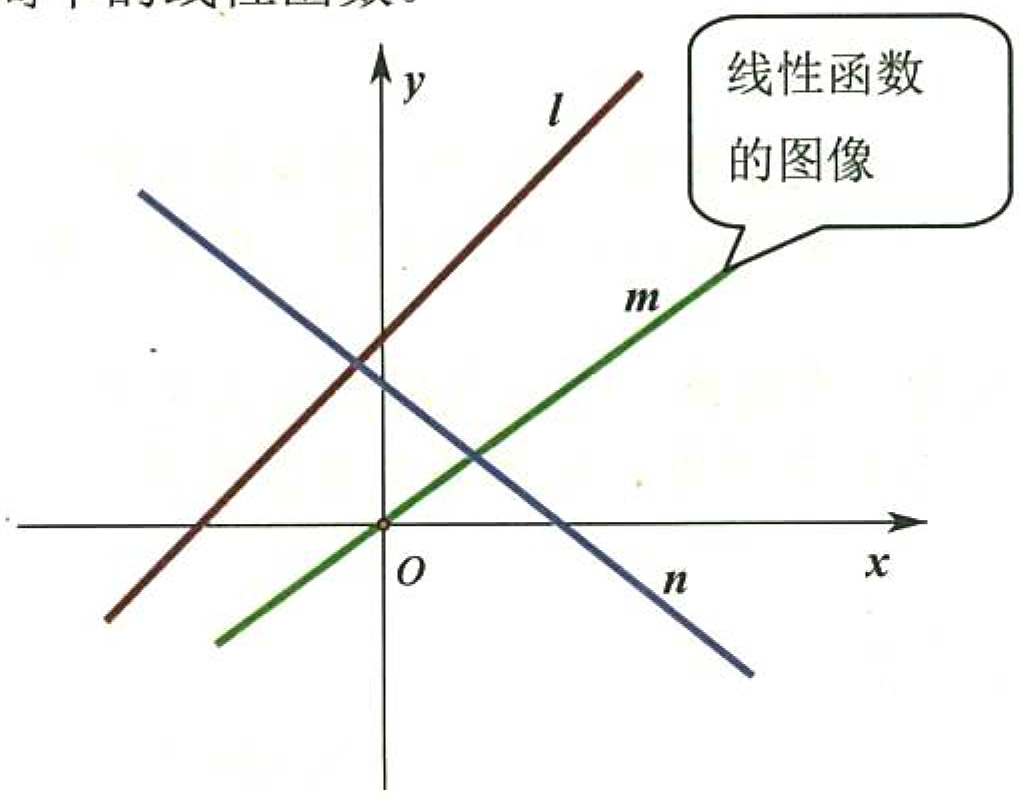


图 1-2 一元线性函数

严格说来，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。

扯啥呢？难道 $f(x)=kx+b$ 不是线性函数吗？是线性的，但不是线性代数里所指的线性含义。只因为不过原点的直线不满足线性代数里对线性函数的比例性的要求（别急，马上有解释哈）。

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。

（1）可加性：即如果函数 $f(x)$ 是线性的，那么有

$$f(x_1+x_2)=f(x_1)+f(x_2)$$

一句话：和的函数等于函数的和。物理意义是说因变量叠加后的作用结果等于各个因变量独自作用结果的叠加。

（2）比例性：也叫做齐次性、数乘性或均匀性，即如果函数 $f(x)$ 是线性的，那么有

$$f(kx)=kf(x) \qquad k\text{为常数}$$

一句话：比例的函数等于函数的比例。物理意义是说因变量缩放，因变量的作用结果也同等比例地缩放。

对于函数 $f(x)=ax+b$ 而言不满足此比例性，$f(kx)=akx+b$，$kf(x)=akx+kb$，因此 $f(kx)\neq kf(x)$。因此严格地讲，$f(x)=ax+b$ 不能再叫线性函数了。或者说，线性代数的线性变换不直接研究坐标系的移动。

可加性与比例性组合在一块就是“线性”的全部意义了，即有

$$f(k_1x_1+k_2x_2)=k_1f(x_1)+k_2f(x_2) \qquad k_1、k_2\text{为常数}$$

一句话：线性组合的函数，等于函数的线性组合。这里面既有缩放又有叠加的物理含义。

可加性和比例性的物理意义上的事例有哪些呢?

线性函数的可加性表明函数所描述的事物具有累加性，所有起因的累加所导致的结果完全等于每个起因独自所引起的结果的累加。可加性看起来简单，似乎没有内涵，其实它界定了所描述的事物是线性还是非线性的。举两个例子：

一个是晶体三极管放大器（见图 1-3，设晶体管的电流放大倍数是 100)，晶体管的电流放大特性可简单分两个区间：线性区和饱和区。在线性区里面，如果基极电流 I_b 是 1mA，则集电极电流 I_c 就是 100mA；如果基极电流是 2mA，则集电极电流就是 200mA。进一步地，如果输入的基极电流是 3mA(可看作 1mA 和 2mA 相加)，则集电极电流就是 100mA 和 200mA 之和 300mA。线性区里面的电流放大过程满足可加性。但在饱和区（见图 1-3，设晶体管的饱和电流是 300mA）里面就不满足这个可加性。当基极电流是 4mA 时，集电极电流是 370mA（不是 400mA，因为不在线性区)；当基极电流是 5mA 时， 集电极电流是 410mA。好了，基极两个输入电流的和为 9mA，但是集电极电流只有 780mA 而不是 900mA，因为到了晶体管的饱和电流区了，晶体管内的载流子不够用，电流增加不上去了。饱和区失去了放大电流的累加性，因而不符合线性关系。

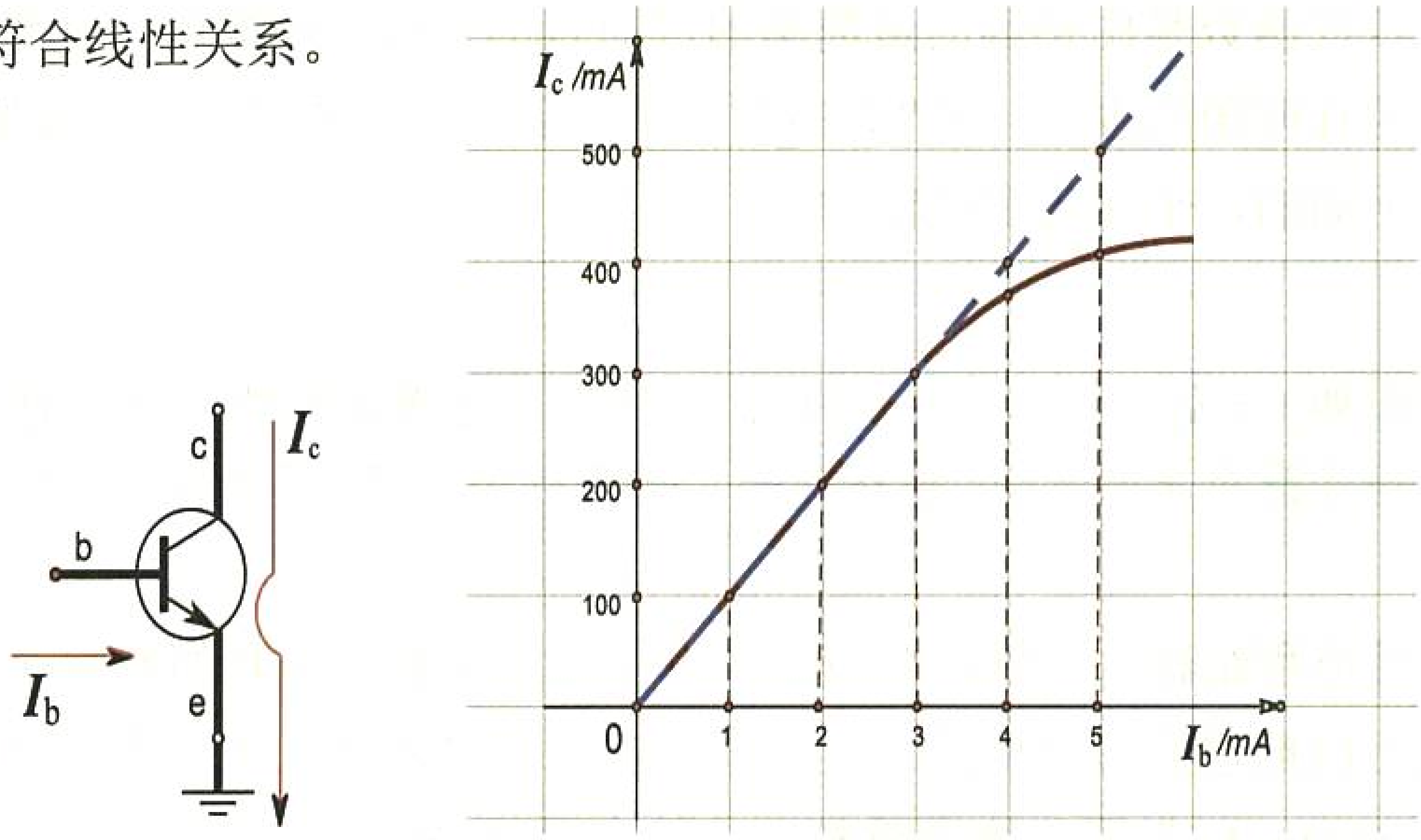


图 1-3 三极管的电流放大示意电路及其电流放大函数示例

二个是人力资源的故事例子，呵呵，属于小学一年级级别的脑筋急转弯：王五在旧上海滩的码头上扛货物麻袋，一天能扛 200 袋。好梦不长，几个月后，陈阿真也来扛麻袋了，谁？据说是大侠陈真韬光养晦化名陈阿真，一天可以扛 300 袋麻袋。好了，问个问题：王五和阿真两个人一天能抗多少袋麻袋?

“500 袋!”天真率性的同学们思维很线性，“200 袋加 300 袋就是 500 袋嘛。”

不对!（对了就不是脑筋急转弯了）据有围观之好事者统计，俩人第一天一共扛了 600 袋麻袋，第二天一共扛了 400 袋麻袋。怎么回事？原来人力资源的事情不是线性问题而是非线性的问题：

第一天，两人摸不透对方的脾气和底牌，为了保住岗位，各施功夫互相竞赛，王五扛了 250 袋，阿真扛了 350 袋，合计 600 袋。

第二天，两人一想，这样下去还不累死掉，一山不容二虎，给对方捣蛋弄走对方。于是两人便扛麻袋边向对方施展拳脚，内耗了，俩人一天共扛了 400 袋。

总结一下，线性的可加性既是没有互相激励的累加，也是没有互相内耗的累加。一加一

就是二，既不大于二也不小于二（对不起哈，一不小心证明了哥德巴赫猜想）。

比例性是啥物理含义呢？比例性又名齐次性，说明没有初始值。比如电路，没有输入信号时输出也没有，有几倍的输入量刚好就有几倍的输出量，增量是倍数关系，存量也是倍数关系。

1.2.2　线性函数概念的推广

实际上，高等的线性概念正是从最简单的比例函数进行推广的，在大学所学习的线性代数里的线性函数概念被扩展成一个多元线性方程组所表示的一个对应关系。如方程组

$$\begin{cases} y_1 = k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n \\ y_2 = k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n \\ \quad\vdots \\ y_m = k_{m1}x_1 + k_{m2}x_2 + \cdots + k_{mn}x_n \end{cases} \qquad (k_{11},\cdots,k_{mn}\text{是不变量})$$

是由 m 个 n 元线性函数组成的，而且这 m 个线性函数还是齐次函数，它们全部过原点。看看，高等概念的线性函数是由初等的最简单的过原点的直线扩展来的，如果 $m=n$，那么这个方程组所确定的几何图形也是一条“直线”！而且，空间中这条直线也是过原点的（$m \neq n$ 的图形是平面或超平面的，平面是多线性的）。

齐次函数

线性齐次函数形如 $y = k_1x_1 + k_2x_2 + \cdots + k_nx_n$，这个正比例函数的式子中每项里的变量出现的次数都是一次的（没有常数项），整齐划一，故此称为“齐次”的，全称为 **n 元线性齐次函数**。

把 n 元齐次线性方程组称为线性函数有点信心不足，看起来方程组和单个方程差别挺大的。其实，我们也可以把它们写得形式一致：重新定义变量就可以把它改写成初等数学中的线性函数的形式“$f(x)=kx$”了。这个重新定义的变量就是向量，扩展如下：

（1）初等线性函数的自变量由一个量 x 扩展定义为一个竖排的数组 $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$，因变量一个量 y 也扩展定义为一个竖排的数组 $\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$，这些 n 元数组或 m 元数组称之为列向量。

（2）初等线性函数的比例系数 k 扩展为由所有的 k_{ij} 构成一个数的方阵，称之为系数矩阵，即

$$\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix}$$

（3）然后再定义一种系数矩阵与向量相乘的运算法则（在“矩阵的几何意义”一章中介

绍），使我们可以把上述的线性方程组改写为如下的等价形式

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

（4）上式的形式可进一步简写为

$$\boldsymbol{y} = f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$$

其中：
$$\boldsymbol{y} = f(\boldsymbol{x}) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix},\quad \boldsymbol{K} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix},\quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

到了这里，我们看到，初等线性函数和高等线性函数的概念终于得到了形式上的统一。当然，一些区别还是有的：系数和变量这时候要用大写/黑体字母。

矩阵出现了

在这里，矩阵横空出世！在我们学习的线性代数课程里面，矩阵的理论及计算占了较大篇幅，原因就是矩阵实际上就是高等线性函数（这里指的是线性方程组）的系数。后面我们会提到，线性函数就是线性变换的表达式。那么矩阵就核心地代表了线性变换。因为 $\boldsymbol{y}=\boldsymbol{K}\boldsymbol{x}$ 就是一个向量 $\boldsymbol{x}$ 通过矩阵 $\boldsymbol{K}$ 变换为一个向量 $\boldsymbol{y}$，向量 $\boldsymbol{x}$、$\boldsymbol{y}$ 各自可以看做一段有向线段，那就是说线性变换就是把一个线段变成另外一个线段。谁决定了这个变换——矩阵。因此一个矩阵对应着一个线性变换，反之亦然。为什么在线性代数课里面大讲特讲矩阵和它的运算就是这个原因。不过令人匪夷所思的是我们的线性代数书（工科）里面竟然很少提及线性变换的概念，为补遗憾，在此我就较多地废话了线性方面的东西。

1.2.3　多元线性函数的几何意义

在前面的线性函数的推广中，从一元线性函数一下子推广到了 n 元线性函数组，跨度有点大了。下面补一下中间过程的课，探讨从一元线性函数如何推广到 n 元线性函数。

首先我们看看从一元线性函数 $f(x)=kx$ 拓展到二元线性函数 $f(x_1,x_2)=k_1x_1+k_2x_2$ 的几何解释。这个几何解释并不是唯一的，目的是让读者认识和理解线性的概念。

1. 拓展的第一步　坐标系由二维扩展到三维

$f(x)=kx$ 的直线图形是在二维笛卡尔坐标系下给出的几何图形，把它放到三维笛卡尔坐标系下，其函数表达式应写为 $f(x_1,x_2)=k_1x_1$ 或者 $f(x_1,x_2)=k_2x_2$。不失一般性，我们这里取 $f(x_1,x_2)=k_1x_1$ 为 $f(x)=kx$ 的扩维表达式。

我们知道，$f(x_1,x_2)=k_1x_1$ 的图形是一个过原点的平面，是由 $f(x_1)=kx$ 的图形扩维后由一根直线变成了一个平面。这是因为函数 $f(x_1,x_2)=k_1x_1$ 与新生长出来的坐标轴 x_2 没有关系，x_2 可以取任意值；换句话说，x_2 的任意值都在函数 $f(x_1,x_2)=k_1x_1$ 的图像上，这是一个过 x_2 坐标轴的平面。

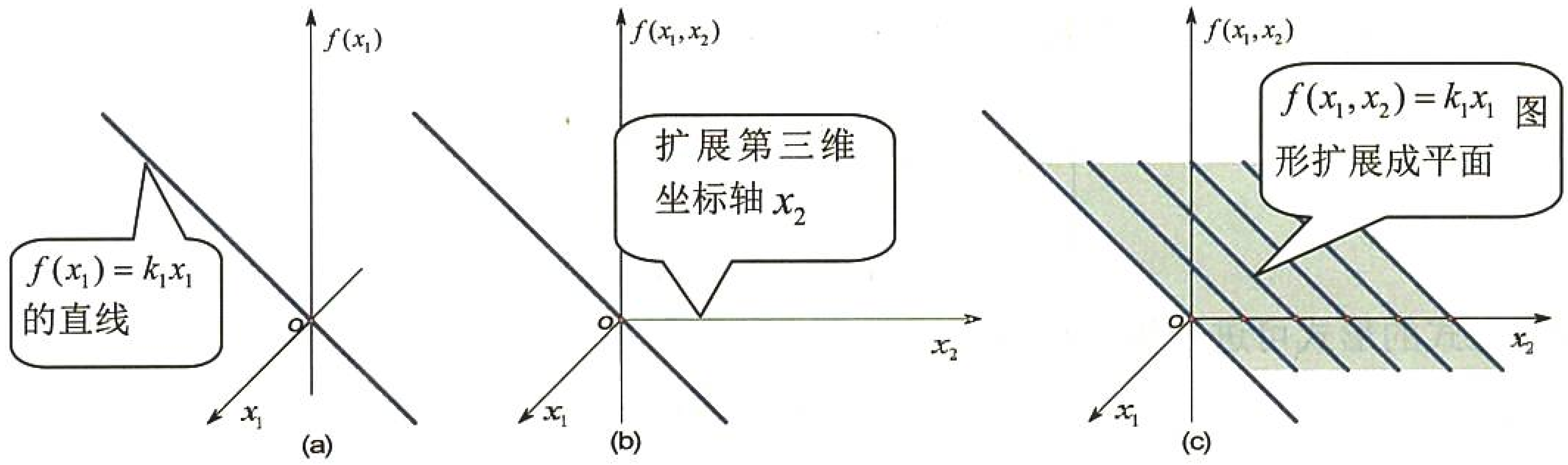


图 1-4 平面是由直线扩展而来的线性函数图形

形象的扩展过程可以这样想象：二维平面坐标系里有一根直线图形（见图 1-4（a）），这时有 x_2 轴过原点以垂直于坐标系 $x_1 \sim f(x_1)$ 的平面向右方向（右手系）生长出来（见图 1-4（b）），然后原来的那条直线 $f(x_1)=k_1x_1$ 沿着坐标轴 x_2 方向向右滑动，无数个平行的直线被 x_2 轴像竹帘子一样串起来，平铺得到了 $f(x_1,x_2)=k_1x_1$ 的平面（见图 1-4（c））。这个平面是由无数的直线铺成的，因此平面也是“线性”的。

2. 拓展的第二步 两个平面加起来

显然，要得到函数 $f(x_1,x_2)=k_1x_1+k_2x_2$ 的图形，只要把三维坐标系下的两个函数 $f(x_1,x_2)=k_1x_1$ 和 $f(x_1,x_2)=k_2x_2$ 所对应的图形加起来即可。一般情形下，两个平面相加仍然是一个平面，如图 1-5 所示。

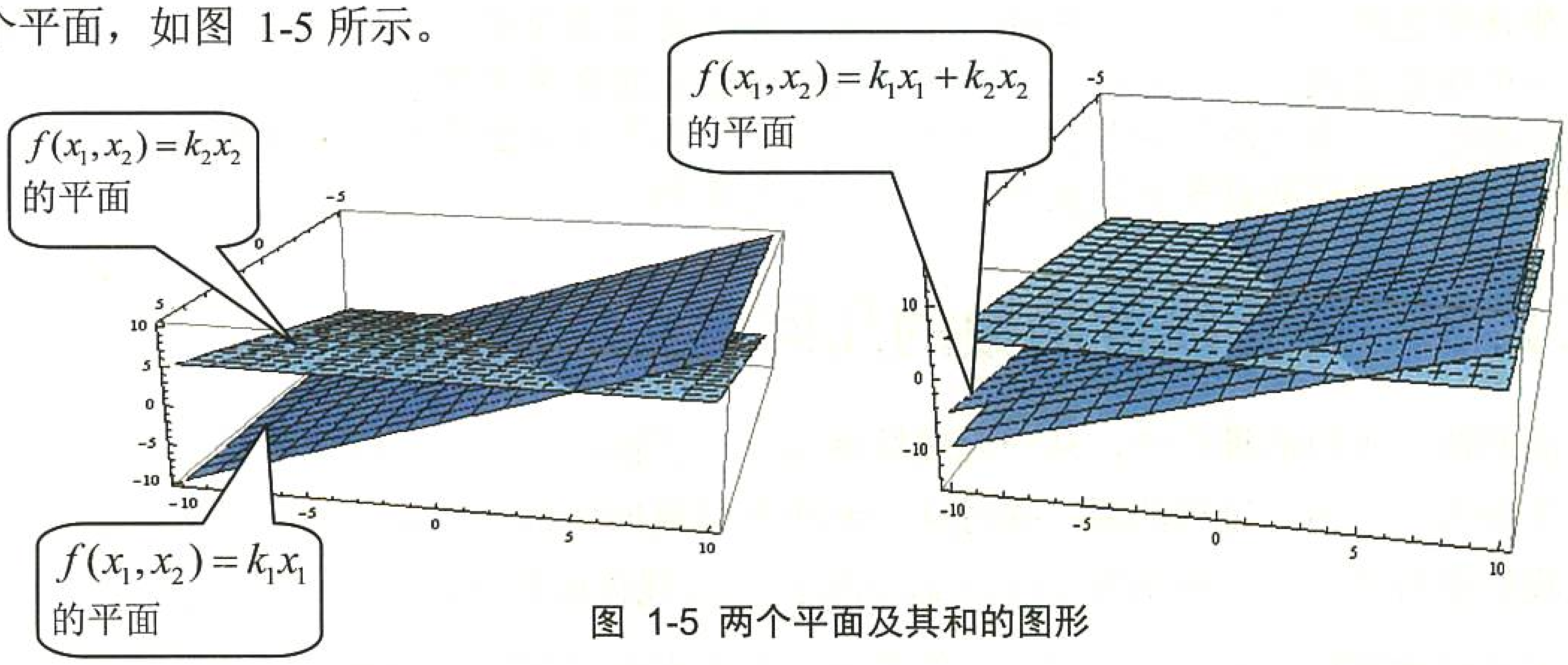


图 1-5 两个平面及其和的图形

因此，线性函数 $f(x_1,x_2)=k_1x_1+k_2x_2$ 的几何图形是一个过原点的平面。这个平面是在三维坐标系下的二维几何图形。

由二元线性函数 $f(x_1,x_2)=k_1x_1+k_2x_2$ 继续扩展到三元线性函数 $f(x_1,x_2,x_3)=k_1x_1+k_2x_2+k_3x_3$ 时，所在的坐标系由三维扩展到四维。可以想象，这个三元变量函数构成了一个三维空间，是由三个空间 $f(x_1,x_2,x_3)=k_1x_1$，$f(x_1,x_2,x_3)=k_2x_2$ 和 $f(x_1,x_2,x_3)=k_3x_3$ 叠加得到的，因此它是一个四维空间中（四维坐标系）的一个三维子空间。

继续扩展到四元及 n 元的线性函数 $f(x_1,x_2,\cdots,x_n)=k_1x_1+k_2x_2+\cdots+k_nx_n$，坐标系空间扩展到五维乃至 n+1 维，其几何图形仍将是一个低于坐标系维度一个维数的“子空间”。

这个 n 元几何图形总是低于坐标系一个维数。我们常常把一个高维的坐标系称为一个“空

间”（比如 n+1 维实数域线性空间 $\mathbf{R}^{n+1}$），那么，只能把这个线性函数低一维的几何图形称为一个“平面”，这是一个扩展意义上的平面，常被称为超平面（对于三维现实“空间里”的我们而言，低一维度的子空间就是平面）。超平面等同于包含在 n 维空间 $\mathbf{R}^n$ 中的 n-1 维欧式空间，它们对应于通常三维空间中的二维平面、平面内的直线、直线上的点等。

把线性函数 $f(x_1,x_2,\cdots,x_n)=k_1x_1+k_2x_2+\cdots+k_nx_n$ 的形式改写为

$$k_1x_1+k_2x_2+\cdots+k_nx_n-f(x_1,x_2,\cdots,x_n)=0$$

或者更一般的形式为

$$k_1x_1+k_2x_2+\cdots+k_nx_n+k_{n+1}x_{n+1}=c$$

这是一个 n+1 维空间 $\mathbf{R}^{n+1}$ 中的一个 n 维超平面，只是这个平面不一定过原点了（注意，不过原点的超平面可称之为空间，但不能称之为线性空间）。

到此我们明白了多元线性函数的“线性”不能单纯地理解为空间中的一条直线了，根据上面的讨论，把线性函数几何图形想象成一个平面更有代表性。实际上，把 n 个 n 元线性函数组成一个满秩方程组才能表示一条直线。

相比较而言，线性函数中含有的参数少，涉及的运算简单，仅为加法和乘法，便于运算，是变量数学中最简单的函数；但另一方面，许多复杂的函数都可以在一定范围和精确度下近似地用线性函数来表示，所以线性函数又是变量数学中最重要的函数。

1.2.4　n 维（高维）空间的直观理解

由三维扩展到四维的空间的确难以想象，下面我想谈谈个人的一些认识供读者参考。

对于笛卡儿坐标系，二维坐标系的两个坐标轴互相正交并构成一个平面空间；三维坐标系的坐标轴互相正交而且第三个坐标轴垂直于其余两个坐标轴构成的平面，三个坐标轴构成一个立体空间；四维坐标系中的四个坐标轴互相正交，第四轴必然与其余的三维立体空间的整体相垂直，四个坐标轴构成一个超多面体空间……以此类推，理解可矣。

比如，四维空间的物理解释的一个例子就是爱因斯坦的时空统一理论，三维物理空间之外增加了一个与之正交的时间轴（垂直或正交的意思应理解为不相关，时间和空间在远低于光速的尺度内就是没有关系的两个事物，这就是牛顿的世界）。你，作为一个有生命周期的高级动物实际上是个四维动物，因为你的肉体既占有了一个三维小空间同时又占有了另外一维的时间轴上的一段长度。

有人总觉得爱因斯坦把时间作为空间的一维有点不爽——时间和物体的尺寸怎么能类同看待呢！即使把时间乘以常数光速作为一维数轴也不好接受。那咱下面就举个爽点的例子。

四维以至 n 维正交空间的物理解释实际上是人们在抽象他所观察到的宇宙物体时出现的概念。在一个银河系外的观察者看来，太阳系不过是视界平面上的一个二维点；当这名观察者快速逼近太阳系时，这个二维平面上的点逐渐变成了三维的太阳系空间，同样，此时的地球在观察者的二维视界平面上也是一个点而已；当观察者来到地球外的大气空间时，地球已是一个三维球体了，而地球上的一个人同样在此时的观察者看来是一个点而已……如果观察者继续体察入微，将会逐步地看到这个人的身体、身体上的细胞、染色体、有机高分子、原

子、原子核、夸克等，一直看到宇宙结构的终结者——振颤着布满所有空间的弦！

这是一个空间套着一个空间的 n 维空间，就像洋葱一样，大的三维空间套着无数个层层缩小的三维空间。整个宇宙空间里的维轴有无穷多个，这些无穷数的维轴都两两正交——正是维轴的尺度范围不同保证了它们的正交性。

佛教的释迦牟尼应称得上是一位几何学家，因为 n 多年前他老人家就率先说过，一粒沙子就是一个大千世界。膜拜了。

现实一点的物例。比如一根水管，近距离观察是个三维物体，远距离观察就是一根无粗细感的二维线段了（见图 1-6）。

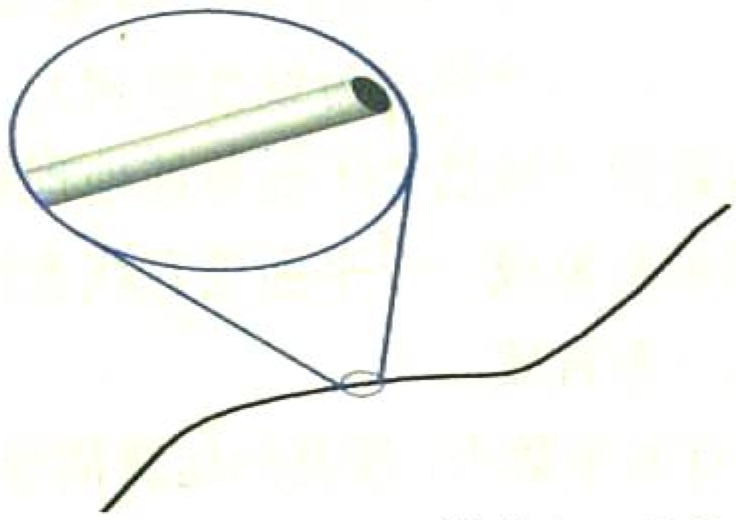


图 1-6　一根水管可以是二维或者三维的

再比如（见图 1-7），近距离的三维空心砖，这样的空心立方体有许多个并排着，立方体的端面并不平坦；如果远距离观察，许多个端面构成一个新的三维物体或图形——椭圆抛物面；这样的整个物体看起来就可以认为是一个四维的物体或四维空间里的图形。

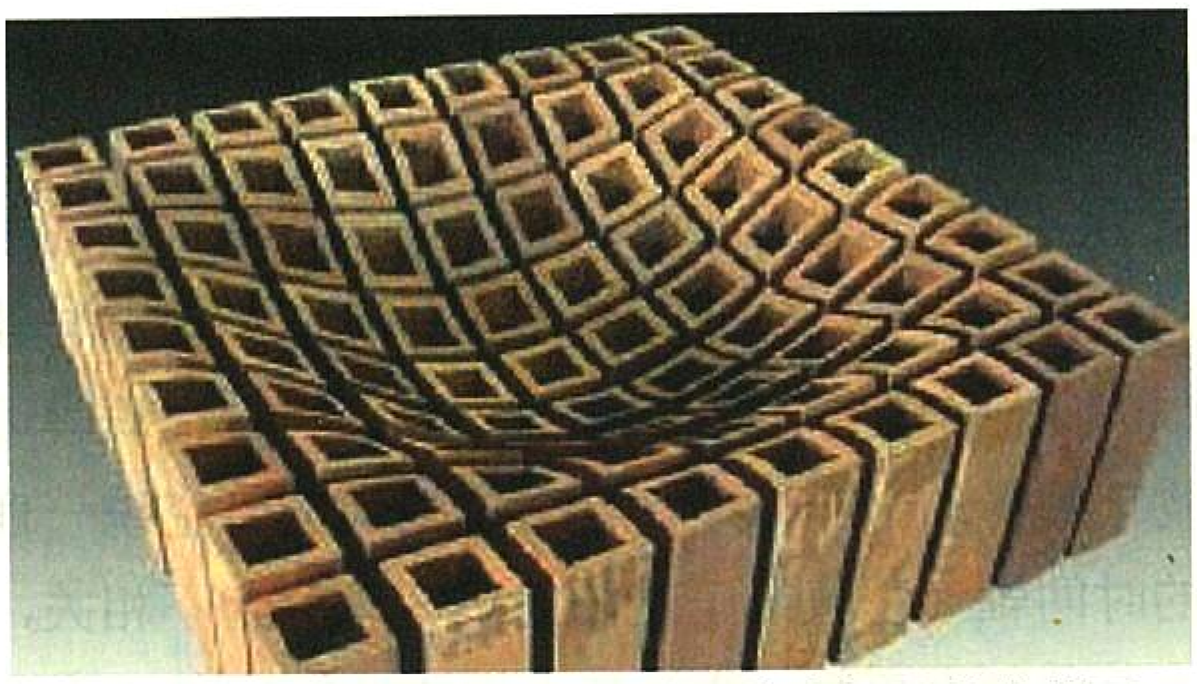


图 1-7　三维空心砖可以构成一个四维空间里的抛物面

这样的一个视点远离就降维的规则其实就是局部空间的看法，世界是由无数个相邻的局部小空间构成的（但如果整个宇宙看成是一个均等的平直空间的话就没有这种说法了，就是前面的最高四维时空了）。

如何画出四维以及更高维的图形呢？

比如我们可以用上述的想法画出一个四维的超立方体（见图 1-8）。

图 1-8（a）是三维空间中的一个立方几何体；图 1-8（b）是把空间向外扩展为四维，同时立方体也向外扩展为四维超立方体。

当然，三维立方体也可以向其内部扩展为四维的超立方体，这样看起来更像个立方体。图就不画了。想当然的，五维超立方体也应该如法炮制，n 维也可以如法炮制了。

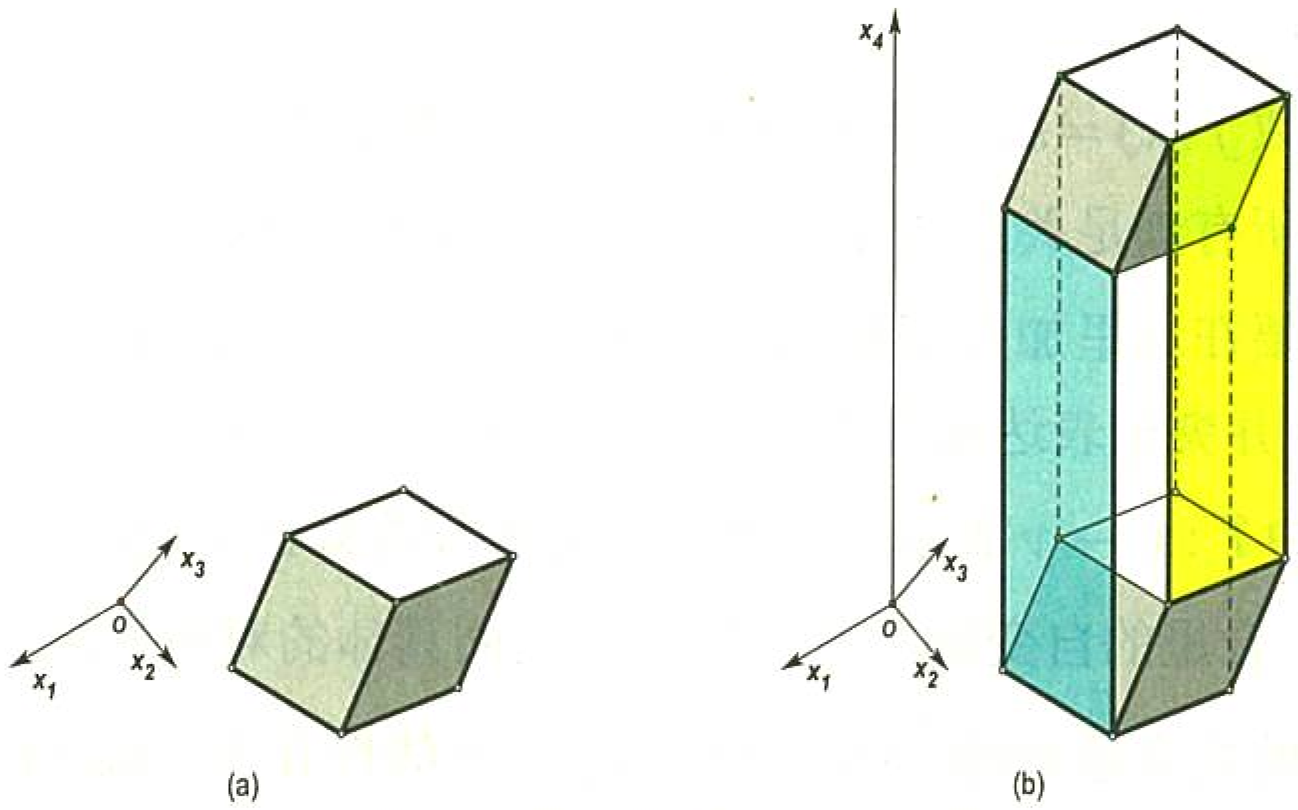


图 1-8 四维立方体的一个画法

除了上述的高维空间套着 n 重低维空间的佛家三千世界的理解外，线性代数的 n 维空间却是众生平等的世界。实际上，在以后的线性代数学习中，n 维空间中的 n 维向量的元素常常是由相对独立的因素组成的，这些相对独立的因素或者完全独立（垂直或正交）或者相对独立（线性无关，我有你没有的部分），我们对此的物理意义是不难理解的。因此对于 n 维空间的几何图形的想象，坐标轴的现实意义上的垂直不是必须的，如果几何意义上的垂直或正交代之以代数意义上的线性无关的概念，取消了几何作图垂直的要求，我们在平面上就可以通过斜坐标系画出 n 维以上的空间了，你就形象地理解了由 n 个向量张成的 n 个空间的理论，进而想象高维空间的图像也就不是一件特困难的事情了。

线性函数的几何意义给人的印象是一个静止的点或线的图形，其实线性函数还有一个运动的概念就是映射！线性映射或线性变换的概念实际上是线性代数的核心内容。

那线性映射大概是个什么东东呢？

1.3 线性映射和线性变换的几何意义

1.3.1　线性映射的几何意义

前面说，初等线性函数 $f(x)=kx$ 和高等线性函数 $\boldsymbol{f(x)=kx}$ 的表达式一致，因此线性函数的概念形式上是统一的。这种统一在数学的实质意义上也是一致的，就是函数的“线性”，实质上就是指变量之间的“线性关系”。

我们再来回味一下这句关于线性函数中心性质的话：线性组合的函数，等于函数的线性组合。详细说来就是，当自变量从 x 变换为它自身的线性组合 $k_1x_1+k_2x_2$ 时，其函数也从 $f(x)$ 变换为函数的线性组合 $k_1f(x_1)+k_2f(x_2)$。因为函数的本意是因变量与自变量之间的对应关系，所以“线性”的本质就是因变量与自变量之间始终保持组合形式不变的一种对应关系，我们把这类特殊的对应关系称之为“线性关系”。因此我们所说的“线性”，实质上就是指变量之间的“线性关系”。

实际上，我们可以引入一种运动的思想，把函数看成一种变换，一种映射，一种从自变量的集合对应变换到因变量的集合的瞬间过程。

这正是线性代数的中心思想之一。

对于初等的线性函数 $f(x)=kx$ 而言，我们需要改变中学老师谆谆教导。中学老师说，线性函数的几何图形是所有满足关系式 $y=kx$ 的点 (x,y) 所累积起来的图形。这个静态的图形概念需要改造改造。要在这里加入变换或映射的动作（注意：是动作，一个瞬时的变化动作，只有开始和结果），并突出表达这种变换和投射的关系，我们把表达式 $f(x)=kx$ 改写成 $T:x\to y, x\mapsto kx$。其中 $T:x\to y$ 表示为一个从自变量数的集合 x 到因变量数的集合 y 的映射，$x\mapsto kx$ 表示两个集合里的自变量 x 到因变量 y 之间具体的对应变换关系。

下面来看一个映射的集合示意图。图 1-9 给出了一元线性齐次函数 $f(x)=kx$ 当 k 取不同的数时的映射对应关系。在三个分图中，有一个共性就是元素 0 必然映射到元素 0。

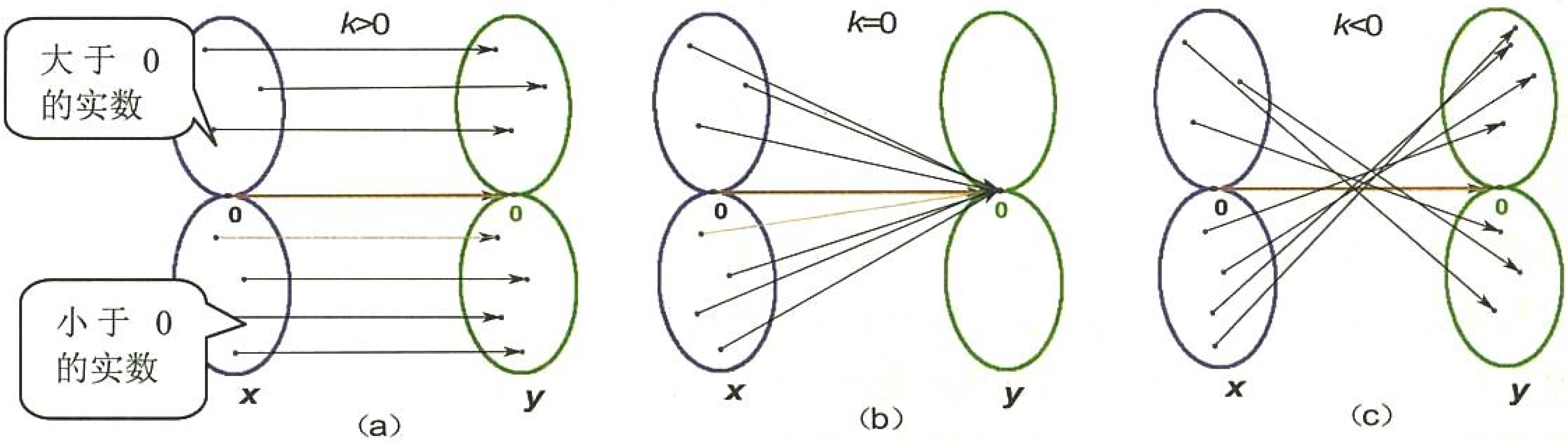


图 1-9 比例函数 $y=kx$ 在集合上的映射关系

在集合上建立坐标系，用坐标系里的点表示集合里的元素，就可以把集合上的映射关系几何化了。

对于一元线性齐次函数 $f(x)=kx$，集合 x 和集合 y 都是实数。大家知道，一个实数域可以用一根坐标轴来表示。因此集合 x 的坐标系就是一根线轴，写为 x 轴；集合 y 的坐标系也是一根轴，写为 y 轴。这样，我们就可以用坐标轴上点之间的映射来替代图 1-9 集合的映射表示法（见图 1-10）。

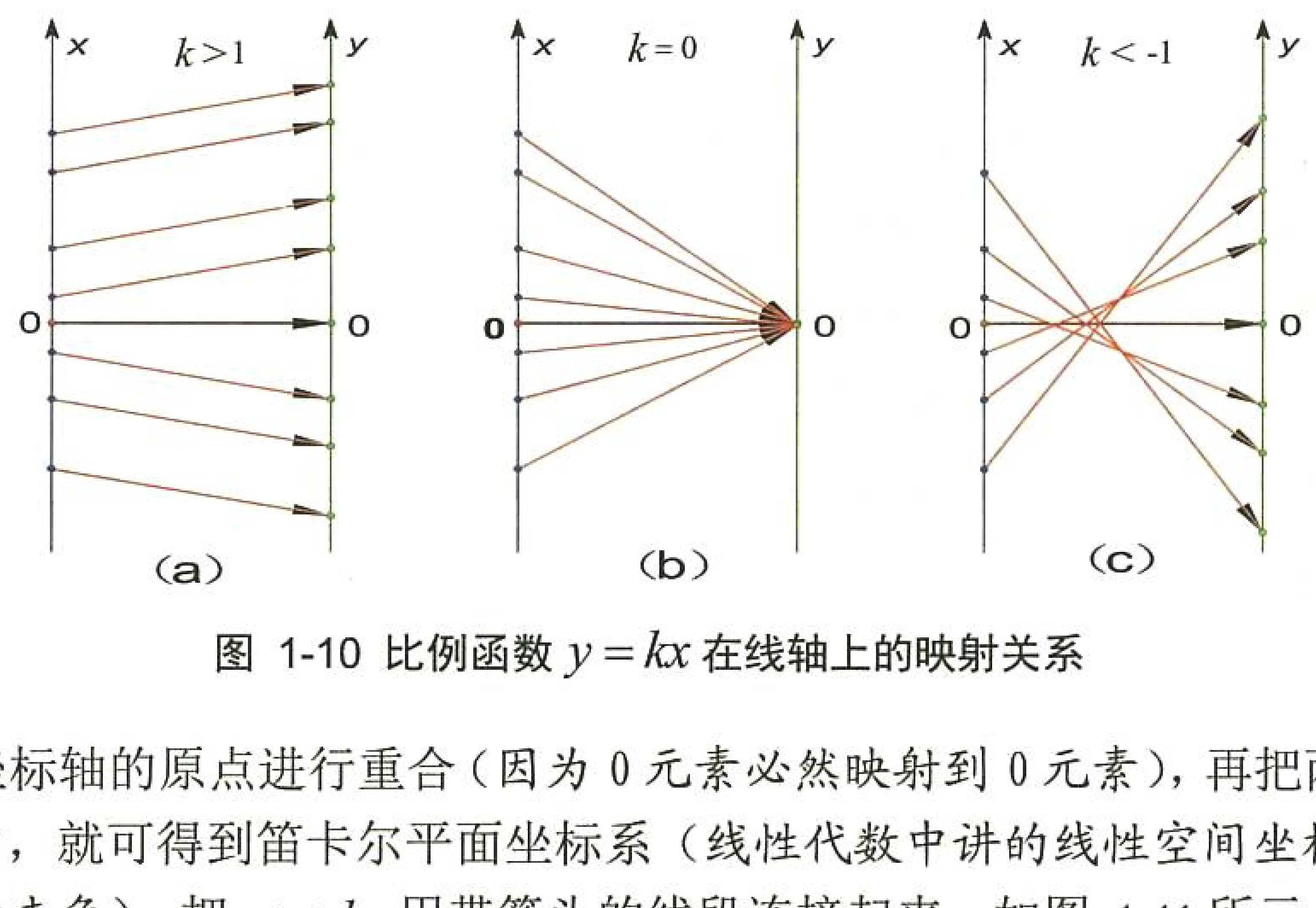


图 1-10 比例函数 $y=kx$ 在线轴上的映射关系

如果把两个坐标轴的原点进行重合（因为 0 元素必然映射到 0 元素），再把两个坐标轴的夹角调整到 $\pi/2$ 角，就可得到笛卡尔平面坐标系（线性代数中讲的线性空间坐标系的坐标轴可以是任意非零的夹角）。把 $x\mapsto kx$ 用带箭头的线段连接起来，如图 1-11 所示（图中只画出

了 $k>0$ 的映射情况）。

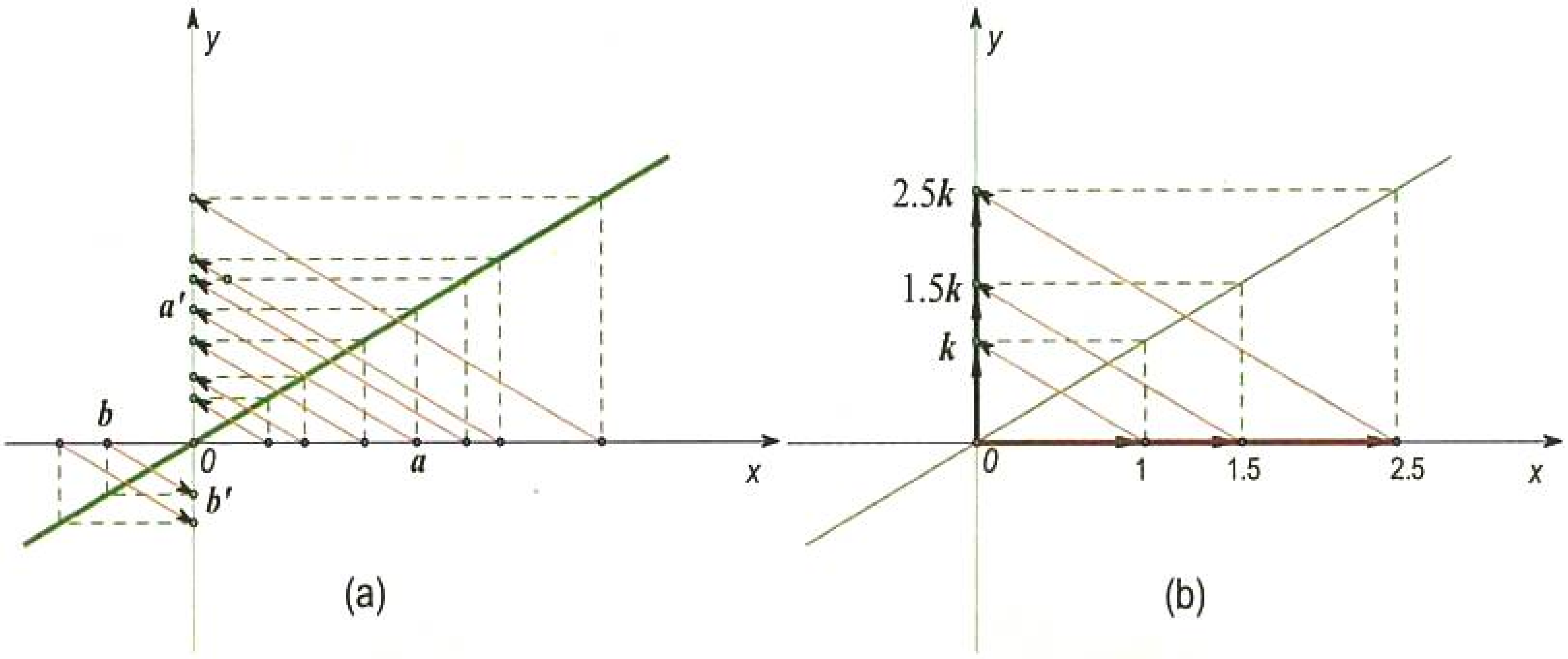


图 1-11 比例函数之点的映射和向量的映射

如图 1-11（a）所示，x 轴上的点 a 和 b 等分别映射到 y 轴上的点 a' 和 b' 等等。过一三象限的直线就是所谓的一元齐次函数 $y=kx$ 的图形，它是由 k 值确定的；就像平行光线通过平面镜进行反射一样，元素通过直线完成了映射。显然，这是点到点的映射。而且，不同的比率 k 值确定了一个不同的映射关系。

如果把点 a、a'、b 和 b' 分别与原点 0 连起来，就会得到线段 $0a$、$0b$、$0a'$、$0b'$。于是线段 $0a$ 映射到线段 $0a'$，线段 $0b$ 映射到线段 $0b'$。

到这里我们有了一个暂时的总结：**线性映射就是把线段映射到线段**（这里，映射终于和“线”扯上关系了）。如果我们把线段改称为向量的话，这个总结就是：**线性映射就是把向量映射成向量(** 见图 1-11（b）**)**。线性映射把向量变成另外一个向量，或者说把“线”变成“线”，因此得名。

当然，这个线性映射也满足线性的可加性和比例性的性质：可加性就是 x 轴上的两向量的和映射得到的 y 轴向量等于两个 x 轴向量分别映射得到的 y 轴向量的和，比例性就是 x 轴向量的倍数映射得到的 y 轴向量等于 x 轴向量映射的 y 轴向量的倍数（见图 1-11（b））。

用一般的数学表达式来描述线性映射的定义就是：

$$T(\boldsymbol{a}+\boldsymbol{b})=T\boldsymbol{a}+T\boldsymbol{b}$$
$$T(k\boldsymbol{a})=kT\boldsymbol{a}$$

其中，T 是映射运算，$\boldsymbol{a}$、$\boldsymbol{b}$ 是任意两个向量。

算子

你看这里，T 本来表示一种线性映射的动作关系（或函数关系），但在上式中就像一个实数或变量一样参与运算。如 $T(\boldsymbol{a}+\boldsymbol{b})=T\boldsymbol{a}+T\boldsymbol{b}$，就像乘法对加法的分配律一样展开运算，因此 T 在这里也叫线性算子。具体的算子有微分算子、积分算子、拉普拉斯算子等。

对于高等的线性函数 $f(\boldsymbol{x})=\boldsymbol{Kx}$ 而言也有同样的结论，我们把此表达式改写成

$$T:\boldsymbol{x}\to\boldsymbol{y},\boldsymbol{x}\mapsto\boldsymbol{Kx}$$

其中 $T:\boldsymbol{x}\to\boldsymbol{y}$ 的形式表示一个从自变向量的集合 $\boldsymbol{x}$ 到因变向量的集合 $\boldsymbol{y}$ 的映射。

为了方便看到 $T:\boldsymbol{x}\to\boldsymbol{y},\boldsymbol{x}\mapsto\boldsymbol{Kx}$ 的几何解释，我们看看二维线性函数的坐标式：

$$\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}a_1&a_2\\b_1&b_2\end{bmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

下面给出一般情况下的映射图像（细节在后续的章节里有详细描述）。

因为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 都是二维向量，所有任意的向量 $\boldsymbol{x}$ 的集合将构成平面 Π_1（见图 1-12），在平面 Π_1 上构建二维坐标系 $\{x_1ox_2\}$；同理，所有任意的向量 $\boldsymbol{y}$ 的集合构成平面 Π_2，在平面 Π_2 上也构建二维坐标系 $\{y_1oy_2\}$。所以，二维线性函数就构成了两个二维平面之间由矩阵 $\boldsymbol{K}=\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}$ 所确定的映射关系（矩阵是二维的比率）。

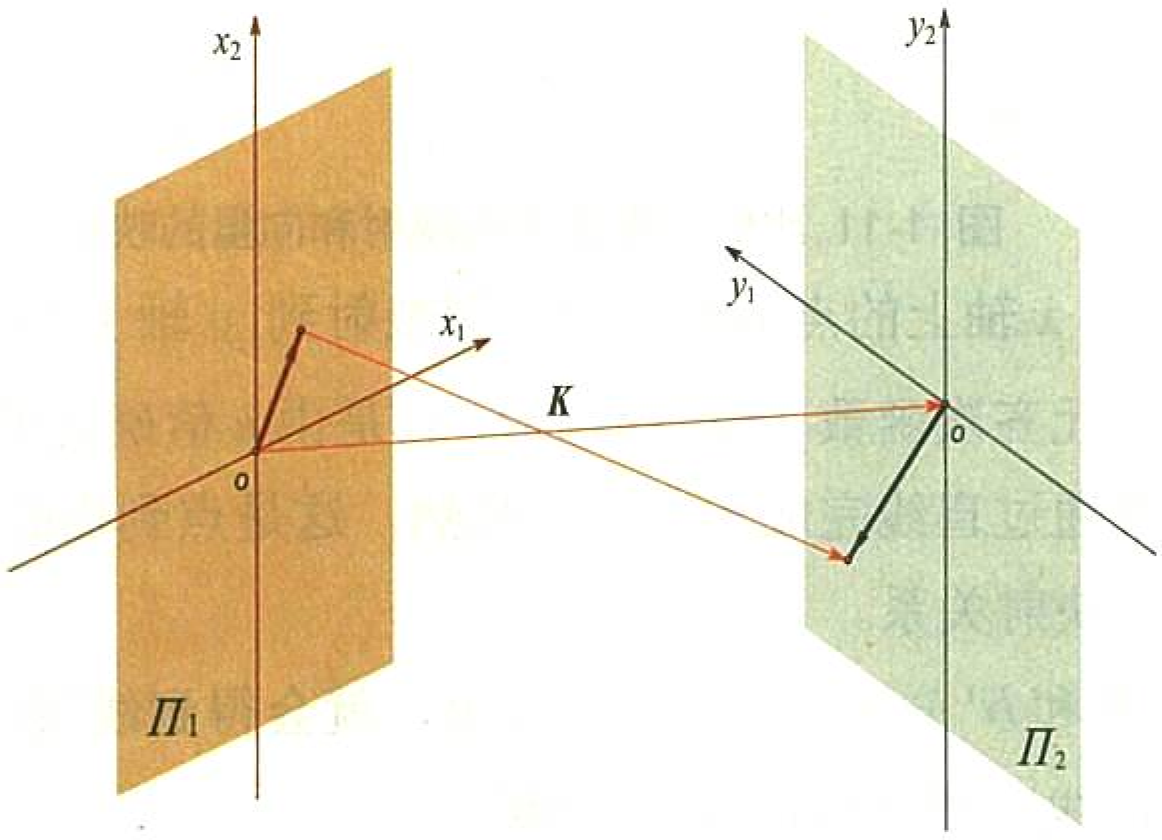


图 1-12 由矩阵 $\boldsymbol{K}$ 所确定的两个平面上二维向量的映射关系

平面 Π_1 的原点 0 始终映射到另一个平面 Π_2 的原点 0，这是线性映射的最基本要求。

为了更紧密地观察映射之间的关系，我们把平面 Π_1 放平，并使两个平面的原点 0 重合，就得到了一个由两个相交平面所构造的三维空间。

图 1-13 中，把平面 Π_1 上的向量 $\boldsymbol{x}$ 标注为 $\boldsymbol{a}_i$，把平面 Π_2 上的向量 $\boldsymbol{y}$ 标注为 $\boldsymbol{b}_i$ (为了和坐标系的标注区别开来)。

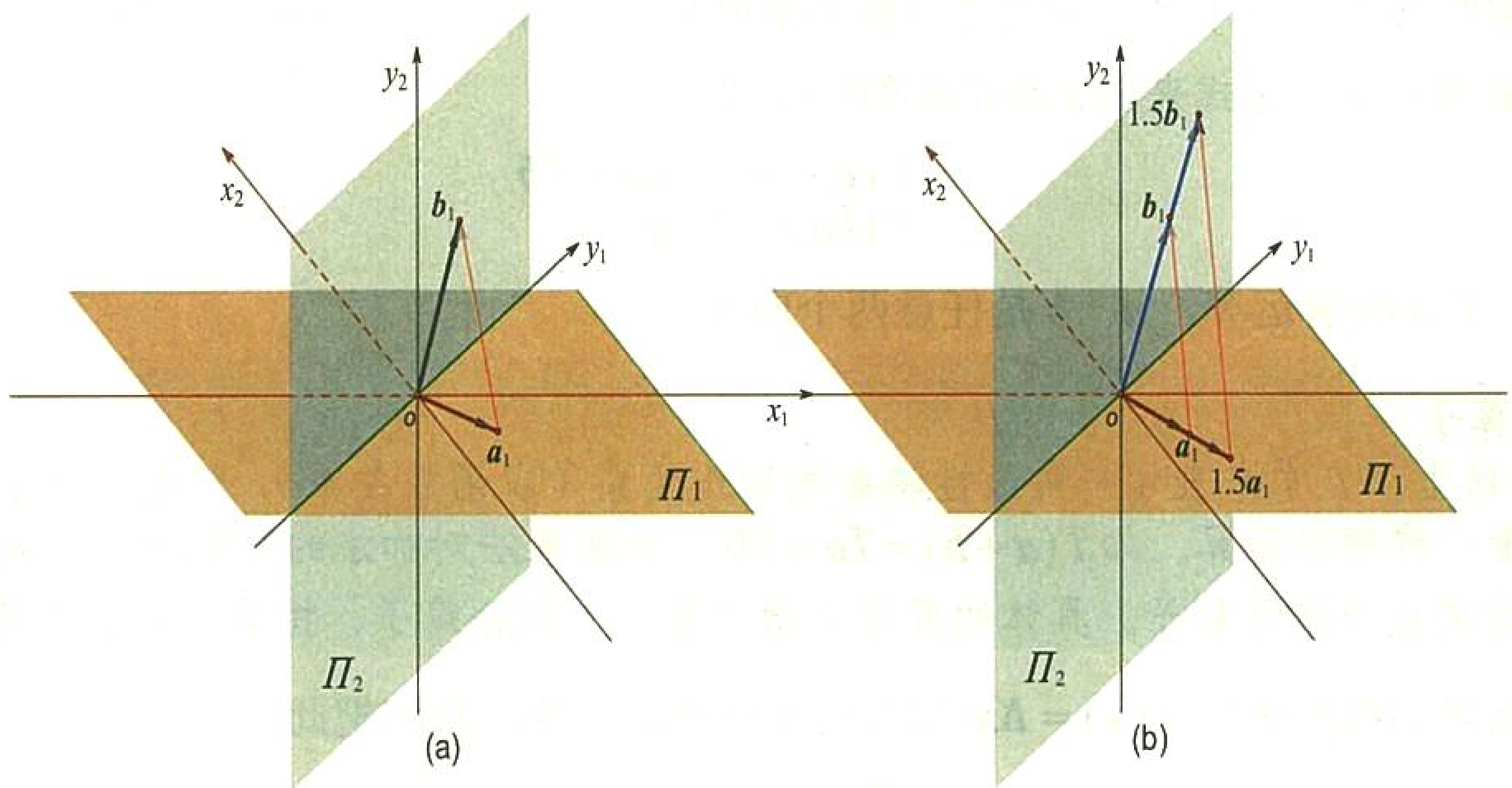


图 1-13 向量之线性映射的比例性

图 1-13（a）表示矩阵 $\boldsymbol{K}$ 把平面 Π_1 上的一个向量 $\boldsymbol{a}_1$ 映射到平面 Π_2 上的向量 $\boldsymbol{b}_1$，也即有向线段映射为有向线段；图 1-13（b）表示把一个向量 $\boldsymbol{a}_1$ 比例放大到 1.5 倍后被矩阵 $\boldsymbol{K}$ 映射到平面 Π_2 上的向量 $1.5\,\boldsymbol{a}_1$，这满足线性映射的比例性。实际上，不论矩阵 $\boldsymbol{K}$ 的元素是什么实数，对于任意的矩阵 $\boldsymbol{K}$，都有这个结论。因为

$$\boldsymbol{a}_1 \to \boldsymbol{K}\boldsymbol{a}_1 = \boldsymbol{b}_1$$
$$1.5\boldsymbol{a}_1 \to \boldsymbol{K}(1.5\boldsymbol{a}_1) = 1.5(\boldsymbol{K}\boldsymbol{a}_1) = 1.5\boldsymbol{b}_1$$

对于线性可加性，我们也给出了图形。由图 1-14 看出，一个平面上两个向量分别映射到另一个平面上的两个向量；对应地，它们的和 $\boldsymbol{\alpha}_1$ 也作了同样的映射得到 $\boldsymbol{\beta}_1$。两个平面上的向量及其和各自构成一个平行四边形。

向量的平行四边形加法法则决定了一个平面上的平行四边形被矩阵映射成为另一个平面上的平行四边形，这两个平行四边形可能全等，可能相似，大部分情况下既不全等也不相似。

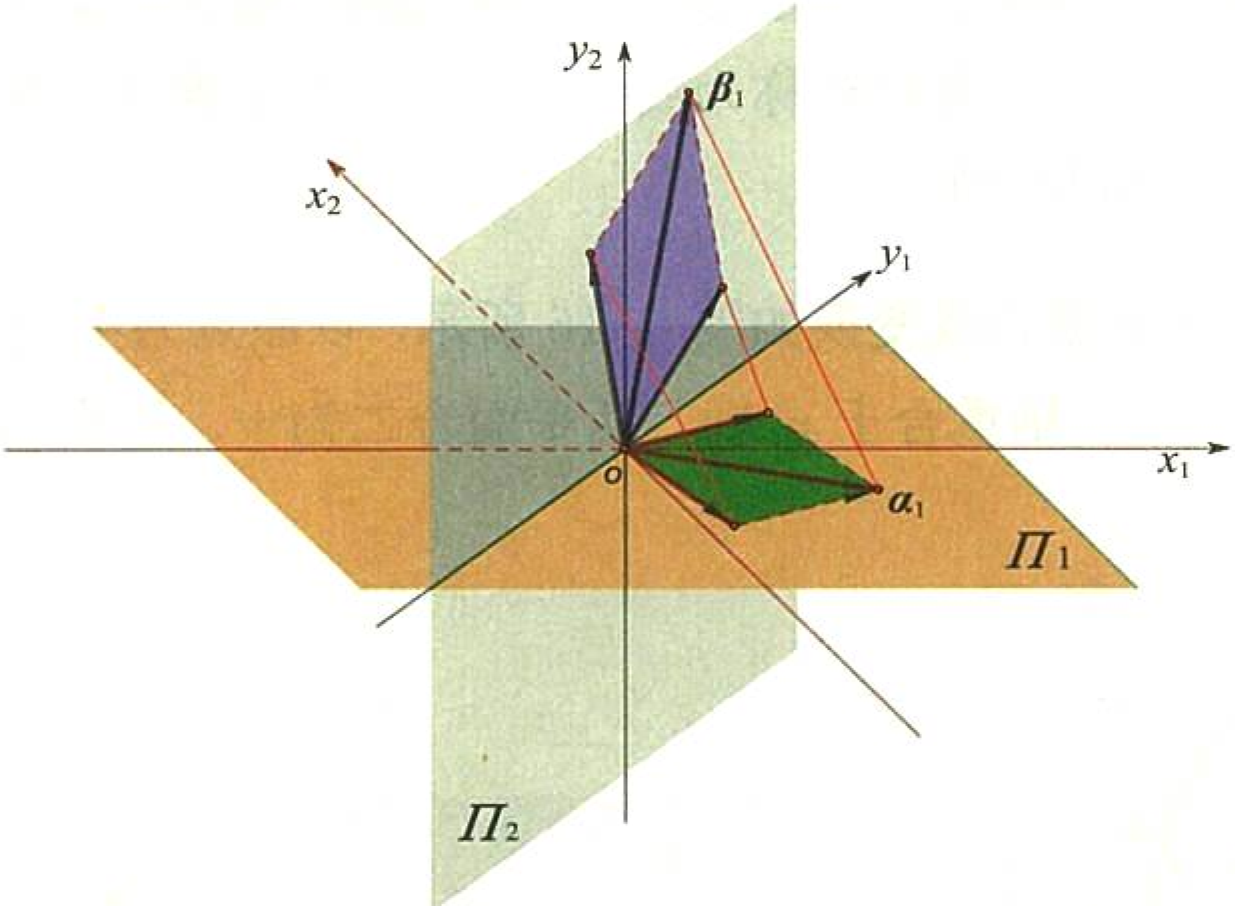


图 1-14 向量之线性映射的可加性

前面我们讨论了一个向量的映射或者是两个向量的和的映射情况。如果由无数等长而异向的向量构成的一个圆，那么这个圆会由某一个矩阵映射成什么图形呢？

实际上，可以被映射成圆、椭圆或者一根线段，特殊情况下被映射成一个点（这个点必然落在原点上）。大多数情况是被映射成一个椭圆，见图 1-15。图中圆上任意一个向量 $\boldsymbol{a}_1$ 映射成椭圆上的 $\boldsymbol{b}_1$ 。

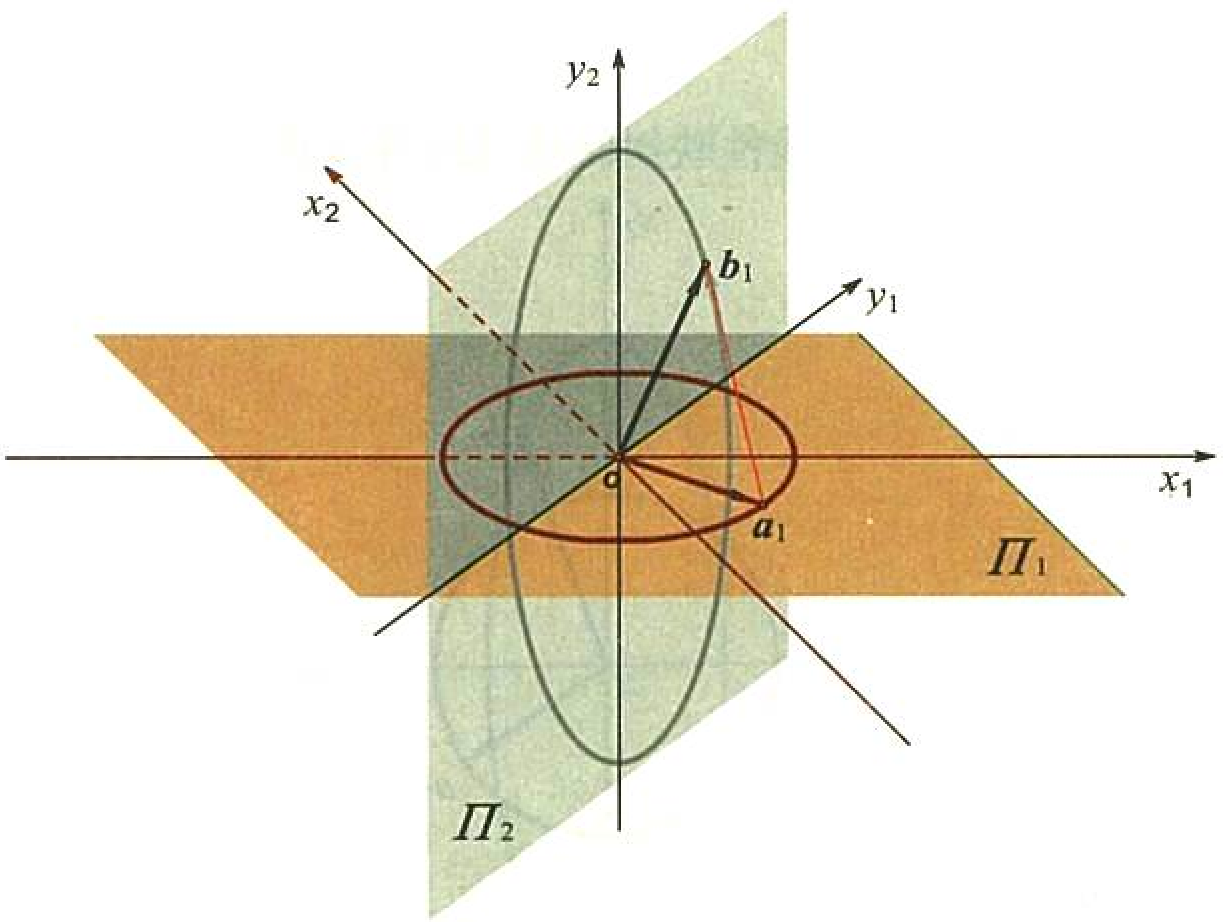


图 1-15 线性映射把平面 Π_1 上的圆映射成平面 Π_2 上的椭圆

在这里，把线性函数中 $f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$ 的变量 $\boldsymbol{x}$ 看成了一个图形“圆”而不只是一个向量，那么函数值 $f(\boldsymbol{x})$ 也就成了一个变换后的图形“椭圆”。

把一个线性映射放到二维平面及三维空间中去考察，细心揣摩其几何意义，就不难理解

概念的本质。例如，对于数乘变换$T(\boldsymbol{a})=k\boldsymbol{a}$，除了把$\boldsymbol{a}$看做向量外，我们可以把$\boldsymbol{a}$看做一个几何图形(其实向量也是一个几何图形，是一个简单的几何图形——有向线段，细节请阅第二章内容)，T 在 $k>1$时就是对$\boldsymbol{a}$放大，在$0<k<1$时就是对$\boldsymbol{a}$缩小，在$k=-1$时就是把$\boldsymbol{a}$反方向变化。

1.3.2 线性变换的几何意义

在大多数的教科书中，线性映射和线性变换被区别为两个概念。如果映射是发生在一个集合上的同一个坐标系中，线性映射就被称为线性变换。线性变换作为线性映射的特例，就是把集合上的两个坐标系合并为一个。

例如把二维平面圆的映射整合成变换如图 1-16 所示，以原点为不变轴心，把Π_2平面旋转放平，y_1轴重合于x_1轴，y_2轴重合于x_2轴，两平面合二为一。整合后的图形用平面直角坐标系表示就是图 1-17。

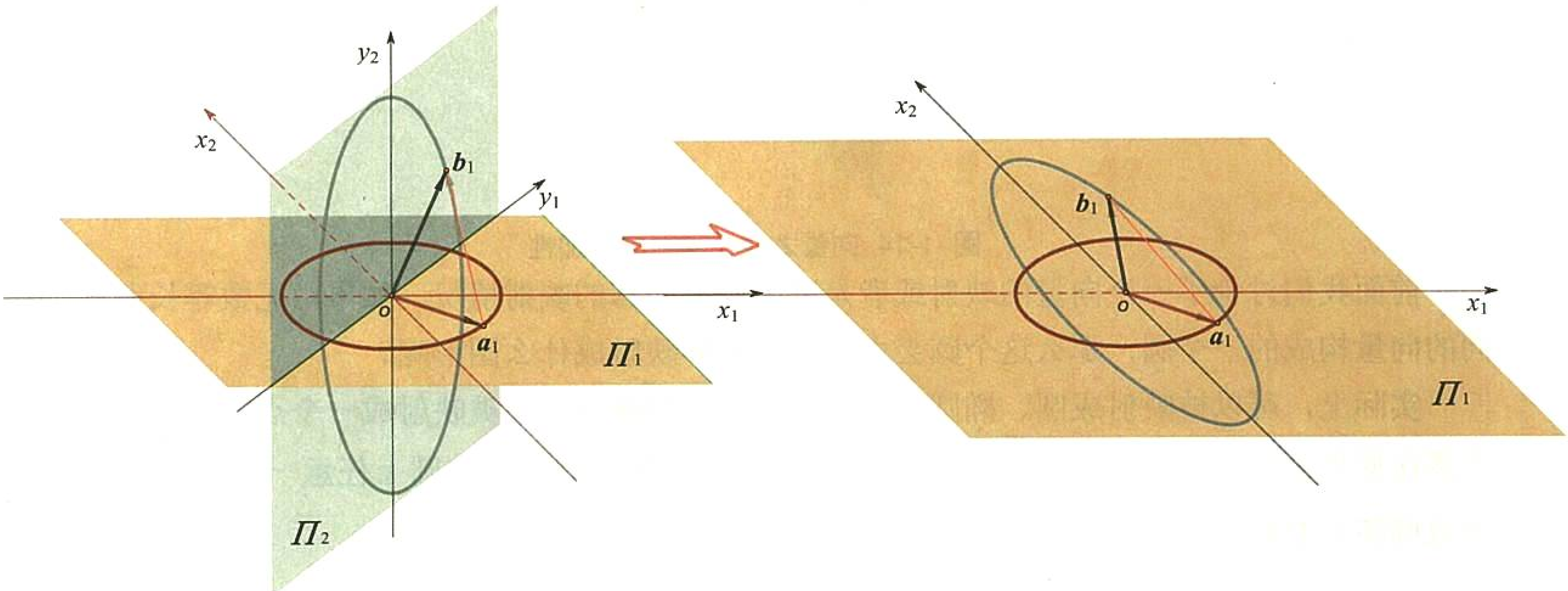


图 1-16 线性映射转换成线性变换

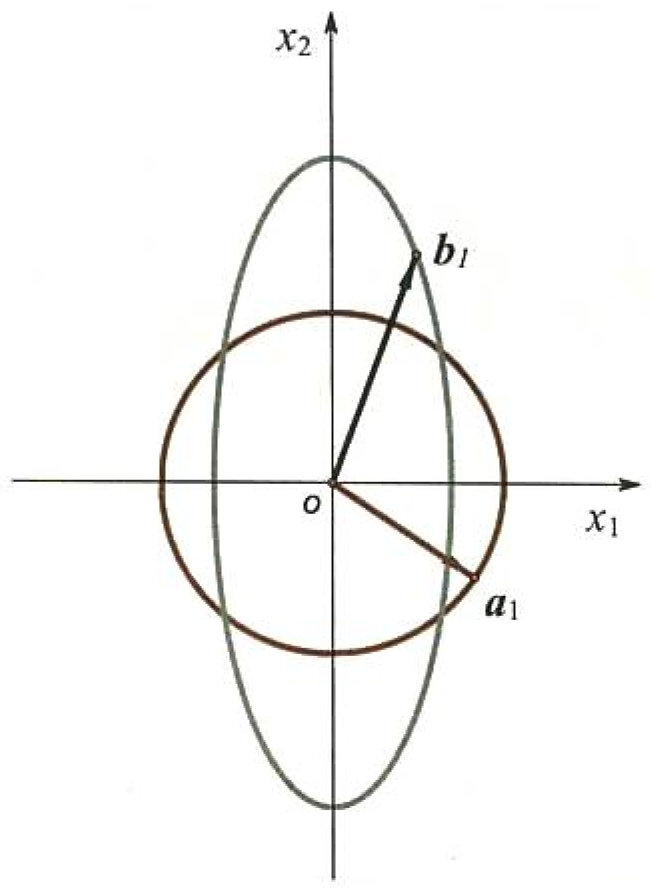


图 1-17 平面直角坐标系下线性变换把圆变成椭圆

直角坐标系下的图形清楚地显示了一个圆被线性变换为一个椭圆。相应地`，圆上的一个向量$\boldsymbol{a}_1$映射为椭圆上的向量$\boldsymbol{b}_1$。圆和椭圆都在同一个线性空间——平面上(平面就是空间)。

在线性代数中，我们主要讨论的是由矩阵所决定的线性变换的各种特性。下面看两个具体的线性变换的例子（下面的“'”为线性变换 $T(\ )$ 的另一种简单写法）。

在平面上所有从原点出发的向量构成的二维线性空间中，把所有向量绕原点作同样角度的旋转是一个线性变换。

如图 1-18 中向量旋转角度是直角。这时向量的和 $\boldsymbol{a}_1+\boldsymbol{a}_2$（$=\boldsymbol{a}_3$）旋转所得到的向量 $\boldsymbol{a}_3{}'$ 恰好等于 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 旋转所得到的向量 $\boldsymbol{a}_1{}'$ 和 $\boldsymbol{a}_2{}'$ 之和 $\boldsymbol{a}_1{}'+\boldsymbol{a}_2{}'$；数 k 与向量 $\boldsymbol{a}_1$ 的乘积 $k\boldsymbol{a}_1$ 旋转所得到的向量 $(k\boldsymbol{a}_1)'$ 恰好等于数 k 与向量 $\boldsymbol{a}_1$ 旋转所得到的 $\boldsymbol{a}_1{}'$ 的数乘积 $k\boldsymbol{a}_1{}'$。这就是说：

$$(\boldsymbol{a}_1+\boldsymbol{a}_2)'=\boldsymbol{a}_1{}'+\boldsymbol{a}_2{}',$$

$$(k\boldsymbol{a}_1)'=k\boldsymbol{a}_1{}'。$$

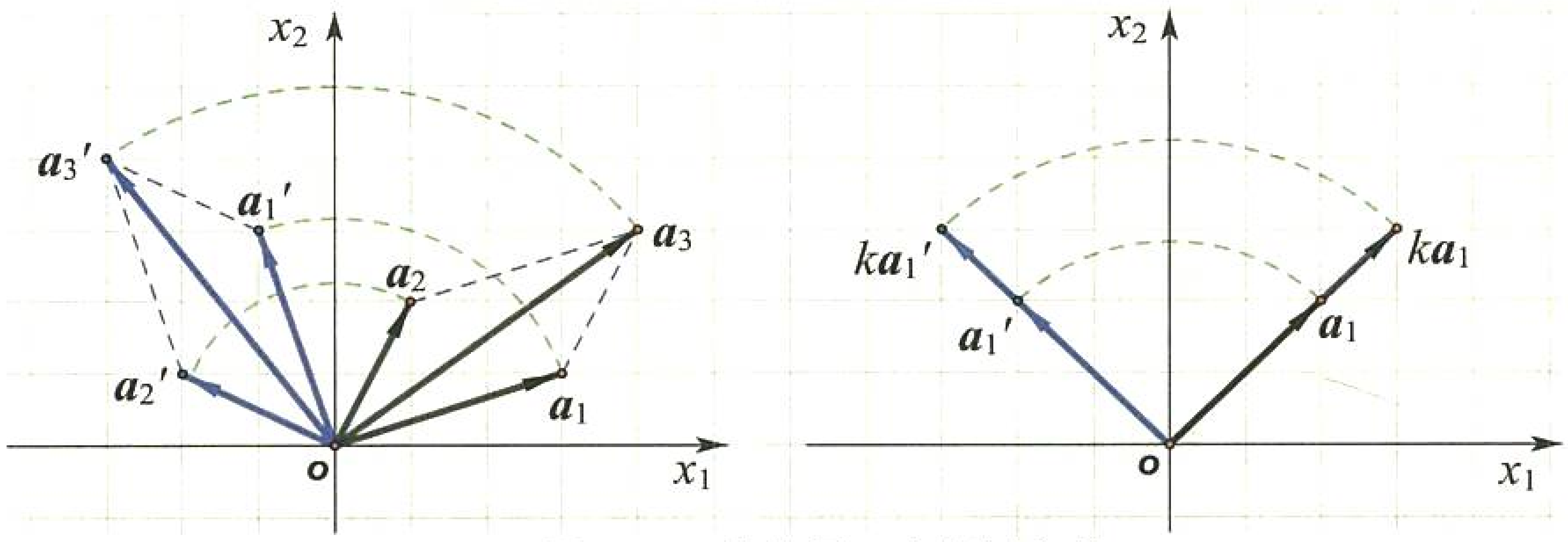


图 1-18 旋转是一个线性变换

另一个例子（见图 1-19）：在建立了空间笛卡尔直角坐标系的三维向量空间中，把每一个向量投影在坐标面 xoy 上，也是一个变换（投影变换）。这时，向量 $\boldsymbol{a}=(a,\ b,\ c)$，在此变换下的像为 $\boldsymbol{a}'=(a,\ b,\ 0)$，不难推知，此时也有 $(\boldsymbol{a}_1+\boldsymbol{a}_2)'=\boldsymbol{a}_1{}'+\boldsymbol{a}_2{}'$，$(k\boldsymbol{a})'=k\boldsymbol{a}'$。

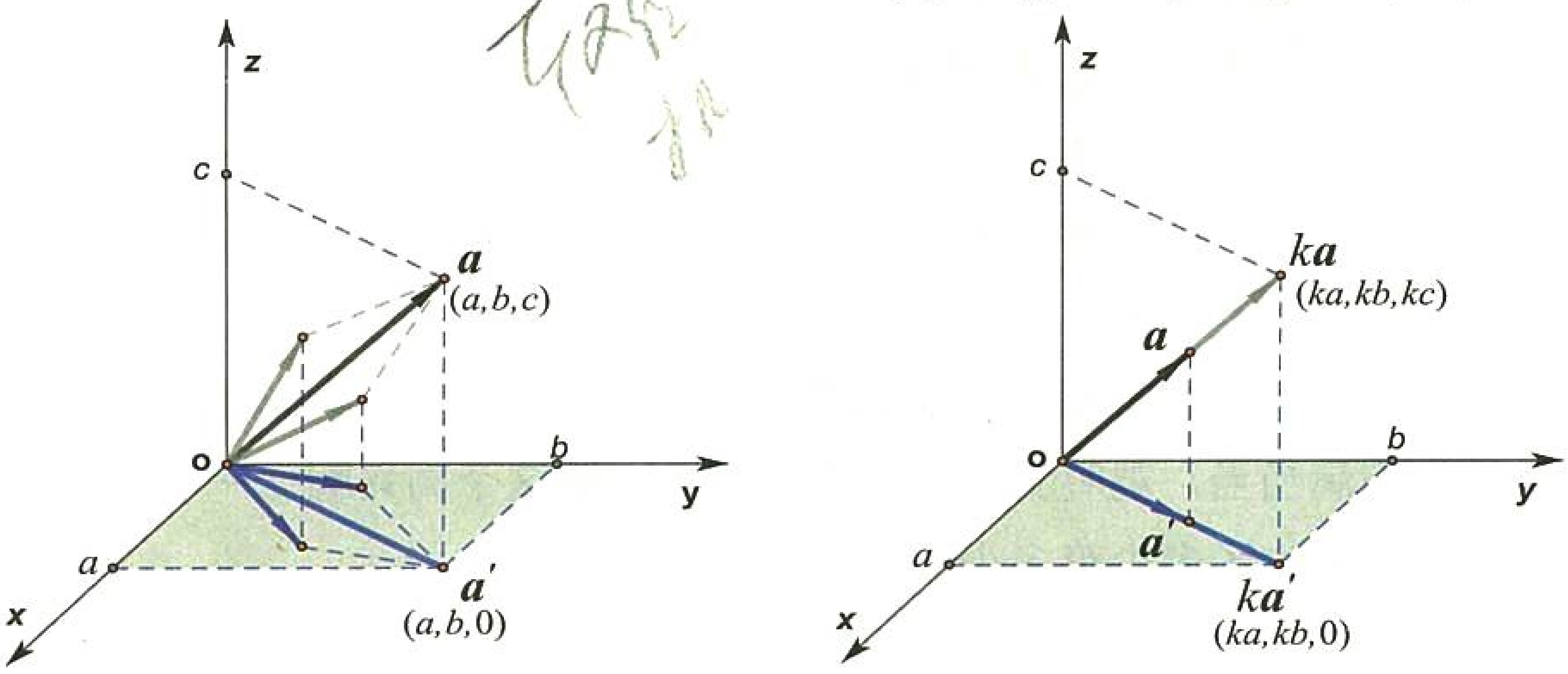


图 1-19 投影是线性变换

这两个变换有一个共同的性质：两个向量之和变换后，所得的向量恰好是把这两个向量变换后所得向量之和；数 k 与一个向量数乘后进行变换，所得的向量恰好是数 k 与把此向量变换后所得向量的乘积。此即所谓的可加性和比例性。这种使向量之间加法与数乘法关系都不受影响的变换——线性变换，它与线性空间的运算相适应，能够反映线性空间中向量的内在联系，是线性空间的重要变换。故线性变换的定义如下：

数域 F 上线性空间 V 中的变换 T 若满足条件：

$$T(\boldsymbol{a}+\boldsymbol{b})=T\boldsymbol{a}+T\boldsymbol{b}\quad(\boldsymbol{a},\boldsymbol{b}\in V)$$
$$T(k\boldsymbol{a})=kT\boldsymbol{a}\quad(k\in F,\boldsymbol{a}\in V)$$

则称 T 为 V 中的**线性变换**。

同线性映射一样，线性变换把向量变成另外一个向量，或者说把“线”变成“线”。在平面上，线性变换把原点仍变为原点（参考零点没有移动），直线仍然变为直线（没有打弯），平行线仍然是平行线，当然平行四边形仍然是平行四边形。

在工程中常用的差分运算、微分运算及积分运算都属于线性变换，都满足以上的可加性和比例性的关系。

通过后续的学习，我们将会逐步认识到线性变换可以有两个方面的含义：**变换空间里的向量，空间坐标系不变；或者变换坐标系而向量不变。两者是相对的，结果等价。**

1.4线性代数的故事

线性代数是高等代数的一大分支。我们知道，在研究关联着多个因素的量所引起的问题时，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。

线性问题或方程里的变量就是向量，因此一说“线性”必提“向量”。一般的线性代数课本里的主要内容及排列顺序大概为行列式、向量组、矩阵、线性方程组及二次型等，这些内容都是对向量的函数或组合。看看线性代数的发展历史，其发祥首先是从线性方程组开始的，而且可以说是从咱中国开始的。

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。中国是方程组解法的老祖宗，汉代重写的《九章算术》（秦代以前的数学书估计都被秦始皇坑了）一书上就有和高斯消元法一样的解法。这本书两千年来大都是中国的代数课本，这有点类似于欧几里得的《几何原本》两千年来一直是希腊乃至世界的几何课本一样。而在后代的《孙子算经》里面比较经典的解方程组问题是“鸡兔同笼”问题。 可以想象，这门学科的发展逻辑场景应该是这样的：

大概三国后面的魏晋南北朝时期，有个贩卖动物的小贩收购了一大车笼的鸡和兔子，鸡和兔子在竹笼里挤得满满当当，鸡鸣狗叫，好不热闹。小贩这个高兴啊——今天财运好啊，能挣不少钱。好吧，数数收了多少鸡多少兔子？

这下小贩该傻眼了吧？整个一大车笼，鸡靠鸡兔挨兔，无法分清。还不能打开笼盖数，一开盖鸡飞兔跑，只能隔着透明的竹笼数。这也不好数啊，中间也不透光，只能模糊地看见“上有三十五头，下有九十四足”，不知道兔的头还是鸡的腿。

其实小贩一点也不犯难，因为这行当退休的老爹已经告诉了他几个传了 n 多年的如何数鸡和兔子的算法：

解法一：

总脚数 ÷ 2-总头数=兔的只数；

总只数-兔的只数=鸡的只数。

解法二：

（个兔脚数 × 总只数 − 总脚数）÷（个兔脚数 − 个鸡脚数）=鸡的只数；

总只数 − 鸡的只数=兔的只数。

解法三：

（总脚数 − 个鸡脚数 × 总只数）÷（个兔脚数 − 个鸡脚数）=兔的只数；

总只数 − 兔的只数=鸡的只数。

嘿，这事儿不但有解，解法还真不少。

春秋战国末期的军事爱好者孙武是个穿越爱好者，那天刚好穿越到南北朝游历天下，打那路过，知道了这件事，就记录在自己的《孙子算经》里。

（鸡兔同笼问题最早记录于《孙子算经》，此书成书于约公元 473 年，朝代是属于南北朝，比春秋战国末期孙武的《孙子兵法》晚了 600 多年，因此《孙子算经》实际上是后人托孙子之名写成的——这好事现在绝迹了，要不就是孙武穿越了哈。）

孙武如果不小心穿越到现代，估计会有点郁闷，因为现在小学生都会轻易解这个问题。解法就是设鸡、兔分别为 x、y 个，列解下面的方程组即可：

$$\begin{cases} x+y=35 \\ 2x+4y=94 \end{cases} \tag{1-1}$$

我们看，解法一就是方程组的②式除以 2 减去①式求得 y；这是小学生们使用的消元解法，大学生们称之为高斯消元法。

类似的更复杂的消元算法在其他行当也有。比如农民在算“三种等级的稻捆数可打多少斗稻谷”的三元方程组问题用的也是消元法，即把数据排成行列矩阵，然后对增广矩阵实施行的初等变换，这事是在更早的汉朝《九章算术》里记载的。

解法二和解法三雷同。这个解法实质上就是线性代数里面行列式的解法——克莱姆法则。我们列出方程组(1-1)的矩阵/向量形式：

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 35 \\ 94 \end{pmatrix} \tag{1-2}$$

该向量方程的克莱姆法则的解答为：

$$x=\frac{\begin{vmatrix} 35 & 1 \\ 94 & 4 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix}} \tag{1-3}$$

$$y=\frac{\begin{vmatrix} 1 & 35 \\ 2 & 94 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix}} \tag{1-4}$$

解法二实际上说的就是式(1-3)的解法。计算内容就是行列式的求解过程。类似地，解法三实际上说的就是式(1-4)的解法。

与鸡兔同笼的问题类似，我小时候常听父亲讲“板凳鏊子”的问题（见后注），印象深刻。板凳鏊子的解法雷同于鸡兔同笼的解法二和解法三，同样属于行列式解法的克莱姆法则。这些形象的解法可被看做是克莱姆法则物理意义的解释。

可惜咱中国这么多年没有进一步发展这些东东，估计这与中国没有发明简明的数

学符号有关。虽然初创了大量新东西却不善于发展这些新东西。

后来这些书传播到了日本和朝鲜。日本古代的数学叫和算，和算是在中国古代数学的影响下发展起来的。公元六世纪始，中国的历法和数学就直接或间接地（通过朝鲜）传入日本，日本多次派留学生到中国学习数学。到八世纪初，日本已仿照隋唐时期的数学教育制度设立算学博士并采用《周髀算经》、《九章算术》、《孙子算经》、《缀术》等中国古算书作为教材……一直到了十七世纪末期，通过关孝和等人的工作，逐渐形成了日本数学体系——和算。

关孝和在日本被尊为“算圣”，他比较聪明，善于抽象。约 1683 年，他竟然在《九章算术》、《孙子算经》里这些线性方程组的解法中提炼出行列式的概念及其算法（也可能有印度数学的影响吧）。可以这样理解为何关孝和没有先发现矩阵：因为关算圣着重点是“算术”——计算的技术，他发现了在解不同线性方程组时共同的计算方式——行列式；而矩阵的概念是要发现函数、变换等更高级的概念后才会发现的抽象东西。

“板凳鳌子”的故事

1980 年初，正是俺上小学的时候。老爸是个乡镇企业会计，但常常感叹自己没有当老师：“唉，瞎了一个好老师的材料!”。为弥遗憾，常常在晚饭后给我们这些孩子们讲上一段，内容包括语文、数学、天文、地理历史典故等。声音洪远，常引得众多的街坊邻居来旁听休闲。其中一个板凳鳌子的数学题的解法正是二元线性方程组的解法。原题如下：

板凳鳌子三十三，一百条腿都朝天，问几个板凳几个鳌子?

题意说明一下，我们那里的板凳是四条腿的，鳌子(一种摊煎饼的平底锅)是三条腿的。板凳鳌子总个数是 33 个，把板凳和鳌子都反过来数一数腿的总条数是 100 个，问板凳鳌子各有多少个。

这个问题对于我们学过代数方程的小学生也是超简单：直接设个 x、y，列方程组，带入，求解，搞定。不过老爸的解法有创意但有点残忍。他的讲法是这样的：

“抓住一条板凳，砍掉一条腿；再抓住一条板凳，砍掉一条腿……所有板凳都砍完了。这时候板凳是三条腿，鳌子也同样是三条腿，那么三十三个板凳和鳌子共有几条腿？”(稍顿，环顾四周)“三乘以三十三等于——九十九条腿!”

“很好，原来共有一百条腿，你砍了板凳后变成了九十九条腿，那么你共砍掉了板凳几条腿？一条腿吧。好了，结果快出来了：你实际上砍掉了一条板凳腿；呃，别忘了砍法：一个板凳是砍掉一条腿，把所有板凳都砍完；所有的板凳数是几条？一条对不对？（砍掉的）一条腿对应一条板凳！共有一条板凳!”

“知道了共有一条板凳，又知道板凳鳌子总数三十三条，鳌子数是——三十二条！问题解决了!”

老爸意犹未尽，接着讲：

“你也可以不用砍板凳，可以给鳌子绑腿：抓住一条鳌子，绑上一条腿；再抓住一条鳌子，绑上一条腿……”。

和关孝和同时代的德国的莱布尼茨也参与了进来，只比关孝和晚 10 年，同样提出了行列式的概念及其有关定理。这次的参与意义重大，西方的数学/科学家们再一次开始发威了。

我个人觉得莱布尼茨对外来文化的学习吸收和归纳能力超强，偷师能力也超强，数学符号的设计能力也超强，进一步开发能力更强。比如，研究中国古代《易经》八卦的卦象而发

明二进制（他的二进制比传教士把《易经》传入欧洲的时候约晚 10 年），研究英国牛顿的“流数术”而发明微积分的写法（他的微积分比牛顿的微积分——流数术晚 10 年），估计日本关孝和的和算他也研究了（他的行列式比关孝和的行列式晚 10 年。呵呵，千古之谜的 10 年现象啊），随后提出了行列式并进一步发展，这并不是一件奇怪的事情。

后面的事情发展得越来越快，看起来顺理成章：

1750 年，瑞士的克莱姆，对行列式的定义和展开法则给出了完整、明确的阐述，并给出了现在解线性方程组的克莱姆法则。英国的西尔维斯特为了将数字的矩形阵列区别于行列式而发明了“矩阵”的术语，它的朋友凯莱进一步创立、发展了矩阵论。

……

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**表 1-1** 所示。

表 1-1　线性代数的英雄榜

序号	人物	理论贡献	力量指数
1	《九章算术》的作者们(汉代)，中国	线性方程组的消元解法	★★★★
2	关孝和（1642—1708），日本	最早提出行列式概念	★★★
3	哈密尔顿（1805—1865），爱尔兰	使用向量和标量的第一人	★★★
4	柯西（1789—1857），法国	1815 年启用行列式名词，1841 年提出特征方程概念，谱论重要贡献者	★★★★
5	克莱姆 （1704—1752），英国	给出了解线性方程组的克莱姆法则。	★★★
6	西尔维斯特（1814—1897），英国	1850 年启用矩阵名词，1852 年发现惯性定律，矩阵论创立者之一	★★★★★
7	凯莱（1821—1895），英国	1855 年引入定义矩阵乘法等运算，矩阵论创立者之一，西尔维斯特的朋友	★★★★★
8	雅可比（1804—1851），德国	重新发现并证明惯性定律	★★★
9	格拉斯曼（1809—1877），德国	1844 至 1862 年间创建高维线性空间理论，定义了线性相关、线性无关、维度、基、子空间、投影等	★★★★★
10	魏尔斯特拉斯（1815—1897），德国	1868 年完成二次型理论	★★★

详细的线性代数的发展历史请查阅本书后面的附录的第一部分——线性代数主要内容及其发展简史。

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多元数组，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，表面上看行列式和矩阵不过是一种速记的形式，但它的大多数生动的概念能为新的思想领域提供钥匙。

更具体的问题是，号称工程中强大工具的线性代数用途到底有哪些呢？

1.5线性代数有什么用

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。我稍微仔细地收集了一些学习线性代数的理由，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，必须学好线性代数。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，也必须学好线性代数，因为瑞典的L.戈丁说过，没有掌握线性代数的人简直就是文盲。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，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，也必须学好线性代数：

- 想搞数学，当个数学家（这个还需要列出来，谁不知道线性代数是数学）。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（参考本节后附的一份小资料）。

- 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性代数，因为它就是研究线性网络的主要工具；进行IC集成电路设计时，对付数百万个晶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，比如光调制器分析及研制需要张量矩阵，手机信号处理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性代数；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，电影大片《阿凡达》中大量的后期电脑特效制作没有线性代数这个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。

- 想当领导，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性不等式组的知识啊。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地安排各种商品的进货，以达到最大利润。

- 对于其他工程领域，没有用不上线性代数的地方。如搞建筑工程，那么奥运场馆鸟巢的受力分析需要线性代数的工具；石油勘探，勘探设备获得的大量数据所满足的几千个方程的方程组需要线性代数知识来解决；飞行器设计，就要研究飞机表面气流作用的过程，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；做餐饮业，构造一份有营养的减肥食谱也需要解线性方程组；知道有限元的名号吗？这个工程分析中十分有效的数学方法，其基础就是求解线性方程组。知道马尔科夫链的尊称吗？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。
- 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。有个老外科学家宣称，有振动的地方就有特征值，这句话曾让我国的数学网友毛骨悚然，大汗淋漓（网上语录）；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在各个工程领域里被用来把实验中得到的大量测量数据拟合到一个理想的直线或曲线上，其算法实质就是超定线性方程组的求解。线性代数中的二次型常常出现在工程（标准设计及优化）和信号处理（输出的噪声功率）的应用中，它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。
- 还有……

说实在的，咱也没有足够经验讲清楚线性代数在各个工程领域中的应用，只能大概人云亦云地讲述以上一些基本应用。因为如果要真正地讲清楚线性代数的一个应用，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，线性代数的概念和应用范围几乎可以涵盖所有的自然科学和工程技术领域。线性代数的理论是如此重要的科学概念，如果我们仍然坚持像当今的教学大纲做法一样只是把它作为计算工具进行练习，显然只是捡到了几粒芝麻（如果没有掌握如MATLAB一类的计算机工具软件的应用，我们捡到的是几粒陈谷子烂芝麻）！掌握线性代数的广泛应用案例、掌握线性代数的几何及物理意义、掌握线性代数的矩阵工具才是掌握线代的必由之路。

其实中国有水平的线性代数教材不少，比如本书后附的众多参考书籍。但俺最想推荐的适合初学者自学的教材是《线性代数及其应用》，这是美国David C. Lay 教授写的超棒的当代应用实例丰富的好教材，属于不看则损失巨大的读书事故。本书属于山枣野果之类，资以开胃，也热情建议偷空瞅那么一眼，属于不看白不看，看了不白看的情况。

附　来自金融危机时期的学好数学的利好消息

2009 年 1 月 26 日华尔街日报刊登一篇关于职业优劣评比的报导，标题为 "Doing the Math to Find the Good Jobs"，该文引述 Les Krantz 根据美国劳工统计局与人口普查局

(U.S.Bureau of Labor Statistics and the Census Bureau) 的资料所做的一项整理研究，依据五项指标：工作环境、所得、职业前景、体力要求和压力，针对200种职业进行综合评比排名。

美国的职业评比排名结果可能出乎大多数读者意料之外，现将排名最前和最后的十种工作抄录如下（数字代表排名）：

前十名职业：	后十名职业：
1. 数学家 (Mathematician)	200. 伐木工 (Lumberjack)
2. 保险统计师 (Actuary)	199. 酪农 (Dairy farmer)
3. 统计学家 (Statistician)	198. 出租车司机 (Taxi driver)
4. 生物学家 (Biologist)	197. 船员 (Seaman)
5. 软件工程师 (Software engineer)	196. 屋顶工 (Roofer)
6. 计算机系统分析师 (Computer systems analyst)	195. 清洁队员 (Garbage collector)
7. 历史学家 (Historian)	194. 焊工 (Welder)
8. 社会学家 (Sociologist)	193. 码头工 (Roustabout)
9. 工业设计师 (Industrial designer)	192. 钢架工 (Ironworker)
10. 会计师 (Accountant)	191. 建筑工 (Construction worker)

数学家之所以排名领先的部分原因是他们的工作环境相对舒适，不会接触有毒物质，不用提重物或弯腰爬行，而且收入颇丰。据估计，美国数学家的平均年薪达94，160美元。

文中提及19年前珍妮弗.寇特因为考虑低工作压力而选择研习数学，现年38岁的她，年收入高于前述平均薪资。她参与一个以数学为基础的计算机程序设计团队，过去所开发的程序曾经应用于电影<黑客任务 (The Matrix)>和<骇速快手 (Speed Racer)>。她在家里和同事以网络联系，绝少超时工作，没有什么工作压力。珍妮弗说：“解决问题牵涉许多思考，我发现那会让人冷静下来”。

(阅读原文 http: //online.wsj.com/article/SB123119236117055127.html)

第2章 向量的基本几何意义

向量的概念始终贯穿于当代科学中，也始终贯穿于线性代数中，所以我们不妨回顾一下中学学过的向量概念及其运算的几何意义，以期为更清晰地理解线性代数的几何本质。

2.1 向量概念的几何意义

2.1.1　自由向量的概念

向量（Vector）和标量的概念是发明四元数的爱尔兰数学家 W. R. 哈密尔顿给出的。向量是一个既有大小又有方向的量，这个量本身就是个几何的概念。我们常常把它与标量（只有大小的量）相区别。抓住向量的**大小**和**方向**这两个特征，一般用一个有向线段来表示一个向量(见图 2-1，显然向量本身就是一个几何图形)，记为 $\overrightarrow{AB}$。

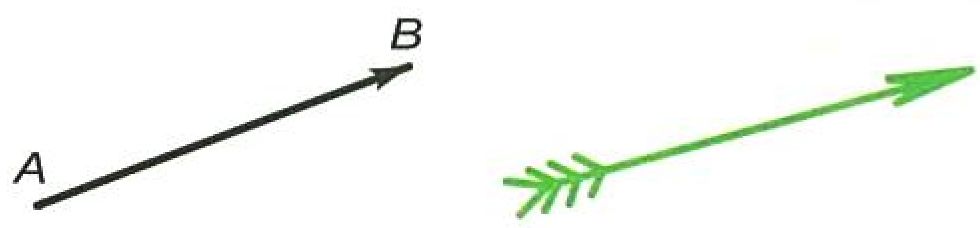


图 2-1 向量如箭，其几何图形表示为一个有向线段

在物理学中把向量叫矢量。矢就是箭，向量如一根箭一样有头部和尾部，箭在空间自由的飞行中箭杆的长度不会变，这一点与向量相同；同时箭在无重力作用的理想情况下方向也不改变。既不变长度也不变方向的理想之箭就是一个向量。所以向量的“飞行”称为平移，这种允许在一条直线上平移的向量称为自由向量（见图 2-2，物理学中常称为滑动向量）。

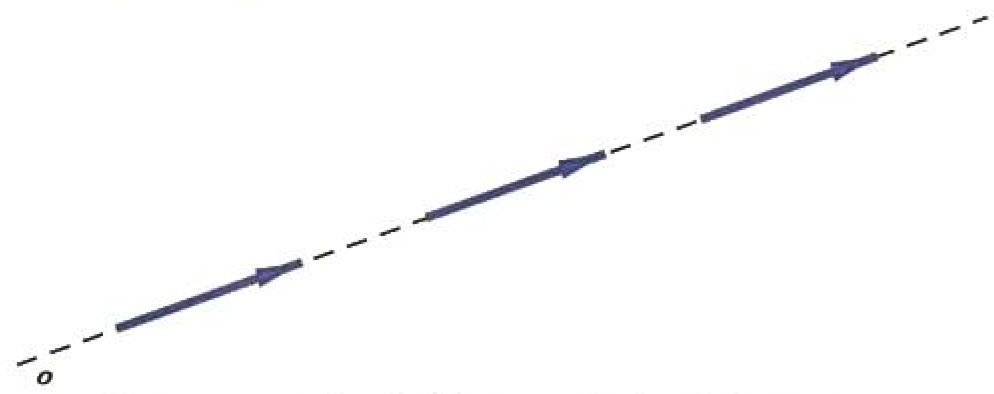


图 2-2 理想之箭是一个自由向量

沿着直线飞行的箭簇在每一时刻所表示的无数向量归属于同一个向量，这些无数的向量实际上是平行的向量。另外，还有不在一条直线上的平行而相等的向量，如下例：

考察一个刚体的平行移动。当刚体从一个位置平行移动到另一个位置时（比如说这个刚体是麦吉小姐过河坐的小船，见图 2-3，小船从河流的一边驶向对岸），刚体上各质点在同一时间段内有相同的位移，各点所画出的位移向量 $\boldsymbol{a}$ 有相同的大小和方向，它们每一个都反映了刚体位移的情况，因此刚体的平移运动可以用这些向量中的任一个来表示。基于这样的原因，凡是两个向量大小相等、方向相同的，我们就说这两个向量是相等的。因此，一个向量在保持长度和方向不变的条件下可以自由平移。如有必要，也可以将几个向量平移到同一个出发点或者坐标原点。

从上面的例子，我们感悟到自由向量为何是自由的。实际上，就是因为向量没有确定的位置，它们不依赖于任何坐标系而存在。因此从逻辑上看，无数的向量可能有相同的表述，所有

的这些向量都互相平行、相等，并具有相同的量值和方向。

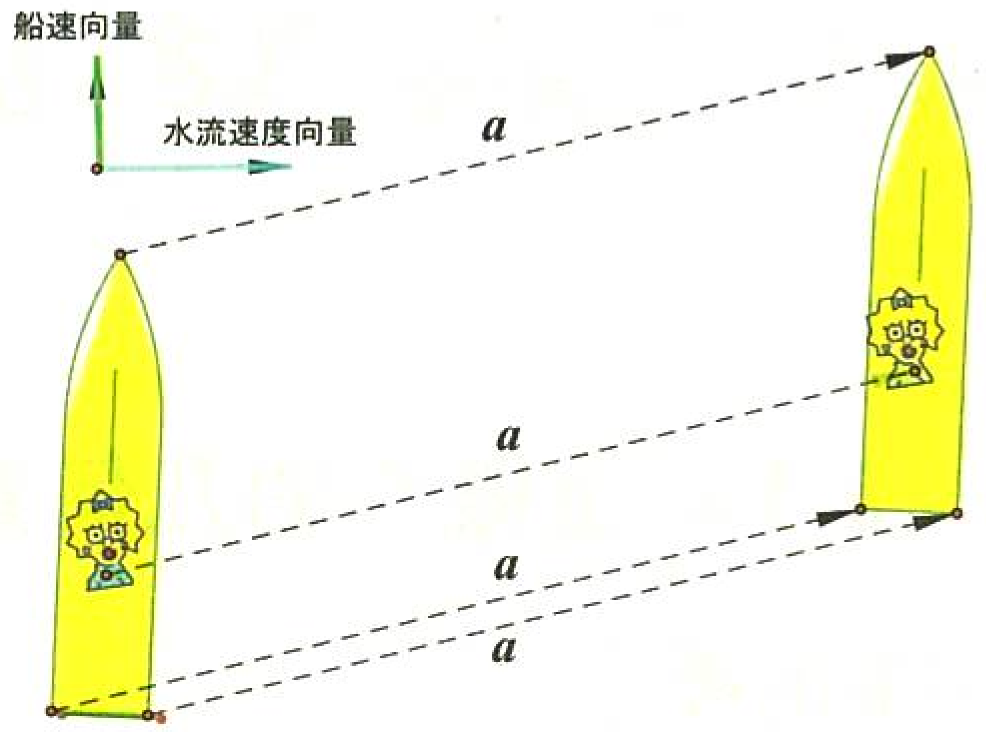


图 2-3 平移的向量也是一个自由向量

向量就是矢量

Vector 的中文译名是“向量”。顾名思义，就是指“既有方向又有大小的量”。可以说译得十分贴切。不过，中国在清末引进 Vector 概念时，物理学家称之为“矢量”，究其原因，早年的向量，只是物理学专门用来表示力和速度等物理量的工具，并不为数学家所重视。因为物理学用得多，矢量的译名自然流行。

物理界把 Vector 叫“矢量”，而数学界叫“向量”。其实是一回事。在早期，数学界也曾经把 Vector 定名为“矢量”，而物理界则把它定名为“向量”。后来，可能是为了尊重对方，却对换了一下，物理用矢量，数学用向量。矢量、向量的分歧，因为各自学科发展繁多，学科有分支，术语有派生，统一起来没那么容易，所以分歧一直维持到今。事实上，台湾物理界至今用的还是“向量”。

虽然两种译名并存，但由于向量在数学领域的大量应用之后，渐渐有取代“矢量”之势。

2.1.2 向量的代数表示

$\overline{AB}$ 是某一个向量的几何表示，它的代数表示一般是小写的黑体字母 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 或者 $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}$ 等。手写时，因为黑体的粗笔画书写不方便，所以常在字母上面加上箭头来与其它字母区别，如 $\vec{a}$、$\vec{b}$、$\vec{c}$。

以上的表示不便计算，如何对向量像数字一样进行运算呢？

在物理学科中，向量被处理为自由向量。虽然向量独立于任何坐标系之外，但为了与解析技术联系起来以实现对向量的计算，数学上我们还必须把向量放在某一个坐标系下来研究。**如果把空间中所有的向量的尾部都拉到坐标原点，这样 n 维点空间就可以与 n 维向量空间建立一一对应关系**：n 维点空间中的点(0，0，0，…)取作原点，那么每一个点都可以让一个向量和它对应，这个向量就是从坐标原点出发到这个点为止的向量。

其实，一旦我们确定好一个坐标系，一个向量就与一个点相对应，而点是用所谓坐标的有序数组表示的，因此我们就也可以把向量用有序数组表示。有了有序数组就可以运算了。使用有序数组或者解析式表述的向量是用以原点为起点的向量末端的坐标值表示，并把坐标值用圆括号括起来，如 $\boldsymbol{a}=(x,y,z)$。在这里这个有序数组 (x,y,z) 称之为向量。

在二维平面中，由原点引出的向量用两个有序实数表示；在三维空间中，由三个有序数表示三维向量。那么 n 维向量就可以由以上二维和三维向量的定义推广得到。虽然 n 维向量的几何意义难以想象，但其现实意义我们还是可以把握的。比如，在三维空间中，我们只要知道一

个球的球心位置和半径的大小就可以确定这个球面。把球心坐标和半径值写成有序数组，我们就得到了一个四维向量。

一个向量可以被分解为三个单位坐标向量的线性表示（实际上这个概念很重要，在今后的向量的运算和矩阵运算理解中起着关键作用）。例如，向量(1, 1, 1)分解如下：

$$(1,1,1)=(1,0,0)+(0,1,0)+(0,0,1)=\boldsymbol{i}+\boldsymbol{j}+\boldsymbol{k}$$

如图 2-4 所示，把单位坐标向量$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$分别首尾连接相加，就得到了(1, 1, 1)的图像。

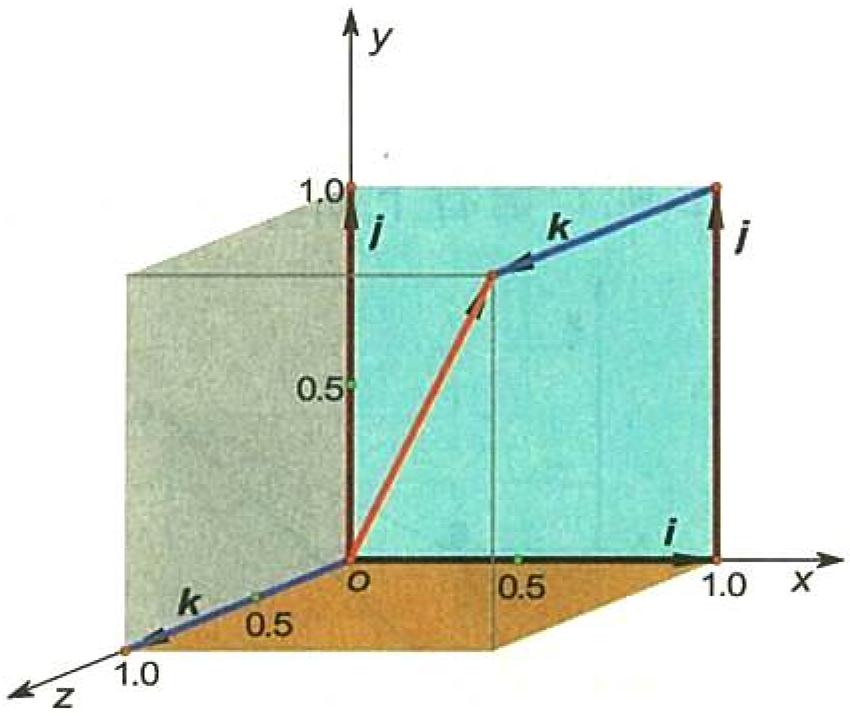


图 2-4 向量的分解

那么，任意一个向量$\boldsymbol{a}=(x,y,z)$就可以表示为$\boldsymbol{a}=(x,y,z)=x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}$，即单位坐标向量的线性表示。显然，分别对单位坐标向量进行缩放x、y、z倍然后相加，就得到了这个向量(x,y,z)的图像。和图 2-4 相似，我们就可以得到如图 2-5 所示的任意一个向量的分解图像。

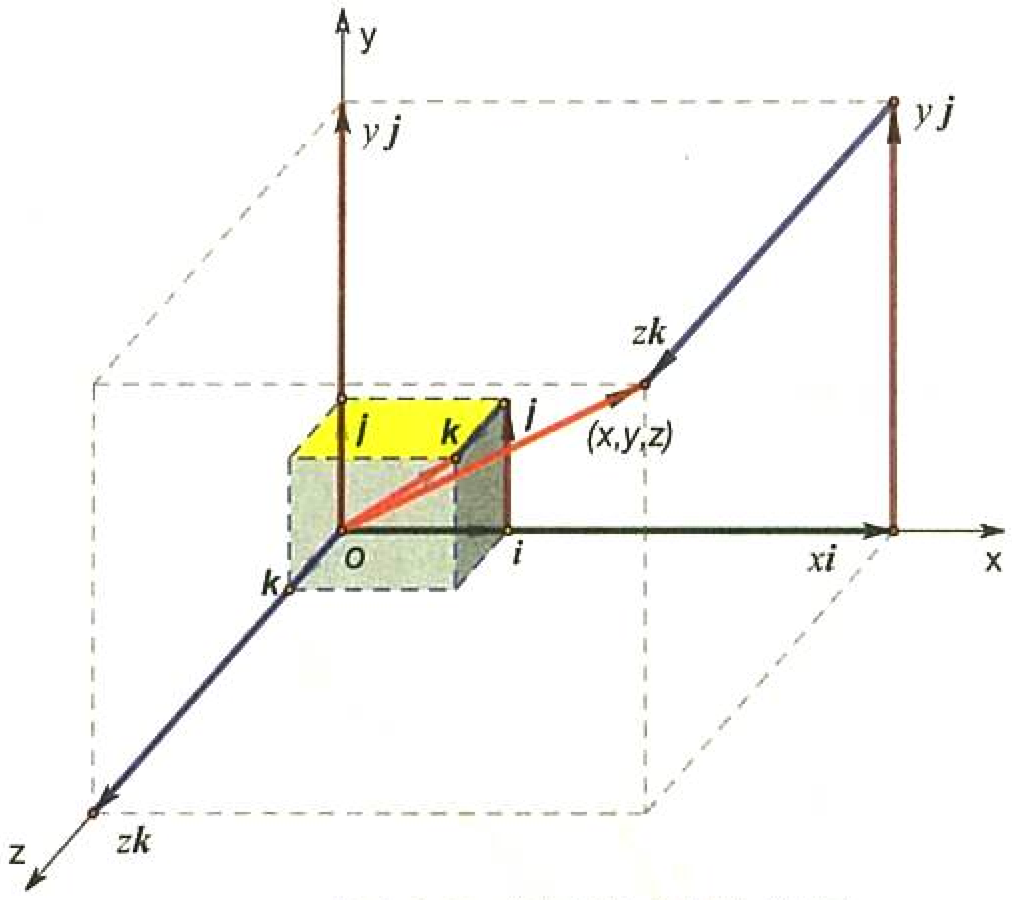


图 2-5 任意向量的分解

向量的运算有加法、减法和乘法，向量乘法有点积和叉积两种，但没有除法。下面我们分别介绍这些运算的细节。

向量与点的关系

向量被看做线性空间或向量空间中的一个元素。但向量与点不同，向量表示的是两点之间的位移而不是空间中的物理位置，它是独立于坐标系的，这就是为什么我们可以在描述向量的加法、数乘等运算的几何解释时常常不用画出坐标系，但一个点离开坐标系就无法表示；向量还可以确定方向，而一个点就不能。

向量允许进行线性运算（加法、数乘），但把两个点加起来则没有几何意义。

向量实际上是用一个点对来表示的，比如$\overline{AB}$，表示起点为A，终点为B。之所以把一个点

与一个向量相对应，是因为我们默认所有的向量是从原点出发的，不要忘了这个约定。

2.2 向量加法的几何及物理意义

设两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，它们的二维分量解析式为 $\boldsymbol{a}=(a_x,a_y)$，$\boldsymbol{b}=(b_x,b_y)$；三维分量的解析表达式为 $\boldsymbol{a}=(a_x,a_y,a_z)$，$\boldsymbol{b}=(b_x,b_y,b_z)$。则我们定义这两个向量的加法为 $\boldsymbol{a}+\boldsymbol{b}=(a_x+b_x,a_y+b_y)$，或者 $\boldsymbol{a}+\boldsymbol{b}=(a_x+b_x,a_y+b_y,a_z+b_z)$。向量加法的定义看起来很简单，就是两个向量的各分量分别对应相加形成了和向量的分量。

那么 $\boldsymbol{c}=\boldsymbol{a}+\boldsymbol{b}$ 的几何意义是什么呢？请看下面二维向量的图解（见图 2-6）。

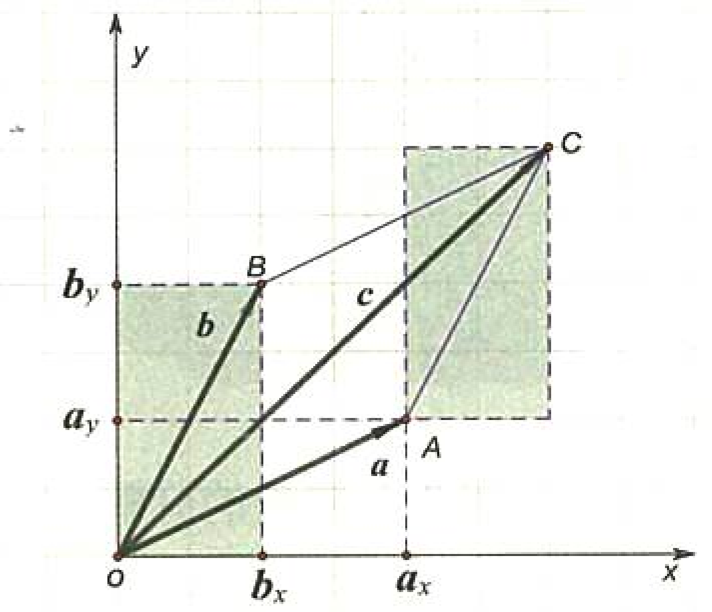


图 2-6 向量加法的平行四边形法则之几何解释

图 2-6 中的图形可以这样解读，表示向量 $\boldsymbol{b}$ 的分量的矩形□$\boldsymbol{ob}_x B\boldsymbol{b}_y$ 被斜上方向平移，放到表示向量 $\boldsymbol{a}$ 的分量的矩形□$\boldsymbol{oa}_x A\boldsymbol{a}_y$ 上面，两个矩形相接于 A 点，叠加后矩形的顶端 C 就是和向量的头端。

连接 BC、AC 后，□$oACB$ 是个平行四边形，和向量 $\boldsymbol{c}$ 是平行四边形的对角线，这就是**平行四边形法则**的几何解释。

当然，如果把向量 $\boldsymbol{b}$ 平移（平行移动）到 AC 的位置，与向量 $\boldsymbol{a}$ 的头部相接，就是三角形法则，见图 2-7。

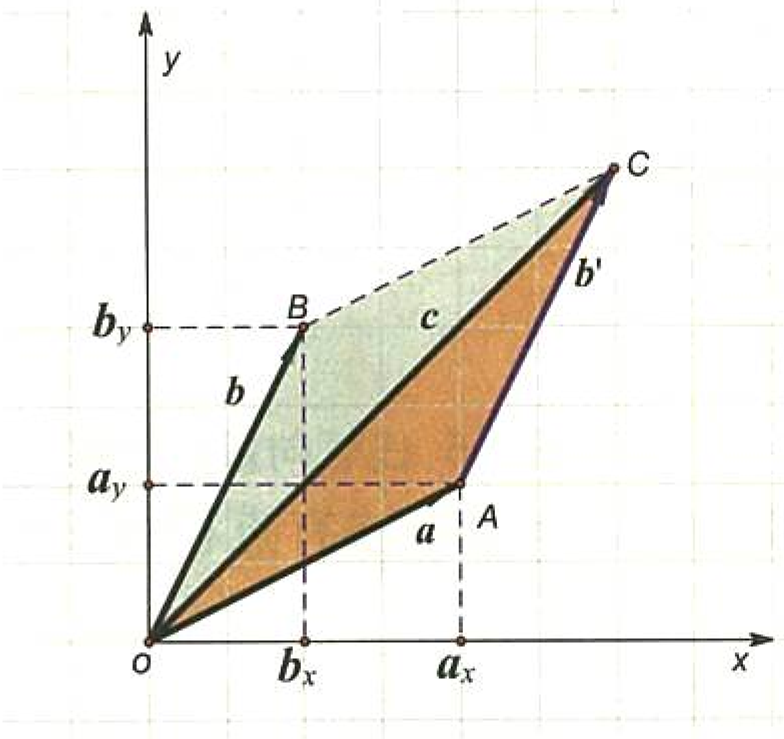


图 2-7 向量加法的三角形法则之几何解释

向量的所谓三角形或平行四边形法则不是人们凭空想当然的数学规定，而是从物理世界中抽象出来的向量运算法则。比如我们前面提到的船只过河的例子，船头指向的方向是船的马力驱动得到的位移 $\boldsymbol{S}_{motor}$（不考虑水流影响），水流的方向是水的冲击力对船造成的位移 $\boldsymbol{S}_{water}$（不考虑船的马力影响），那么，实际情况是船的真正位移 $\boldsymbol{S}_{boat}$ 是一条斜线，这条斜线就是 $\boldsymbol{S}_{motor}$ 和 $\boldsymbol{S}_{water}$ 的合成。它们的合成关系就是平行四边形的关系。

如果水的流速和船的马力不变，其中三个时刻（任意）的位移的合成图如图 2-8 所示。

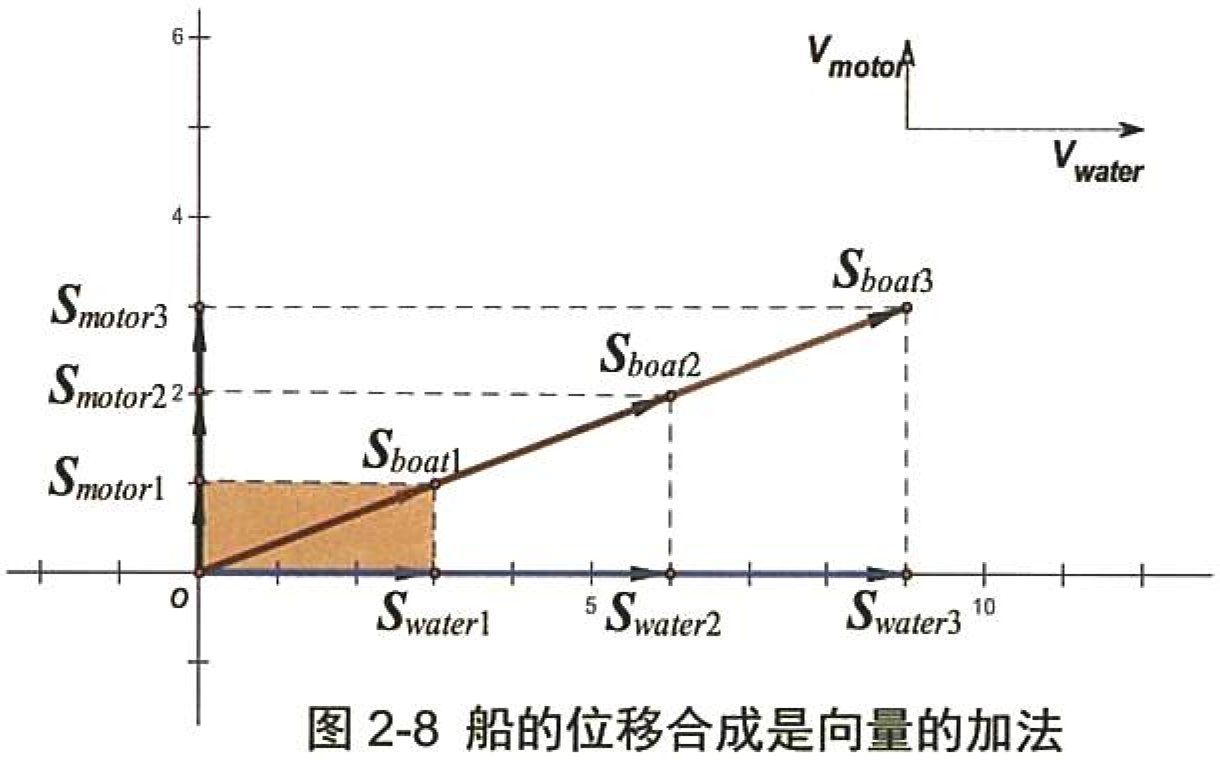


图 2-8 船的位移合成是向量的加法

如果水的流速不变，但在第二时刻和第三时刻船的马力逐步变大，那么三个时刻（任意）的位移的合成图如图 2-9 所示。

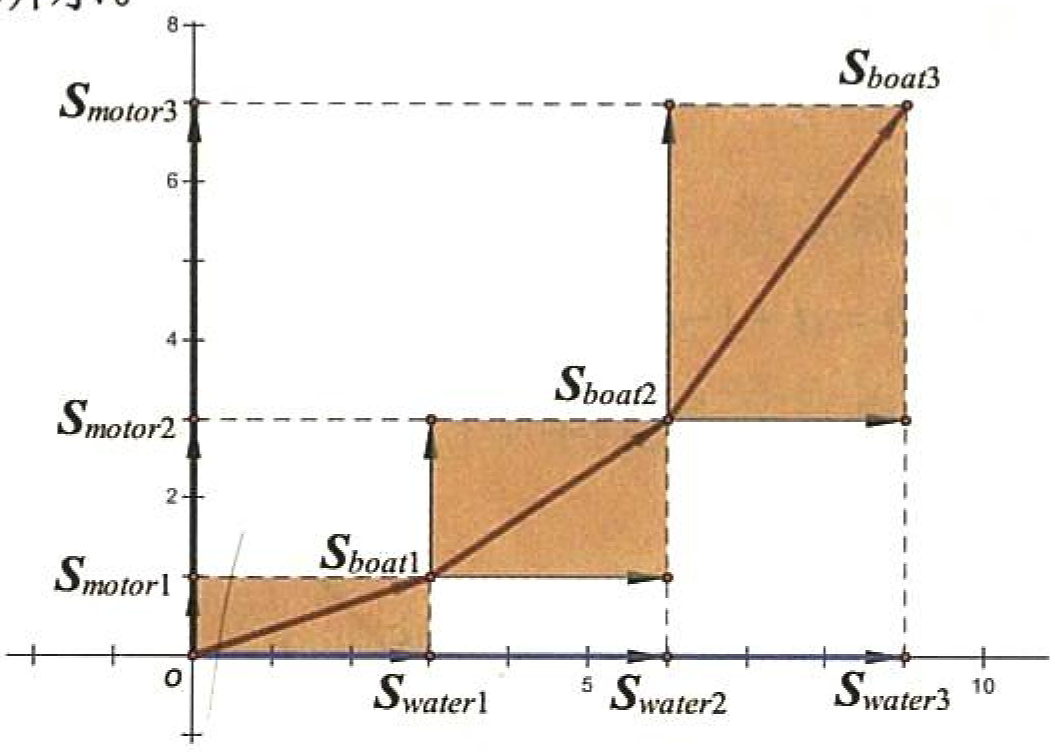


图 2-9 船的位移合成是向量的加法（另一种情形）

两个向量的加法叫做三角形法或者平行四边形法，那么多个向量的加法同样也满足这些法则，并可以由三角形法则得到多个向量的多边形法则。下边我们画出多个向量的加法和减法的图例。

图 2-10 是把 $\boldsymbol{a}$ 、$\boldsymbol{b}$ 、$\boldsymbol{c}$ 、$\boldsymbol{d}$ 四个向量按照三角形法则相加的图例，在图中，我们把 $\boldsymbol{a}$ 、$\boldsymbol{b}$ 、$\boldsymbol{c}$ 、$\boldsymbol{d}$ 四个向量依次首尾相接，直到画完所有向量，最后只是把第一个向量的尾部 o 指向最后一个向量的首部 P，画出的向量就是 4 个向量的和。这个画法可以称做多向量的多边形加法法则。

多边形法则很容易从三角形法则推导出来，如图 2-11 中增加的和向量即是应用三角形的向量法则画出了中间向量逐次相加的结果，最后推出了图中的多边形法则。

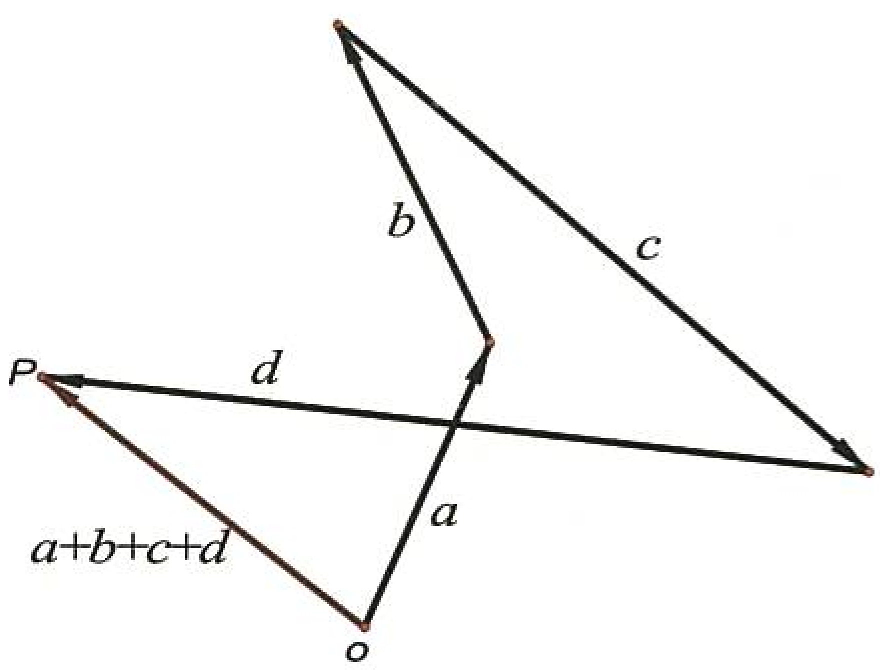


图 2-10 多向量的多边形加法法则

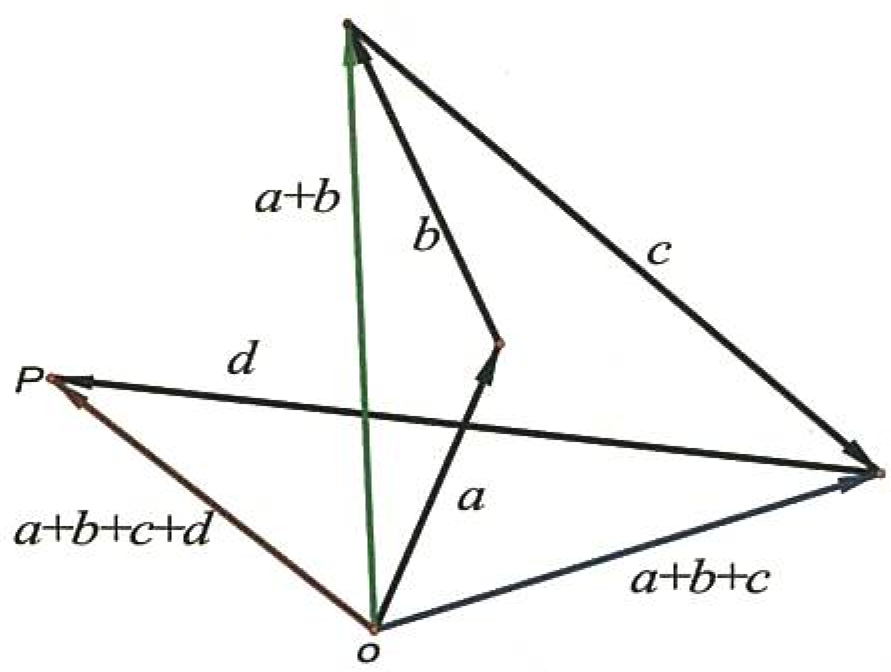


图 2-11 用三角形法则推导出多边形法则

多变形法则体现在船只过河的例子就是把船的每时刻的位移进行合成，如图 2-12 所示。

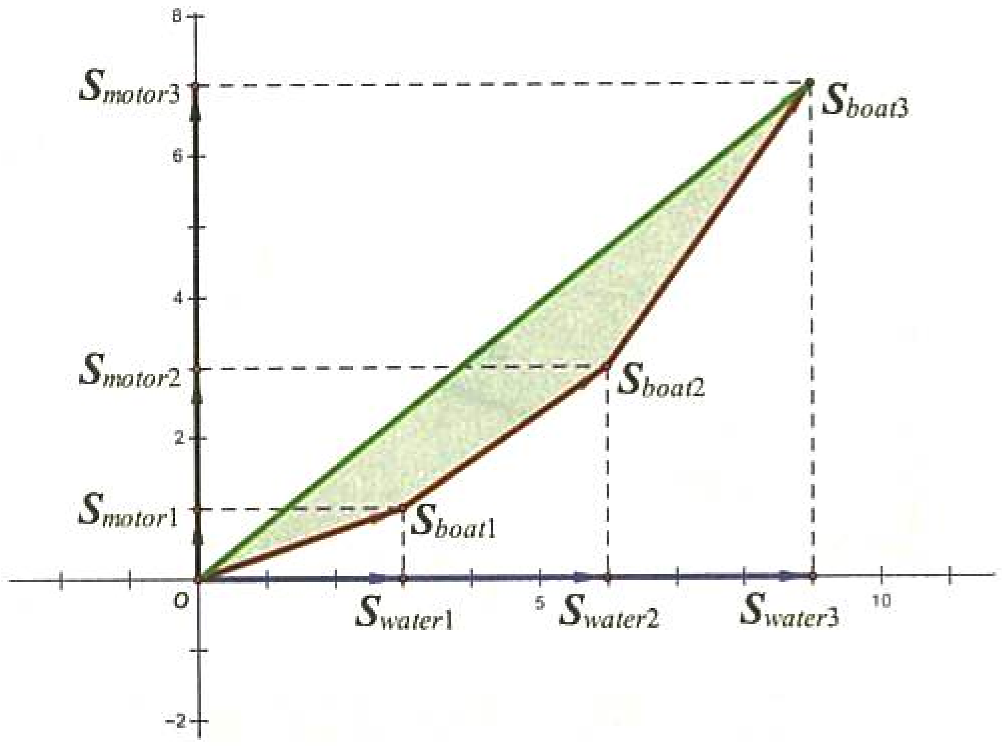


图 2-12 船的位移的合成就是多向量加法

多向量加法的数学本质，实际上是这些向量在坐标轴上（以 o 点为坐标原点的坐标系）的投影（或坐标分量）的合成（相加或相减）后的结果。向量的更高一级的运算如点积、叉积的定义也是这个数学本质的体现。

关于减法，实际上是加法的特殊形式，是加法的逆运算。向量减法，我们可以用定义加法的方式定义减法，例如定义 $\boldsymbol{c}=\boldsymbol{a}-\boldsymbol{b}=\boldsymbol{a}+(-\boldsymbol{b})$，只要把被减向量 $\boldsymbol{b}$ 反向后再与向量 $\boldsymbol{a}$ 相加即可。实际上在平行四边形法则中，和向量与差向量构成平行四边形的两个对角线(见图 2-13)。

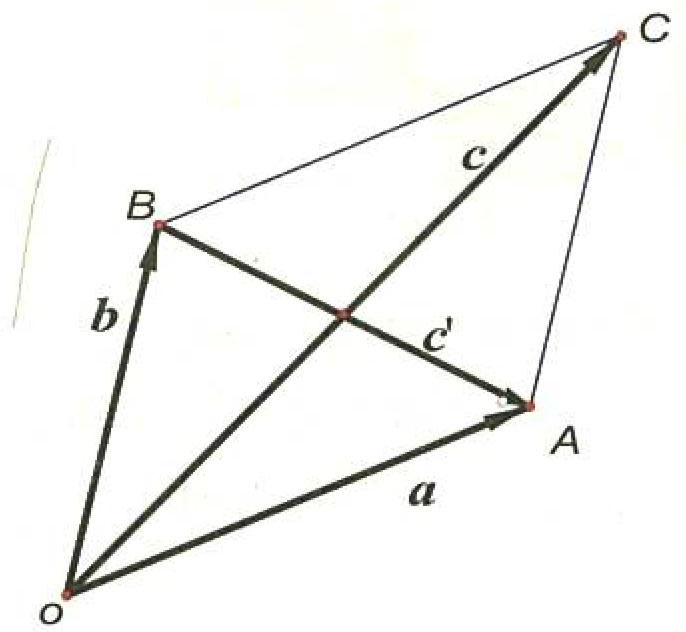


图 2-13 和向量与差向量构成平行四边形法则的两个对角线

多向量的减法也是多次使用三角形法则得到的，如图 2-14 所示。图中分别给出了多向量的加减法分解为多次的三角形法则。

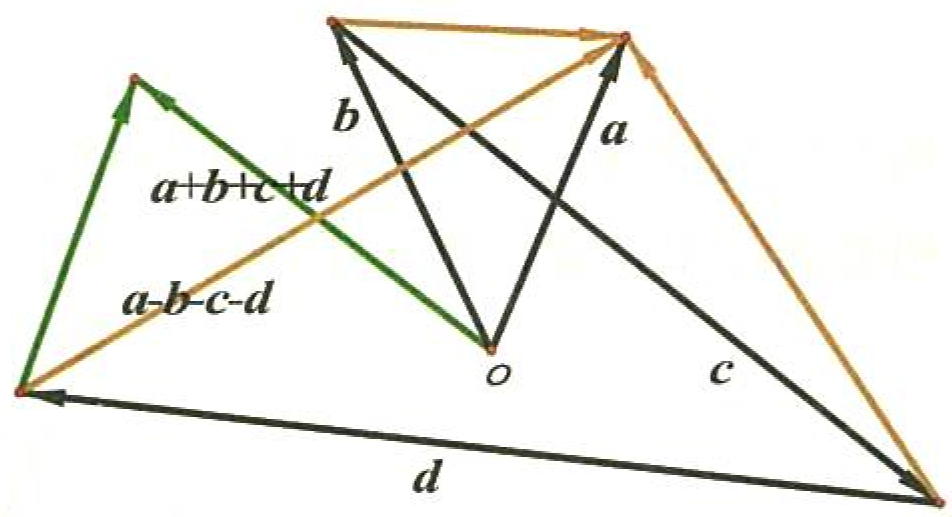


图 2-14 多向量加减法的多边形法则对比

格拉斯曼与 n 维向量

海曼·格拉斯曼是德国的几何学家，他 1844 年发表的《延拓论》创立了现今的 n 维几何学。格拉斯曼在构建 n 维几何代数理论时是以一个非常简单的公式 $AB+BC=AC$(见图 2-15（a）)作为研究起点的。他发现，如果不考虑线段点 A、B、C 的顺序，只要不把 AB、BC 这样的因子仅仅理解为长度，并且赋予它们“方向”(例如 $BA=-AB$)，公式依然正确。举个例子：如果 C 位于 A 和 B 之间（见图 1-15（b）)，那么 $AB=AC+CB$，但是由于 $CB=-BC$，我

们将发现 $AB=AC-BC$。此时只要在这个公式两边简单地加上BC就能得到最初的公式 $AB+BC=AC$。

图 2-15 直线上的有向线段加法

也就是说图 1-15 的两个图例都有 $AB+BC=AC$ 成立。

把图 1-15 的两根直线中间的 B 点和 C 点各自向上拉伸一些距离，等式 $AB+BC=AC$ 依然成立（见图 2-16），我们刚刚知道这就是向量的三角形加法法则。

图 2-16 三角形上的有向线段加法——向量加法法则

继续向空间中拉伸就是三维向量乃至 n 维向量的加法了，等式同样成立。

这是一个非常有价值的发现。但格拉斯曼对向量的延伸和拓展更加让人吃惊，他拓展到向量的外积运算。比如，如图 2-17 所示的正方形或平行四边形 $ABCD$，如果正方形或平行四边形的面积 $S_{ABCD}=AB\cdot AD$，那么就有 $AB\cdot DA=-S_{ABCD}$ 成立。这正是 2.4 节将要介绍的向量的外积（叉积）运算及其几何意义。格拉斯曼由此最终发明了一种全新的被称之为外积代数的几何理论。

图 2-17 有向线段的乘积——向量叉积运算

2.3 向量内积的几何和物理意义

2.3.1 向量内积的几何解释

向量的内积也叫数量积、标积、点积（点积的名称来自于内积的计算符号·）等，都是一个意思，就是内积的结果是个数量或者标量。内积的定义有两个，下面我们把它们列举出来并探讨一下它们的关系。以三维向量为例：

$$\boldsymbol{a}\cdot\boldsymbol{b}=ab\cos\theta \tag{2-1}$$

$$\boldsymbol{a}\cdot\boldsymbol{b}=a_xb_x+a_yb_y+a_zb_z \tag{2-2}$$

公式（2-1）是说向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的长度之积再乘以它们之间的夹角的余弦；公式（2-2）的意思是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的坐标分量分别对应乘积的和。

定义内积有很多好处，除了物理上的直接应用外，至少我们还可以应用这个定义（即公式（2-2））去计算一个向量的长度（在已知它的坐标时）。比如我们求向量 $\boldsymbol{a}$ 的长度：

$$a=\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}=\sqrt{a_xa_x+a_ya_y+a_za_z}=\sqrt{a_x^2+a_y^2+a_z^2}$$

作为内积定义的两个公式有关系吗？当然有！你来看：

假设我们选一个这样的坐标系（见图 2-18），使 x 轴沿向量 $\boldsymbol{a}$ 的方向，原点重合于向量起

点，那么 $a_x = a$， $a_y = a_z = 0$，代入式（2-2）得

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_x b_x + a_y b_y + a_z b_z = ab_x$$

ab_x 的含义是 $\boldsymbol{a}$ 的长度乘以 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 方向上（亦即 x 轴方向）的分量，这个 b_x 分量就是 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影。我们知道投影 $b_x = b\cos\theta$，因此公式 $\boldsymbol{a} \cdot \boldsymbol{b} = ab\cos\theta$ 得证（如果它对一个坐标系成立，则对所有的坐标系都成立）。

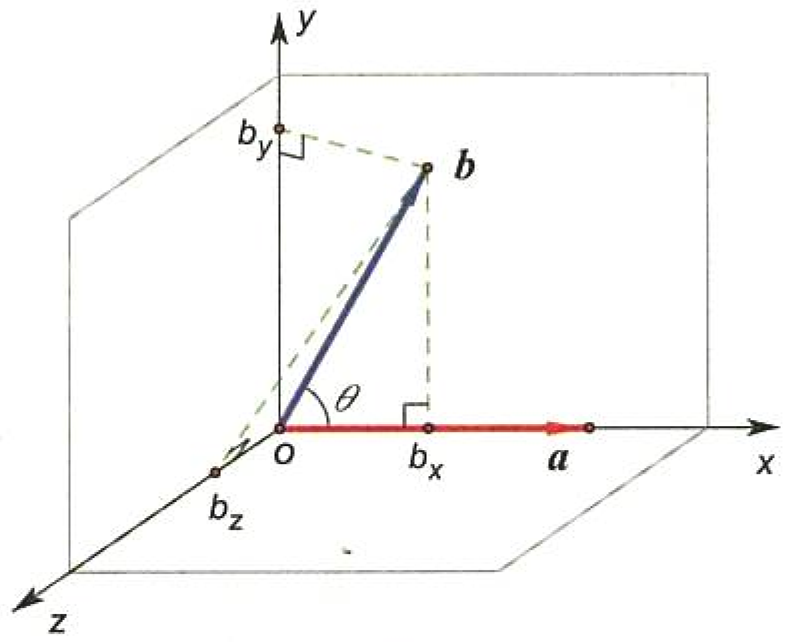


图 2-18 在特定坐标系下向量内积的计算

因此，向量内积的几何解释就是一个向量在另一个向量上的投影的积，也就是同方向的积。

特别地，如果一个向量如 $\boldsymbol{a}$ 是某个坐标轴的单位坐标向量，那么，两个向量的内积 $\boldsymbol{a} \cdot \boldsymbol{b}$ 就是向量 $\boldsymbol{b}$ 在此坐标轴上的坐标值。因此，如果想要将一个向量变换到新的坐标系，那么只要对新坐标系轴向量进行内积运算即可。这个结论很重要，这是傅立叶分析的理论基础之一。

实际上，矩阵的乘法运算本质上也是行向量和列向量的内积运算。两个矩阵相乘，根据矩阵乘积的定义，就是左矩阵的行（向量）与右矩阵的列（向量）进行逐次内积。

另外，对两个向量的内积可以有其他的几何解释，这些解释在应用上就显得比较直观。比如，从内积值上我们可以看出两个向量在方向上的接近程度。当内积值为正值时，两个向量大致指向相同的方向（方向夹角小于 90°）；当内积值为负值时，两个向量大致指向相反的方向（方向夹角大于 90°）；当内积值为 0 时，两个向量互相垂直（这个性质经常在向量几何中作为判断直线与直线是否垂直的依据）。笼统来说，内积值越大，两个向量在方向上的就越接近，内积值越小，两个向量在方向上就越相反。

图 2-19 中，向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}_1$ 的方向最接近，大致相同，夹角为锐角，内积为正；向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}_2$ 的方向垂直，内积为 0；向量 $\boldsymbol{a}$ 与向量 $\boldsymbol{b}_3$ 的方向大致相反，夹角为钝角，内积为负。同样地，向量 $\boldsymbol{b}_1$ 与向量 $\boldsymbol{b}_2$，向量 $\boldsymbol{b}_2$ 与向量 $\boldsymbol{b}_3$，向量 $\boldsymbol{b}_3$ 与向量 $\boldsymbol{b}_4$ 方向都大致相同，夹角为锐角，它们各自的内积亦为正。

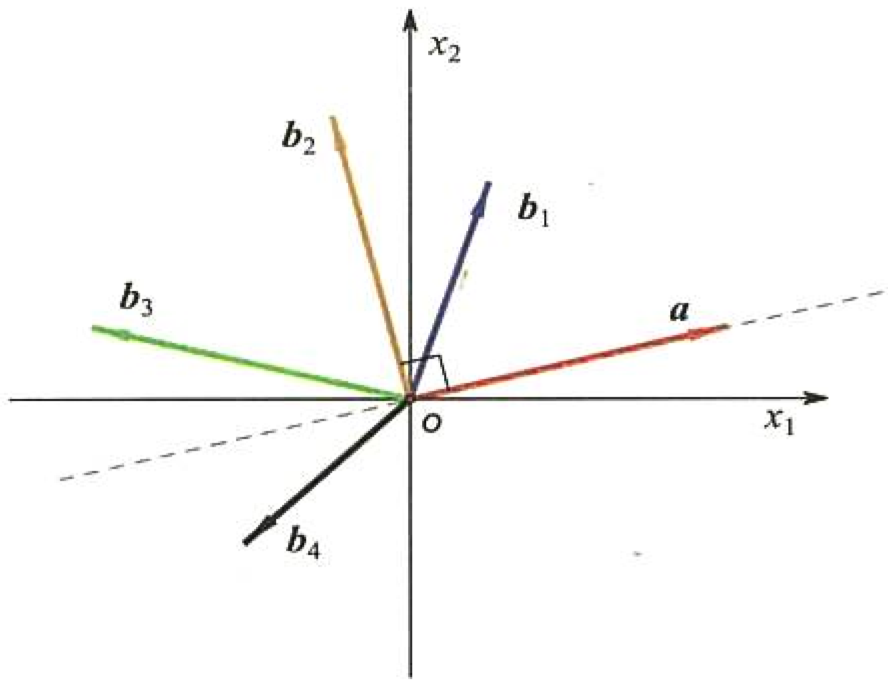


图 2-19 内积大小可确定两个向量的在方向上的接近程度

2.3.2　向量内积的物理解释

所谓向量的内积，其实就在我们身边。比如我上周购买的食物的价格向量是 $\boldsymbol{p}$ =（蔬菜 2 元/斤，大米 1.5 元/斤，牛肉 10 元/斤，啤酒 3 元/瓶），消耗的数量向量为 $\boldsymbol{d}$ =（3.5 斤，5 斤，2 斤，3 瓶）；那么我上周的饮食消费就是向量 $\boldsymbol{p}$ 和 $\boldsymbol{d}$ 的内积：

$$\boldsymbol{p}\cdot\boldsymbol{d}=(2, 1.5, 10, 3)\cdot(3.5, 5, 2, 3)=7+7.5+20+9=43.5\text{ 元}$$

另外，内积的一个经典物理例子就是当一个物体从某处被拉到另一处时所做的功，下面我们把这个做功的图画出来印证以上内积两大公式的一致性。

假设在一个斜坡上用力 $\boldsymbol{F}$ 斜上方拉一个物体，位移为 $\boldsymbol{S}$，那么这个力 $\boldsymbol{F}$ 所做的功为（分量的分解见图 2-20（a））：

$$W=\boldsymbol{F}_x\boldsymbol{S}_x+\boldsymbol{F}_y\boldsymbol{S}_y$$

另外，我们也可以把力 $\boldsymbol{F}$ 沿着 $\boldsymbol{S}$ 的方向和垂直 $\boldsymbol{S}$ 的方向（按照图 2-20（b））进行分解，那么这个力 $\boldsymbol{F}$ 所做的功又可表示为

$$W=\boldsymbol{F}_s\boldsymbol{S}=\boldsymbol{F}\boldsymbol{S}\cos\theta$$

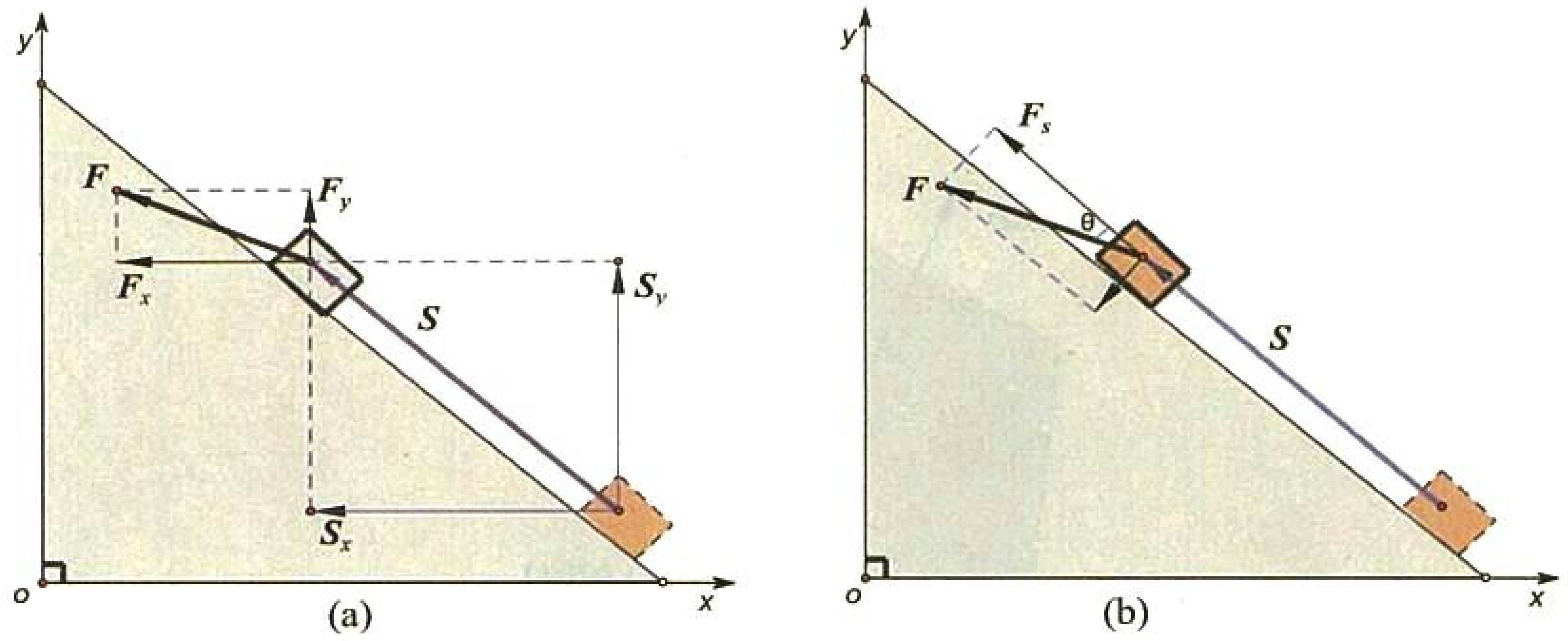


图 2-20 力做功的向量分解

由此，我们从物理原理上印证了内积两大公式的一致性。

用向量求解余弦差角公式

余弦差角公式是中学学过的三角公式，咱用向量重新证明一下，借此感受向量工具简洁有效的威力。如图 2-21 所示，采用向量分解的方法，单位圆上的半径向量 $\overrightarrow{OA}$、$\overrightarrow{OB}$，其在 x 轴和 y 轴上的投影长度，就是余弦和正弦。于是，任意角的三角函数的符号变化就显得一清二楚，正弦、余弦函数的图象也很容易画出。

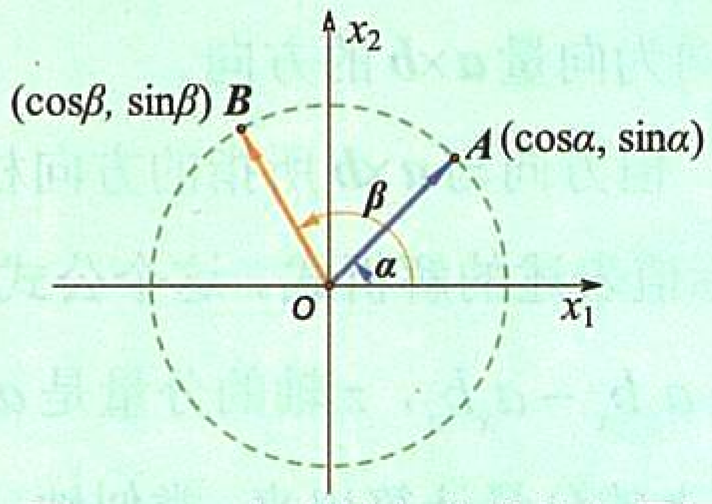


图 2-21 余弦差角公式的向量证明方法

特别是，由向量的内积公式导出余弦的差角公式简直是举手之劳，简单极了：

由向量内积的两个等价定义式得到向量 $\overrightarrow{OA}$ 和 $\overrightarrow{OB}$ 内积一个等式：

$$\overrightarrow{OA}\cdot\overrightarrow{OB}=|\overrightarrow{OA}|\cdot|\overrightarrow{OB}|\cos(\beta-\alpha)=\cos\alpha\cos\beta+\sin\alpha\sin\beta$$

因为 $$\left|\overrightarrow{OA}\right|=\left|\overrightarrow{OB}\right|=1$$

所以 $$\cos(\beta-\alpha)=\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta$$

2.4 向量叉积的几何和物理意义

2.4.1 叉积的定义及其几何解释

向量叉积（Cross product）又译为交叉积（交叉积的名称来自于其运算规则，因为两个向量作叉积运算时，是把向量的元素交叉相乘；当然其计算符号 $\boldsymbol{a}\times\boldsymbol{b}$ 刚好也是叉叉），也可称为外积，因为叉积会产生新的一维向量。两个向量确定了一个二维的平面，叉积又会产生垂直于这个平面的向量。

叉积的定义也有两个，下面我们把它们列举出来并探讨一下其关系。

设三维空间中的两个向量为 $\boldsymbol{a}=(a_x,a_y,a_z)$，$\boldsymbol{b}=(b_x,b_y,b_z)$，则

$$\boldsymbol{a}\times\boldsymbol{b}=(ab\sin\theta)\boldsymbol{n}_0 \text{（其中 } \boldsymbol{n}_0 \text{ 是垂直于 } \boldsymbol{a} \text{ 和 } \boldsymbol{b} \text{ 展成平面的单位法向量）} \tag{2-3}$$

$$\boldsymbol{a}\times\boldsymbol{b}=(a_yb_z-a_zb_y,a_zb_x-a_xb_z,a_xb_y-a_yb_x) \tag{2-4}$$

公式（2-3）是叉积几何意义的定义式。

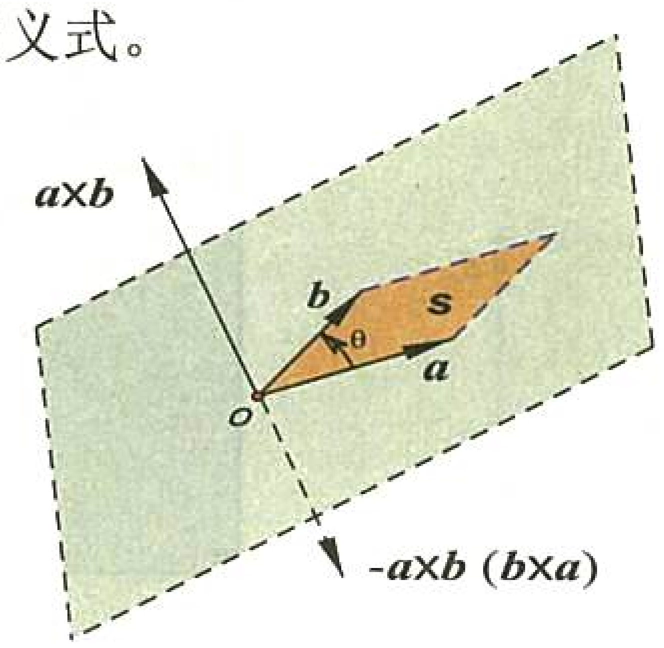


图 2-22 用右手法则定义叉积的方向

$\boldsymbol{a}\times\boldsymbol{b}$ 为一个新生成的向量，这个向量垂直于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 展成的平面（图 2-22 中的灰色大平行四边形，由线段 $o\boldsymbol{a}$ 和 $o\boldsymbol{b}$ 所确定的平面），向量的大小等于以 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 为邻边所张成的平行四边形（图中的深色小平行四边形）的面积 $\boldsymbol{S}$；

垂直于平面有两个方向，我们规定用右手法则来确定叉积的方向：按照乘式 $\boldsymbol{a}\times\boldsymbol{b}$ 的运算顺序，右手的四指平直指向第一个向量 $\boldsymbol{a}$，然后弯曲指向向量 $\boldsymbol{b}$（从向量 $\boldsymbol{a}$ 沿着 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 间较小夹角转向向量 $\boldsymbol{b}$），则右手大拇指的指向为向量 $\boldsymbol{a}\times\boldsymbol{b}$ 的方向。

由此向量 $\boldsymbol{b}\times\boldsymbol{a}$ 也垂直这个平面，但方向与 $\boldsymbol{a}\times\boldsymbol{b}$ 所指的方向相反，即 $\boldsymbol{b}\times\boldsymbol{a}=-\boldsymbol{a}\times\boldsymbol{b}$。

公式（2-4）是用向量的三维坐标值表述的解析式。这个公式的表面含义是叉积 $\boldsymbol{a}\times\boldsymbol{b}$ 的 x 轴的分量是 $a_yb_z-a_zb_y$，y 轴的分量是 $a_zb_x-a_xb_z$，z 轴的分量是 $a_xb_y-a_yb_x$。可以看出，叉积 x 方向的分量由向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 在 yoz 平面上的分量计算出来；类似地，$\boldsymbol{a}\times\boldsymbol{b}$ 的 y 方向的分量由向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 在 xoz 平面上的分量计算出来，$\boldsymbol{a}\times\boldsymbol{b}$ 的 z 方向的分量由向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 在 xoy 平面上的分量计算出来。

换句话说，两个三维的向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的叉积 $\boldsymbol{a}\times\boldsymbol{b}=(c_x,c_y,c_z)$ 在空间中张成一个平行四边形面积块，它在三个二维坐标平面上的投影就是这两个向量叉积的三个分量。比如，第三个叉积分

量 $c_z = a_x b_y - a_y b_x$ 就是在 xoy 平面上的投影，如图 2-23 中 xoy 平面上的灰色平行四边形。

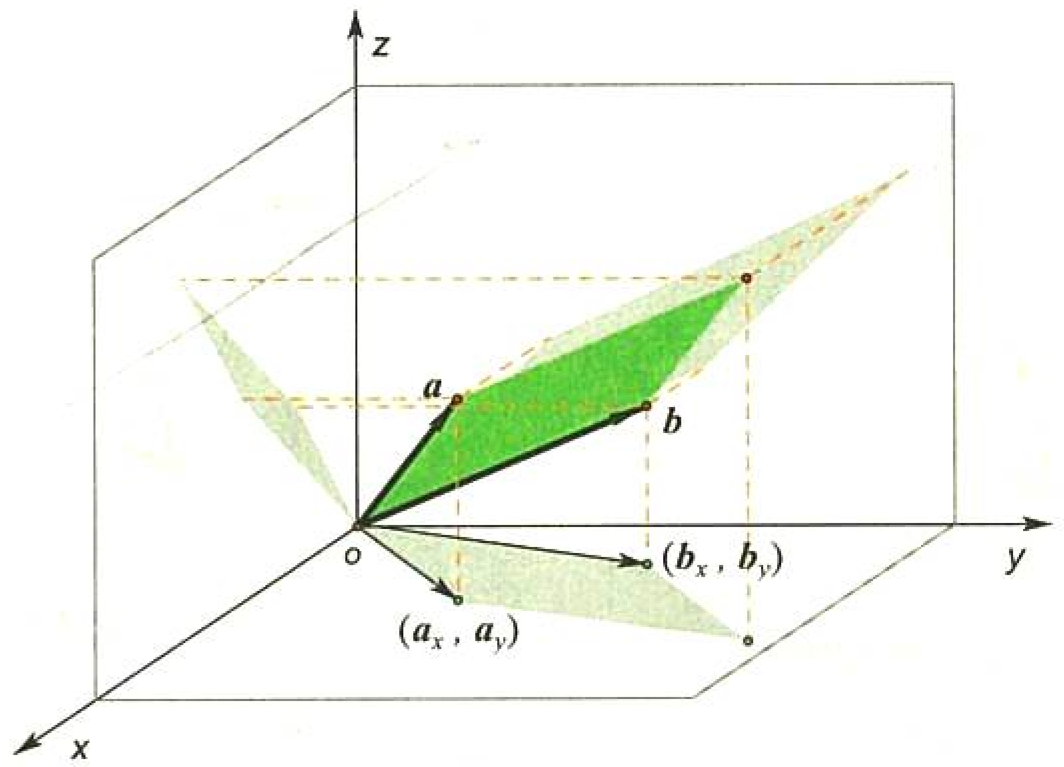


图 2-23 叉积的分量是叉积到三个坐标平面上的投影

实际上，叉积的这三个分向量分别又是三个叉积的结果，是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 投影的叉积。例如，向量 $\boldsymbol{a}$ 在 xoy 平面上的投影是 (a_x, a_y)，向量 $\boldsymbol{b}$ 在 xoy 平面上的投影是 (b_x, b_y)（见图 2-23），那么向量投影的叉积就是 (a_x, a_y) 和 (b_x, b_y) 的叉积，亦即 $(a_x, a_y)\times(b_x, b_y) = a_x b_y - a_y b_x$。

两个向量的叉积只能定义在三维空间中

这个规定看起来让人惊异，确实向量的叉积有点另类，不太数学，它不能推广到高维的空间中。在三维空间里，根据右手法则，定义两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的一个叉乘 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{c}$ 仍然是一个向量，新向量 $\boldsymbol{c}$ 垂直于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 所张成的平面并由右手定则唯一地被确定。但在四维以上空间里无法定义两个向量的叉积，比如在四维空间里，与四维向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 所张成的平面之外与之正交的向量有无穷多（一个平面都是），因此无法以几何方式唯一确定一个向量与 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 相联系，就是说，不能通过一个作图法由 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 唯一地确定一个向量 $\boldsymbol{c}$ 使其在刚体运动下不变。

虽然在四维空间里无法定义两个向量的叉积，但我们可以定义三个向量的连叉积 $\boldsymbol{a}\times\boldsymbol{b}\times\boldsymbol{c}$，因为这样叉积的方向又可以唯一确定了。所以，在 n 维空间里，可以有 n-1 个向量的向量积的公理化定义。作为拉普拉斯展开定理的代数余子式的几何意义，我们将在行列式一章中讨论它。

2.4.2 叉积的物理意义

叉乘的定义看起来有点怪，大家可能感觉到，叉积向量好像不是太真实似的，特别是方向定义的显然是人为的。实际上，叉积这种向量与一般力啊位移啊等的向量确实不同，所以在物理学中又被称为赝向量或轴向量。

但这个定义也是从物理应用方面得来的。举个例子：知道陀螺的原理吗？高速旋转的陀螺会定向（见图 2-24（a））。陀螺所定义的方向就是矢径向量 $\boldsymbol{r}$ 和线速度 $\boldsymbol{v}$ 叉乘积的方向，即角速度 $\boldsymbol{\omega}$ 方向——与旋转面垂直（注：物理中三者关系定义为 $\boldsymbol{v}=\boldsymbol{\omega}\times\boldsymbol{r}$。由三向量相互正交，可以推导得 $\boldsymbol{\omega}=\dfrac{\boldsymbol{r}\times\boldsymbol{v}}{|\boldsymbol{r}|^2}$ 或 $\boldsymbol{r}\times\boldsymbol{v}=|\boldsymbol{r}|^2\boldsymbol{\omega}$。我们可以仿照力矩的定义来命名 $|\boldsymbol{r}|^2\boldsymbol{\omega}$ 为一个新变量“转矩”，显然转矩越大陀螺越稳定、定向越好）。

类似的一个例子是螺钉（见图 2-24（b）），螺钉只要左右向旋转即可在螺孔中前进或者后退。用螺丝刀把这颗螺钉按照 $\boldsymbol{F}$+的方向右旋，那么旋转时的扭力向量 $\boldsymbol{F}$ 和矢径向量 r 这两个

叉乘的结果即是力矩 $\boldsymbol{M}$ 的方向，这颗螺钉就会沿着力矩 $\boldsymbol{M}$ 在螺母孔内前进，反方向就会改变叉积的方向进而退出螺孔（右螺旋螺钉）。也就是力矩或叉乘向量的方向就是螺钉的螺旋前进的方向，这个方向垂直于螺丝刀口和扭力的方向，也就是垂直于被叉积的两个向量的方向。

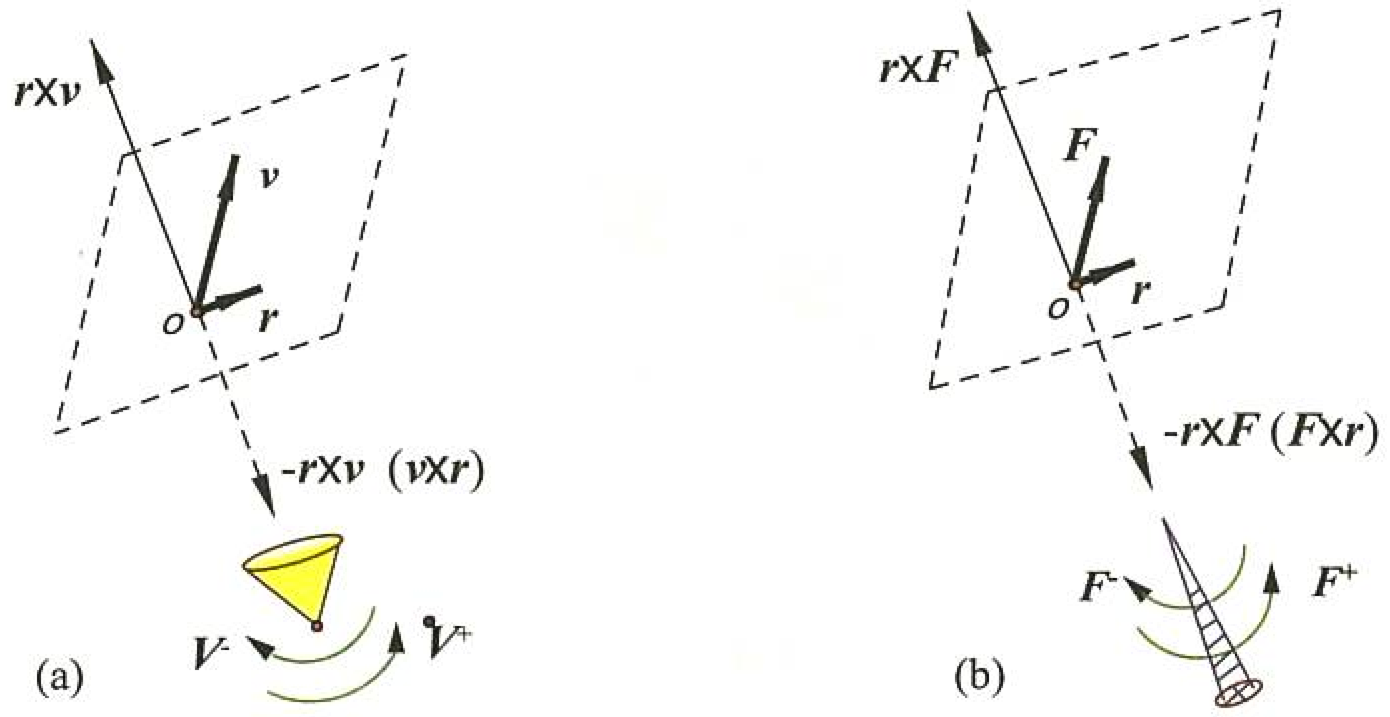


图 2-24 陀螺、螺钉与叉积

力矩就是向量叉积。还有点疑问？好，弄个夸张一点的。我们把螺钉的原理稍微改变一下：假如有一个 100 米长的细钢棒（好长），钢棒架在几个支架上，钢棒一端装有摇臂。当有人用摇臂扭转钢棒时，这个扭转的力（就是力矩）会沿着这个长长的钢棒一直延伸到钢棒的尾端，并且整个钢棒上都有扭转的力存在，无论我们碰触钢棒的任何部位都会感知到这个力矩的存在。这个扭转的力是多大呢？如果摇动的人用力越大，横向摇臂越长，这个力矩沿着纵向轴就会越大，我们就越难用手抓停它（如图 2-25）。

可以想象，力矩沿着 100 米长的、与摇臂和摇动的力垂直的方向，无处不在！

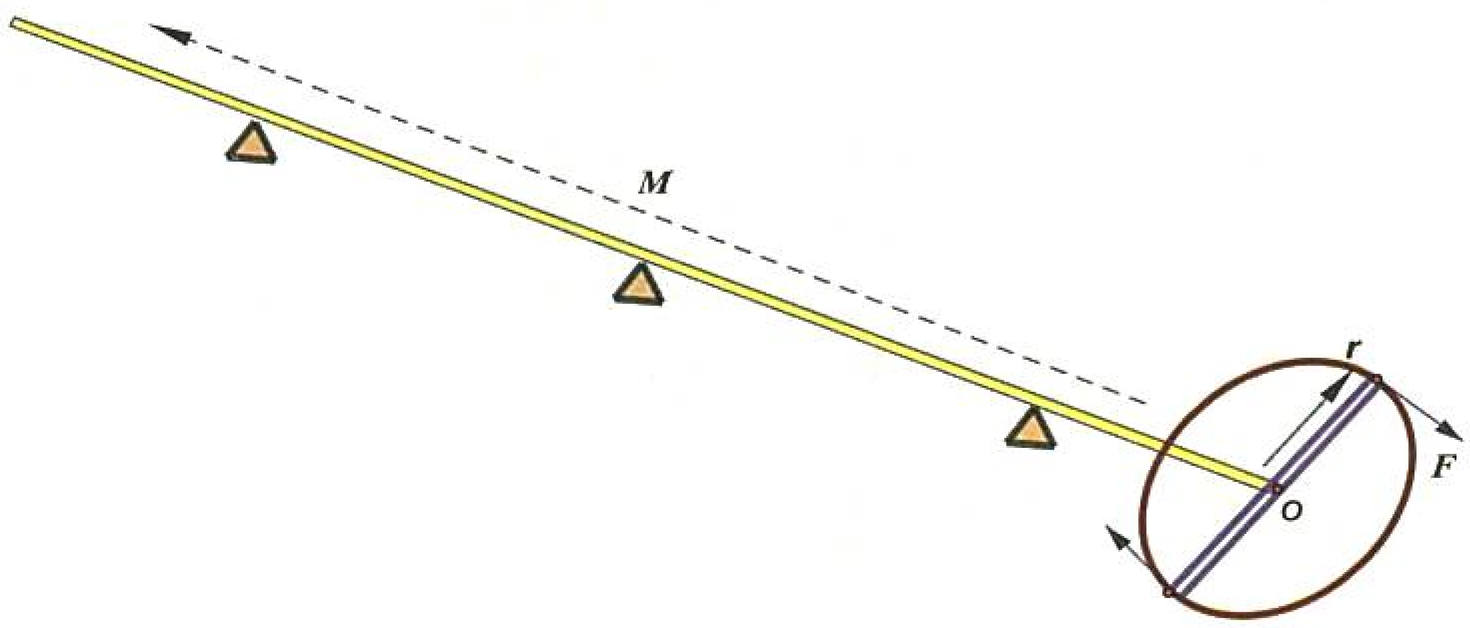


图 2-25 力矩的方向

下面看一看叉积解析式的物理意义的分解。同样，我们也举扭矩的例子。这里再次把叉积解析公式重新列在这里：

$$\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{b}=(a_yb_z-a_zb_y,a_zb_x-a_xb_z,a_xb_y-a_yb_x)$$

我们说， 一个三维向量可以分解为沿着 x、y、z 轴的分向量，或者一个三维向量可以看作是三个分别与坐标轴同向的分向量之和。即：

$$\boldsymbol{v}=(x,y,z)=x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}$$

这里我们同样可以认为 $(a_yb_z-a_zb_y)\boldsymbol{i}$ 是 x 方向的向量，$(a_zb_x-a_xb_z)\boldsymbol{j}$ 是 y 方向的向量，$(a_xb_y-a_yb_x)\boldsymbol{k}$ 是 z 方向的向量。

前面讲叉积的这三个分向量分别又是三个叉积，从何讲起？下面我们以叉积的 z 轴分量的

$(a_xb_y - a_yb_x)\boldsymbol{k}$ 来比对物理上力矩的概念。

假设，在三维空间坐标系下，以我们的书面为 xoy 平面，z 轴垂直书面并指向我们。以这个三维坐标系的原点 o 为转动固定点，在空间力 $\boldsymbol{F}'$ 的作用下对杠杆作趋势为逆时针的转动。

当我们逆着 z 轴向下看时，就会看到空间杠杆转动在平面 xoy 上的投影：(假设) 平面上的杠杆投影向量为 $\boldsymbol{r}$，平面上的扭力投影向量为 $\boldsymbol{F}$，那么力对杠杆的叉积 $\boldsymbol{r}\times\boldsymbol{F}$ 的方向将指向我们。下面来看看这个图解。

图 2-26 中在 xoy 坐标系的四个象限都画出了力矩的一般情况，并且把叉乘的两个向量 $\boldsymbol{r}$ 和 $\boldsymbol{F}$ 都进行了分解(继续投影——向 x 和 y 坐标轴投影)：扭力 $\boldsymbol{F}$ 分解为 x 轴和 y 轴的分量 $\boldsymbol{F}_x$ 和 $\boldsymbol{F}_y$，杠杆 $\boldsymbol{r}$ 也分解为 x 轴和 y 轴的力臂分量 $\boldsymbol{r}_x$ 和 $\boldsymbol{r}_y$。在初中我们就知道力矩等于力乘以力臂，力臂与扭力垂直。

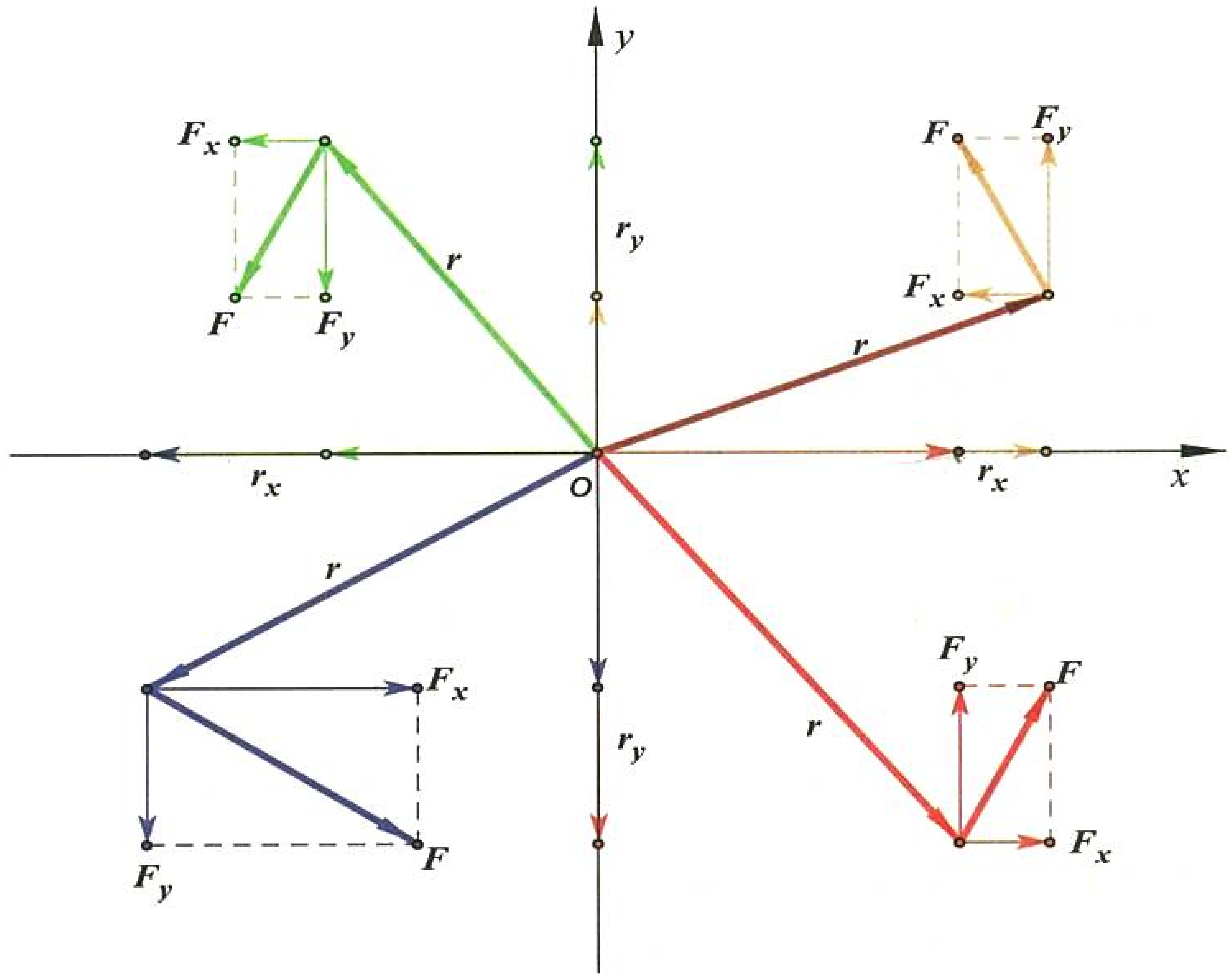


图 2-26 在四个象限下向量 F 和 r 的分解

显然，分解后力臂和扭力的这些分量 $\boldsymbol{r}_x$、$\boldsymbol{r}_y$ 和 $\boldsymbol{F}_x$、$\boldsymbol{F}_y$ 要么平行要么垂直。平行的力臂和扭力的力矩为零，即 $\boldsymbol{r}_x\times\boldsymbol{F}_x=\boldsymbol{r}_y\times\boldsymbol{F}_y=0$，因此，我们对分解的分向量进行叉积求力矩，只有 $\boldsymbol{M}=\boldsymbol{r}_x\times\boldsymbol{F}_y+\boldsymbol{r}_y\times\boldsymbol{F}_x$。所以，我们分别得出第一象限的力矩是 $\boldsymbol{M}_1=\boldsymbol{r}_x\times\boldsymbol{F}_y+\boldsymbol{r}_y\times\boldsymbol{F}_x = xF_y\boldsymbol{k}+y(-F_x)\boldsymbol{k}=(xF_y-yF_x)\boldsymbol{k}$（参见图 2-26）。这里，$\boldsymbol{k}$ 是指向 z 坐标方向的单位向量。

类似地，其它三个象限的力矩分别是

$$\boldsymbol{M}_2=\boldsymbol{r}_x\times\boldsymbol{F}_y+\boldsymbol{r}_y\times\boldsymbol{F}_x=(-x)(-F_y)\boldsymbol{k}+y(-F_x)\boldsymbol{k}=(xF_y-yF_x)\boldsymbol{k}$$
$$\boldsymbol{M}_3=\boldsymbol{r}_x\times\boldsymbol{F}_y+\boldsymbol{r}_y\times\boldsymbol{F}_x=(-x)(-F_y)\boldsymbol{k}+(-y)F_x\boldsymbol{k}=(xF_y-yF_x)\boldsymbol{k}$$
$$\boldsymbol{M}_4=\boldsymbol{r}_x\times\boldsymbol{F}_y+\boldsymbol{r}_y\times\boldsymbol{F}_x=xF_y\boldsymbol{k}+(-y)F_x\boldsymbol{k}=(xF_y-yF_x)\boldsymbol{k}$$

我们看，四个象限的力矩表达式相同，都是 $(xF_y-yF_x)\boldsymbol{k}$。

在这个向量 $(xF_y-yF_x)\boldsymbol{k}$ 中，(x,y) 是矢径向量 $\boldsymbol{r}$ 的坐标，(F_x,F_y) 是扭力向量 $\boldsymbol{F}$ 的坐标；在 xoy 平面上，向量 $\boldsymbol{r}$ 和向量 $\boldsymbol{F}$ 的叉积的大小等于 xF_y-yF_x，方向 $\boldsymbol{k}$ 指向 z 轴方向。

显然，我们把向量 $\boldsymbol{r}$ 和向量 $\boldsymbol{F}$ 改写成通用向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ ，这个结果就变成了 $(a_x b_y - a_y b_x)\boldsymbol{k}$ 。

类似地，我们可以推出 x 轴、y 轴方向的叉积表达式如前所述。

这就是叉积计算公式的物理意义。

实际上，我们把坐标系重新选择或者把坐标系右旋一个合适的角度，就可以得到叉积定义的另一个公式 $\boldsymbol{a}\times\boldsymbol{b}=(ab\sin\theta)\boldsymbol{n}_0$ 。

如图 2-27(a)所示，两个向量 $\boldsymbol{r}$ 和 $\boldsymbol{F}$ 在 xoy 坐标系中 $\boldsymbol{F}$ 分解为 $\boldsymbol{F}_x$ 和 $\boldsymbol{F}_y$ ，$\boldsymbol{r}$ 也分解为 $\boldsymbol{r}_x$ 和 $\boldsymbol{r}_y$ ；如果把 xoy 坐标系右旋一个 φ 角，变为 $x'oy'$ 坐标系，刚好使 x' 轴与向量 $\boldsymbol{r}$ 重合（见图 2-27(b)）。

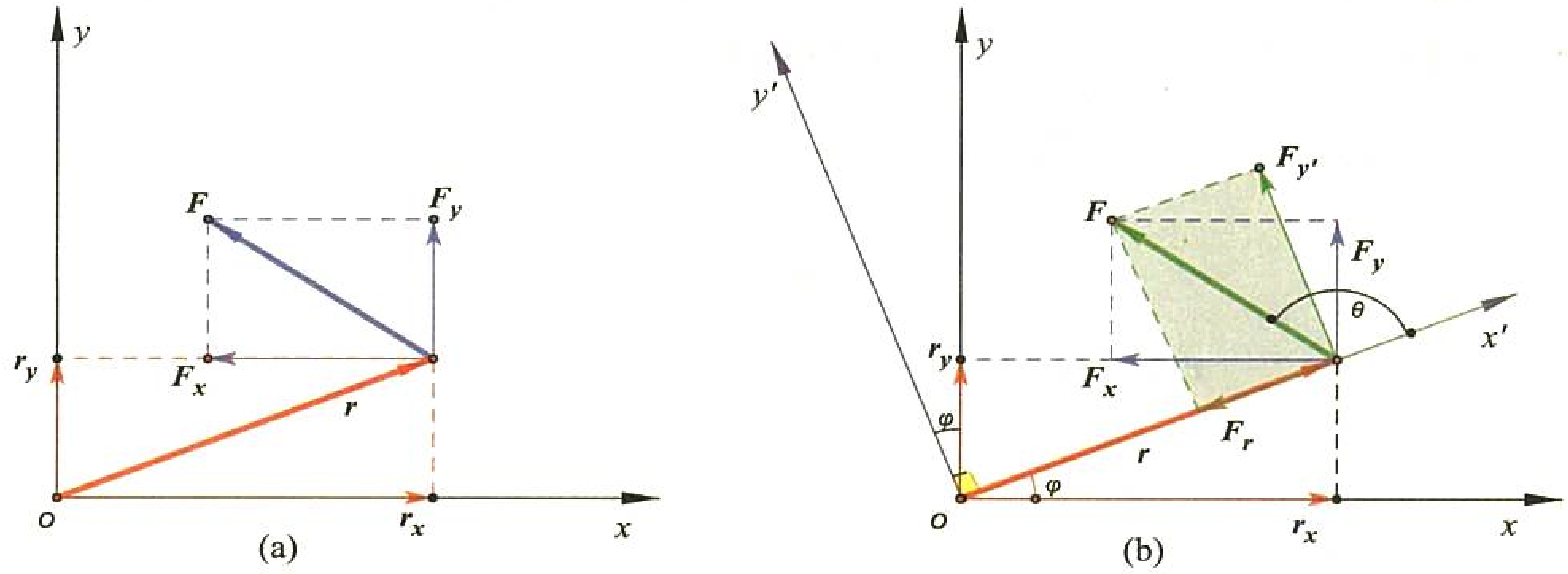


图 2-27 两个叉积定义式是等价的

显然，在新的坐标系下，$\boldsymbol{r}$ 不必分解分量了；$\boldsymbol{F}$ 只需分解为 $\boldsymbol{F}_{x'}$ 和 $\boldsymbol{F}_{y'}$，则新坐标系下的叉积（z 方向的力矩 $\boldsymbol{M}$）为

$$\boldsymbol{M}_z=\boldsymbol{r}\times\boldsymbol{F}=\boldsymbol{r}\times(\boldsymbol{F}_{x'}+\boldsymbol{F}_{y'})=\boldsymbol{r}\times\boldsymbol{F}_{x'}+\boldsymbol{r}\times\boldsymbol{F}_{y'}$$

又因为 x' 轴与 $\boldsymbol{r}$ 重合，所以 $\boldsymbol{r}\times\boldsymbol{F}_{x'}=0$；且 $\boldsymbol{r}$ 到 $\boldsymbol{F}$ 的夹角为 θ，所以上式继续等于

$$\boldsymbol{M}_z=\boldsymbol{r}\times\boldsymbol{F}_{y'}=rF\sin(\pi-\theta)\boldsymbol{k}=rF\sin\theta\,\boldsymbol{k}$$

如果我们把向量 $\boldsymbol{r}$ 和 $\boldsymbol{F}$ 改写成通用向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ ，且不是强调在 xyz 三维坐标下，那么 z 向的单位向量 $\boldsymbol{k}$ 可以写成与叉乘的两个向量垂直且满足右手系的 $\boldsymbol{n}_0$。至此得到了叉乘的第二个公式

$$\boldsymbol{a}\times\boldsymbol{b}=(ab\sin\theta)\boldsymbol{n}_0 。$$

2.5 向量混合运算的几何意义

我们所讨论的向量的混合运算包括向量加法和乘法的混合运算，仔细研究一些混合运算的几何意义有助于我们理解向量的几何本质。

2.5.1 向量加法的结合律的几何解释

三个向量加法的结合律为

$$(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c}).$$

其图解是显然的，图 2-28(a)给出了 $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}$ 的加法图解，图 2-28(b)给出了 $\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$ 的图解，图 2-28（c）把图（a）和（b）两者叠放在一起，显示了两个加法有相同的结果。

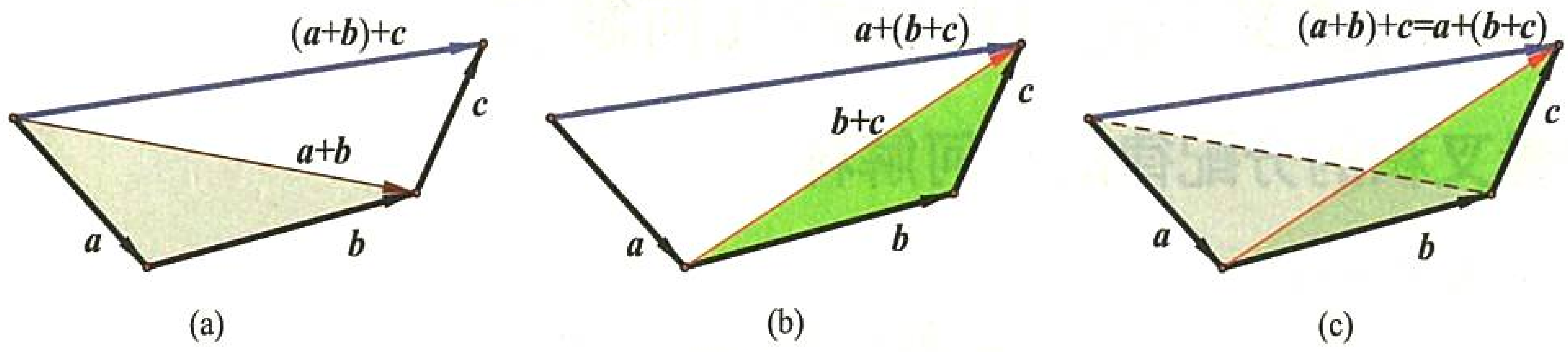


图 2-28 向量加法的结合律

2.5.2　向量数乘的分配律的几何解释

数乘两个向量和的分配律为

$$k(\boldsymbol{a}+\boldsymbol{b})=k\boldsymbol{a}+k\boldsymbol{b}$$

其图解是显然的，图 2-29 中设数量 k 大于 1，$\boldsymbol{a}+\boldsymbol{b}$ 的加法三角形和 $k\boldsymbol{a}+k\boldsymbol{b}$ 的加法三角形是两个相似三角形，因而得到如图 2-29 所示的图形。

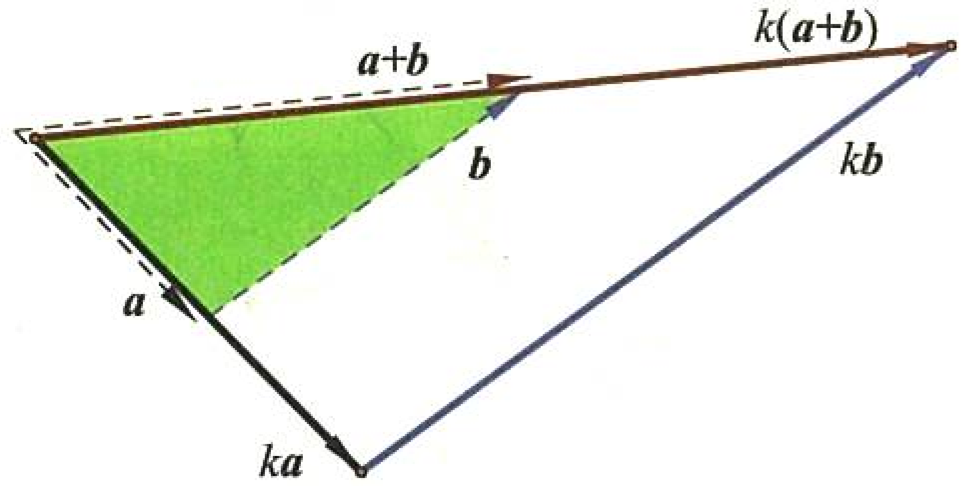


图 2-29 向量数乘的分配律

2.5.3　向量点积的分配律的几何解释

向量点积的分配律为

$$\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}$$

如图 2-29 所示，$\overrightarrow{OB'}$ 为向量 $\boldsymbol{b}$ 在向量 $\boldsymbol{a}$ 上的投影，$\overrightarrow{B'C'}$ 为向量 $\boldsymbol{c}$ 在向量 $\boldsymbol{a}$ 上的投影，$\overrightarrow{OC'}$ 为向量 $\boldsymbol{b}+\boldsymbol{c}$ 在向量 $\boldsymbol{a}$ 上的投影。

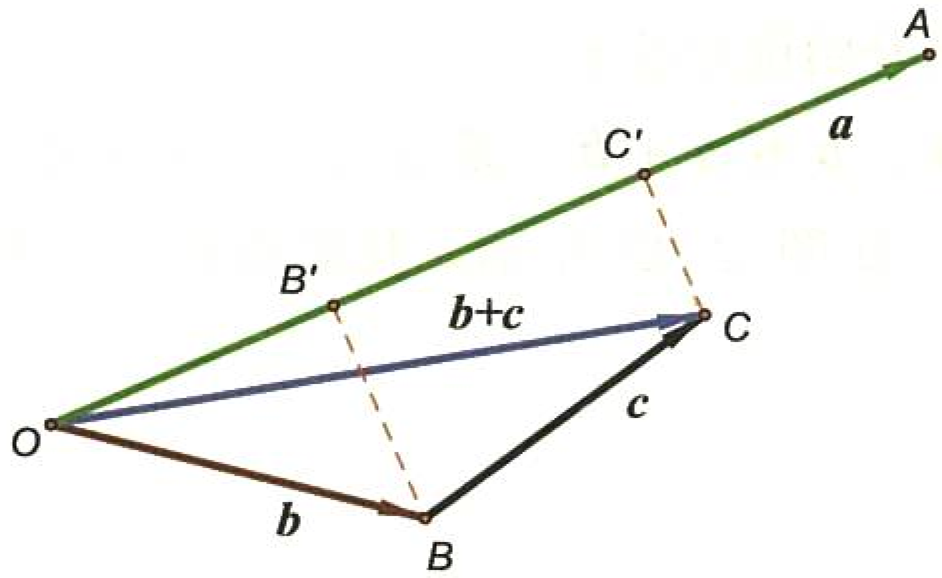


图 2-30 向量点积的分配律

由图有

$$\boldsymbol{a}\cdot\boldsymbol{b}=\overrightarrow{OA}\cdot\overrightarrow{OB'},\quad \boldsymbol{a}\cdot\boldsymbol{c}=\overrightarrow{OA}\cdot\overrightarrow{B'C'}$$

于是得

$$\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}=\overrightarrow{OA}\cdot\overrightarrow{OB'}+\overrightarrow{OA}\cdot\overrightarrow{B'C'}=\overrightarrow{OA}\cdot\left(\overrightarrow{OB'}+\overrightarrow{B'C'}\right)=\overrightarrow{OA}\cdot\overrightarrow{OC'}$$

而又有 $\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})=\overrightarrow{OA}\cdot\overrightarrow{OC'}$，因此有分配律 $\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}$ 成立。

2.5.4 向量叉积的分配律的几何解释

1. 向量叉积的分配律的几何解释 1

向量叉积的分配律为

$$(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}$$

为理解这个分配律的几何意义，我们可以有两个图解的解释。先说一个较熟悉的几何解释。

在详述几何解释之前，先介绍一个引理的几何意义。就是对于单位向量$\boldsymbol{c}_0$，其与另一个向量$\boldsymbol{a}$的叉积$\boldsymbol{a}\times\boldsymbol{c}_0$的几何图形如何作图求出。

如图 2-31 所示，先过单位向量$\boldsymbol{c}_0$的起始o点构造一个垂直于它的平面，然后把向量$\boldsymbol{a}$投影到这个平面上得到向量$\boldsymbol{a}'$，接着把向量$\boldsymbol{a}'$在平面上绕o点顺时针旋转$\pi/2$角度。容易知道，旋转后的向量$\boldsymbol{a}''$同时垂直于$\boldsymbol{a}$和$\boldsymbol{c}_0$。下面我们将说明$\boldsymbol{a}''$恰好等于$\boldsymbol{a}\times\boldsymbol{c}_0$。

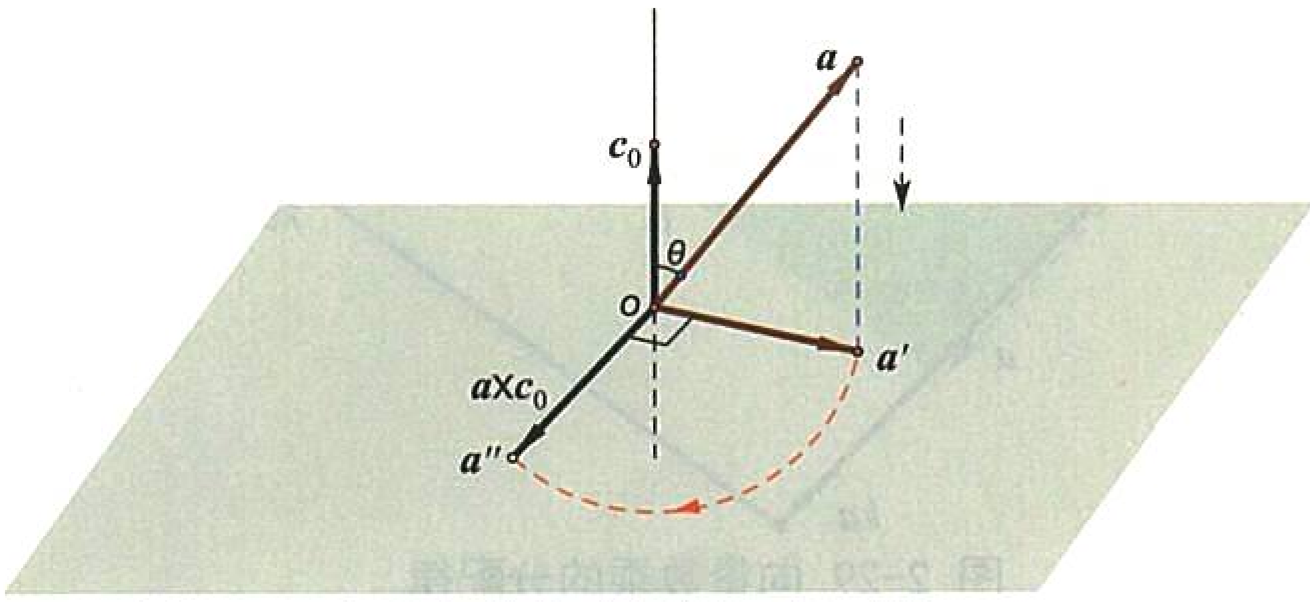


图 2-31 向量与单位向量的叉积图解

因为$\boldsymbol{a}$和$\boldsymbol{c}_0$的叉积向量的长度有

$$|\boldsymbol{a}\times\boldsymbol{c}_0|=|\boldsymbol{a}||\boldsymbol{c}_0|\sin\theta=|\boldsymbol{a}|\sin\theta=|\boldsymbol{a}'|$$

也就是说，任意一个向量$\boldsymbol{a}$和单位向量$\boldsymbol{c}_0$叉积的长度$|\boldsymbol{a}\times\boldsymbol{c}_0|$等于向量$\boldsymbol{a}$向平面（垂直于$\boldsymbol{c}_0$）的投影向量长度$|\boldsymbol{a}'|$。而$|\boldsymbol{a}''|$又是$|\boldsymbol{a}'|$旋转得到的。因此$|\boldsymbol{a}''|=|\boldsymbol{a}'|$。

另外，前面我们说了$\boldsymbol{a}''$同时垂直于$\boldsymbol{a}$和$\boldsymbol{c}_0$，综合得知$\boldsymbol{a}''=\boldsymbol{a}\times\boldsymbol{c}_0$。

引理说完了。下面开始说正题了。

这里有三个向量$\boldsymbol{a}$、$\boldsymbol{b}$和$\boldsymbol{a}+\boldsymbol{b}$，那么这三个向量分别与单位向量$\boldsymbol{c}_0$叉积的向量图形可以模仿上述的过程得到。如图 2-32 所示，我们得到了三个叉积向量$\boldsymbol{a}''=\boldsymbol{a}\times\boldsymbol{c}_0$、$\boldsymbol{b}''=\boldsymbol{b}\times\boldsymbol{c}_0$和$(\boldsymbol{a}+\boldsymbol{b})''=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}_0$。

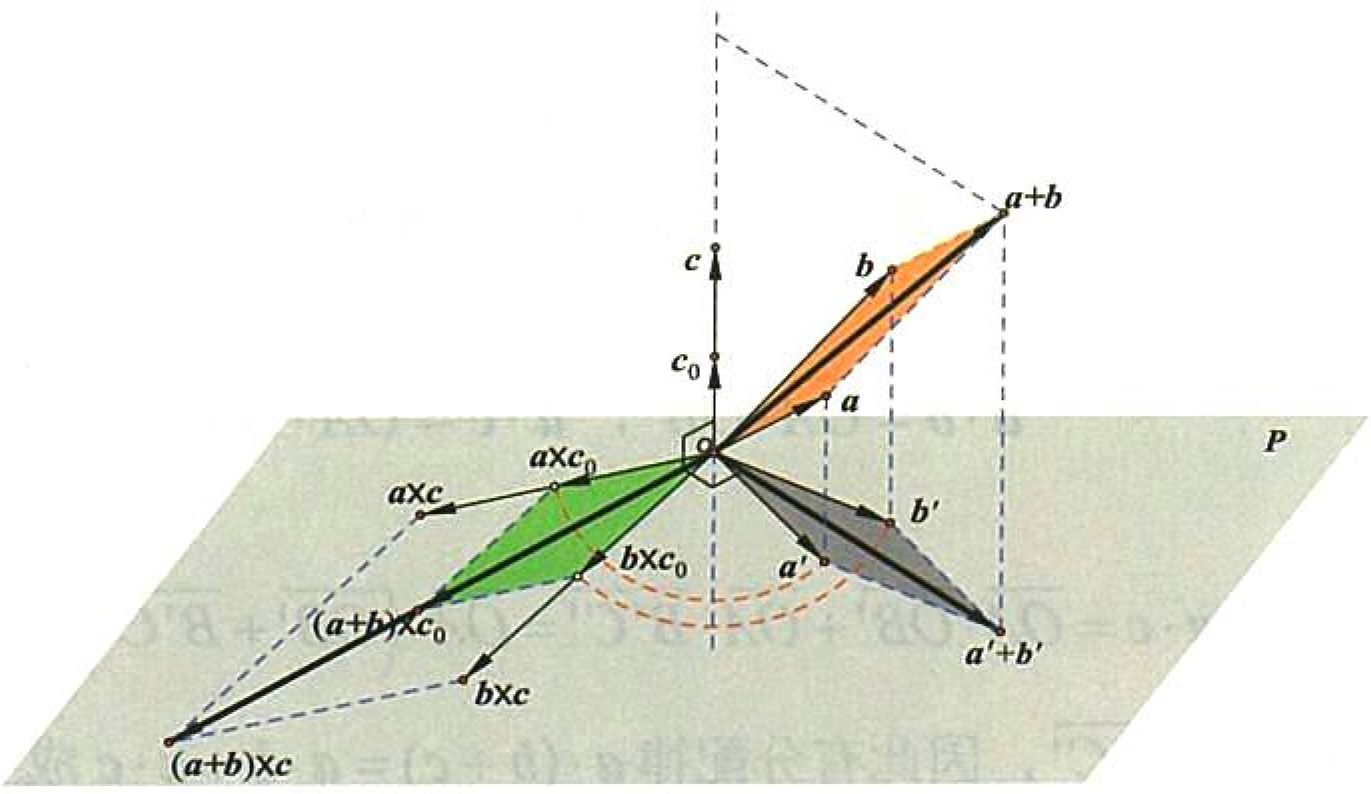


图 2-32 向量叉积分配律的图解

另外，因为三个向量有相加的关系，构成了一个平行四边形，这个平行四边形在投影下不会改变其边的平行或连接的关系，因此投影后的三个向量 $\boldsymbol{a}'$ 、$\boldsymbol{b}'$ 和 $(\boldsymbol{a}+\boldsymbol{b})'$ 在平面上仍然构成一个新的平行四边形，符合平行四边形法则的加法规律，故有 $\boldsymbol{a}'+\boldsymbol{b}'=(\boldsymbol{a}+\boldsymbol{b})'$。

耐心加油呵。接着来，投影平行四边形整体顺时针旋转 90 度后仍然是个平行四边形，因此三个向量 $\boldsymbol{a}''$ 、$\boldsymbol{b}''$ 和 $(\boldsymbol{a}+\boldsymbol{b})''$ 仍然符合平行四边形法则，即有

$$\boldsymbol{a}''+\boldsymbol{b}''=(\boldsymbol{a}+\boldsymbol{b})'' \tag{2-5}$$

把 $\boldsymbol{a}''=\boldsymbol{a}\times\boldsymbol{c}_0$ 、$\boldsymbol{b}''=\boldsymbol{b}\times\boldsymbol{c}_0$ 和 $(\boldsymbol{a}+\boldsymbol{b})''=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}_0$ 代入式（2-5），得

$$\boldsymbol{a}\times\boldsymbol{c}_0+\boldsymbol{b}\times\boldsymbol{c}_0=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}_0 \tag{2-6}$$

好，到此问题基本解决。随后临门一脚是，把单位向量 $\boldsymbol{c}_0$ 伸缩为一般的向量 $\boldsymbol{c}=|\boldsymbol{c}|\boldsymbol{c}_0$ 。即对等式（2-6）两边同乘以一个常数 $|\boldsymbol{c}|$，得 $\boldsymbol{a}\times\boldsymbol{c}_0|\boldsymbol{c}|+\boldsymbol{b}\times\boldsymbol{c}_0|\boldsymbol{c}|=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}_0|\boldsymbol{c}|$，即

$$\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}$$

倒换等式两边的项即得到分配律的结论：

$$(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}=\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}$$

总结一下：上述过程表述为向量的“一投一转，再加一伸”。

2. 向量叉积的分配律的几何解释 2

“一投一转，再加一伸”，叉积分配律整得忒麻烦？再来一个简单点的几何解释。这个简洁的叉积分配图解需要有向面积的概念。下面首先介绍有向面积。

有方向的线段被用来作为向量的图形，那么面积也是有方向的。在微积分对 xy 平面上的曲线与 x 坐标轴围成的面积积分中，x 坐标轴上方的面积积分值为正，轴下方的面积积分值为负（见图 2-33）。这实际上也是面积有方向的表现。

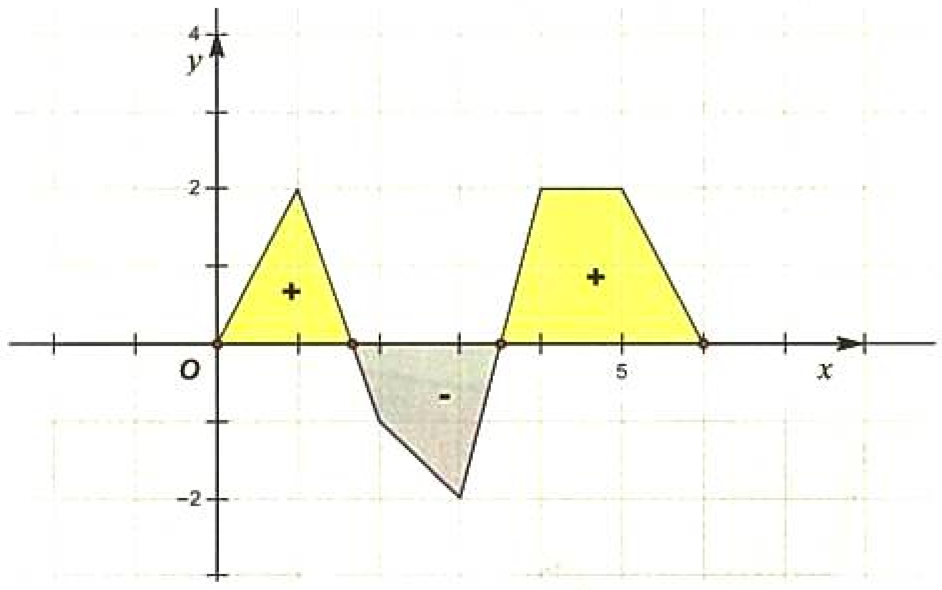


图 2-33 正面积和负面积的方向相反

一个有边界的平面（见图 2-34（a）），它的大小由它的面积决定，它的方向由它在空间的法线的方向确定，因此有向面积也可以用向量的办法来完全刻画：面积向量的方向就是法线的方向，向量的长度正比于它的面积。按照右手规则，如果面积周线的回转方向是 $ABCDE$，那么法线的正向即代表这面积的方向是向上的。根据这种说法，我们可以这样说，向量不但可以表示一个有方向的线段，而且也可以表示一个有方向的面积。反过来讲，**一个有向线段（一定长度的箭头）被用来作为向量的几何图形（这是几乎所有数学书的做法），而一个有方向的面积也可以表示成一个向量。**为了方便，我们可以称前者为线向量，称后者为面向量。

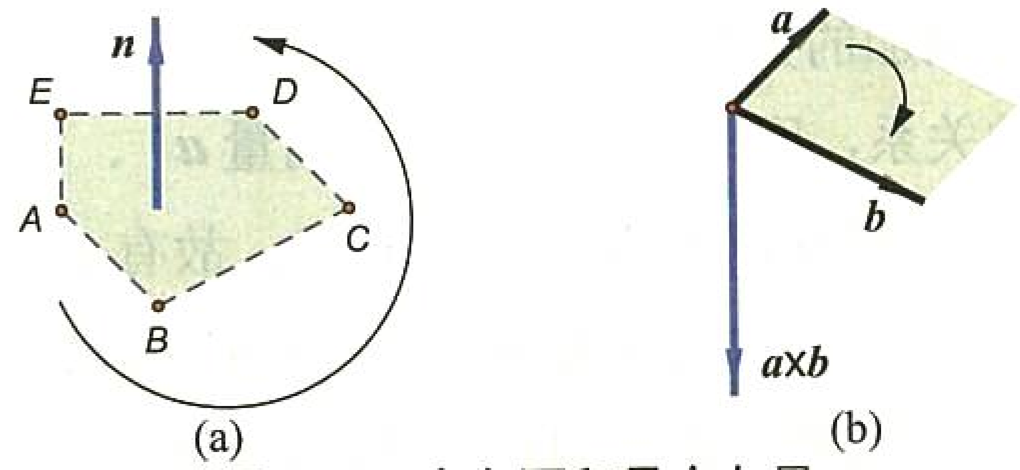


图 2-34 有向面积是个向量

向量$\boldsymbol{a}$和$\boldsymbol{b}$ 的叉积就是一个有向面积的例子（见图 2-34（b）），以向量$\boldsymbol{a}$和$\boldsymbol{b}$ 为边的平行四边形的有向面积是用$\boldsymbol{a}\times\boldsymbol{b}$ 来表示的，因为所有向量都被处理成线向量，所以$\boldsymbol{a}\times\boldsymbol{b}$ 也被刻画成有向线段的图形，这个有向线段垂直于被代表的面，线段长度等于$\boldsymbol{a}$和$\boldsymbol{b}$ 构成的平行四边形的面积。在这里，线向量和面向量混叠在一起。下面的叙述中，我们把线向量扩展成了面向量——具有旋转方向的平行四边形的面。

前面讲过，两个线向量$\boldsymbol{a}$和$\boldsymbol{b}$ 相加得到一个线向量$\boldsymbol{a}\times\boldsymbol{b}$，这个过程满足平行四边形法则或三角形法则。那么，线向量$\boldsymbol{a}$和$\boldsymbol{b}$ 分别和第三个线向量$\boldsymbol{c}$叉积后依次得到了两个面向量$\boldsymbol{a}\times\boldsymbol{c}$和$\boldsymbol{b}\times\boldsymbol{c}$。实际上，面向量的加法运算同样满足平行四边形法则和三角形法则。向量叉积的分配律$\boldsymbol{a}\times\boldsymbol{c}+\boldsymbol{b}\times\boldsymbol{c}=(\boldsymbol{a}+\boldsymbol{b})\times\boldsymbol{c}$ 的图示见图 2-35。

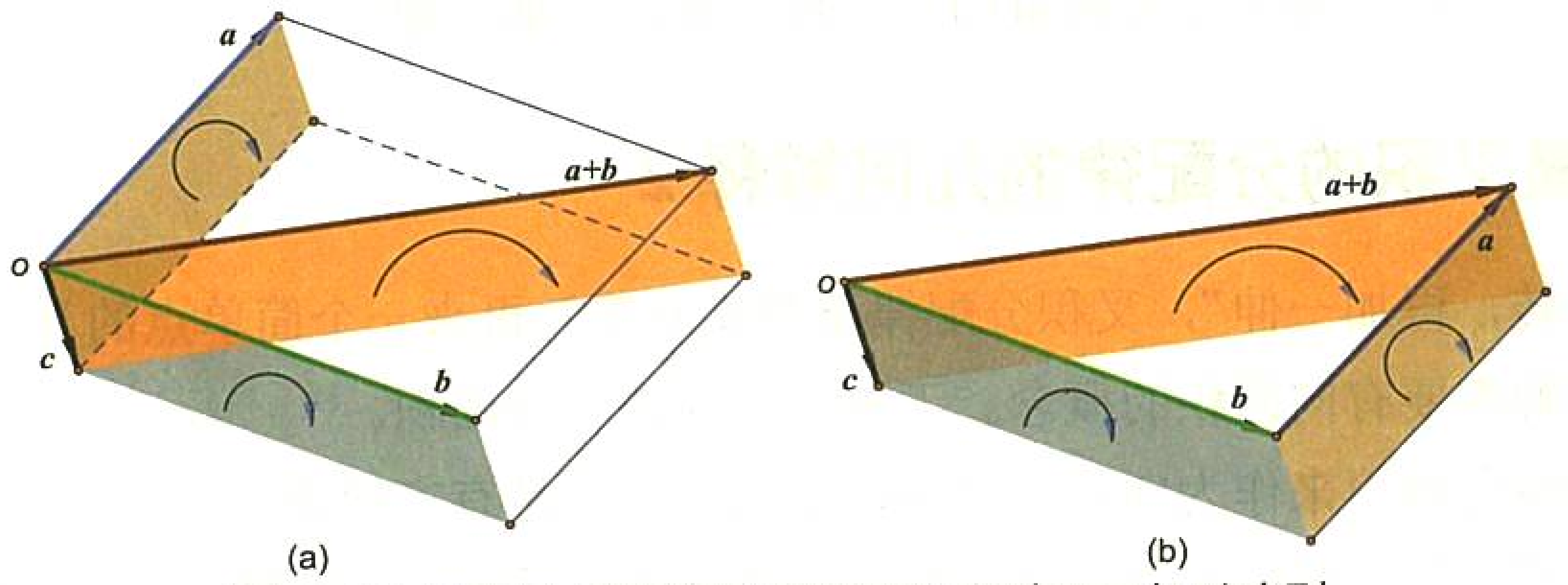


图 2-35 面向量加法的平行四边形法则和三角形法则

面向量的加法法则的证明可以从封闭面的和向量为零着手，这里不再证明了。实际上，咱有更形象的图证来理解这个法则。比如，对于三角形法则（见图 2-35（b）），我们可以想象有向量$\boldsymbol{c}$的长度那么多的三角形上下叠加在一起形成面向量，叠加的方向沿着向量$\boldsymbol{c}$的方向进行，如图 2-36 所示。

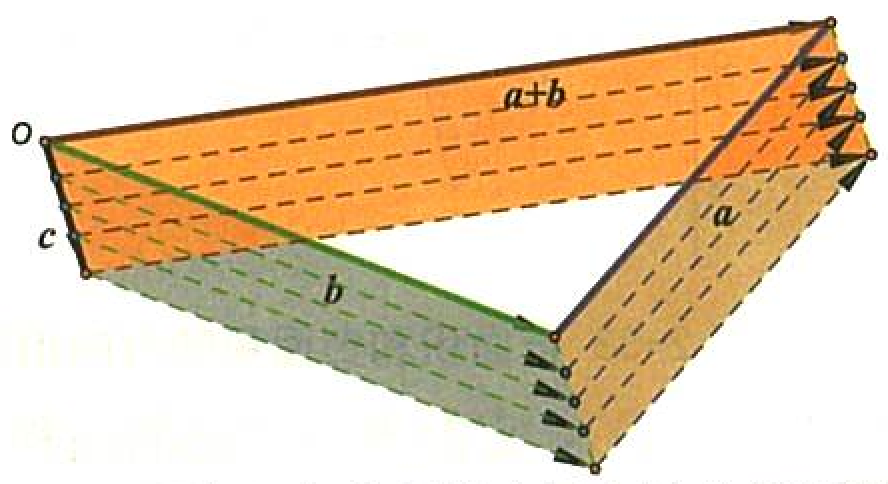


图 2-36 面向量加法是线向量加法的叠加效果

2.5.5 向量混合积的几何解释

三个向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$，如果先作两向量$\boldsymbol{a}$ 和$\boldsymbol{b}$ 的点积$\boldsymbol{a}\cdot\boldsymbol{b}$，这个所得的是数量，再与第三个向量$\boldsymbol{c}$相乘，其结果$(\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}$表述一个新向量，它是向量$\boldsymbol{c}$的伸缩，与向量$\boldsymbol{c}$平行。

如果先作两向量$\boldsymbol{a}$ 和$\boldsymbol{b}$ 的叉积$\boldsymbol{a}\times\boldsymbol{b}$，这个所得的向量与第三个向量$\boldsymbol{c}$再作点积$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$或者

叉积 $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}$，前式结果表示数量，叫做三向量的混合积或三重数积；后式结果表示向量，叫做三重矢积。下面我们仅对向量的混合积的几何意义进行讨论。

三向量的混合积有如下关系：$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=(\boldsymbol{b}\times\boldsymbol{c})\cdot\boldsymbol{a}=(\boldsymbol{c}\times\boldsymbol{a})\cdot\boldsymbol{b}$，且其中 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{c}|\cos\alpha$，$(\boldsymbol{b}\times\boldsymbol{c})\cdot\boldsymbol{a}=|\boldsymbol{b}\times\boldsymbol{c}||\boldsymbol{a}|\cos\beta$，$(\boldsymbol{c}\times\boldsymbol{a})\cdot\boldsymbol{b}=|\boldsymbol{c}\times\boldsymbol{a}||\boldsymbol{b}|\cos\gamma$。混合积是这样的一个数，它的绝对值表示以向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 为棱的平行六面体的体积。如果向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 以顺序组成右手系，那么积的符号是正的；如果组成左手系，那么积的符号是负的。

下面咱推导一下上述结论。

我们知道，向量积 $\boldsymbol{a}\times\boldsymbol{b}$ 是一个向量，它的模在数值上等于以向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 为边的平行四边形 $OADB$ 的平面（见图 2-37），它的方向垂直于这个平行四边形的平面，且当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 以顺序组成右手系时，向量 $\boldsymbol{a}\times\boldsymbol{b}$ 与向量 $\boldsymbol{c}$ 是朝着此平面的同一侧；当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 以顺序组成左手系时，向量 $\boldsymbol{a}\times\boldsymbol{b}$ 与向量 $\boldsymbol{c}$ 是分别朝着此平面的两侧。设向量 $\boldsymbol{a}\times\boldsymbol{b}$ 与向量 $\boldsymbol{c}$ 之间的夹角为 α，则当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 以顺序组成右手系时，夹角 α 为锐角；当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 以顺序组成左手系时，夹角 α 为钝角。

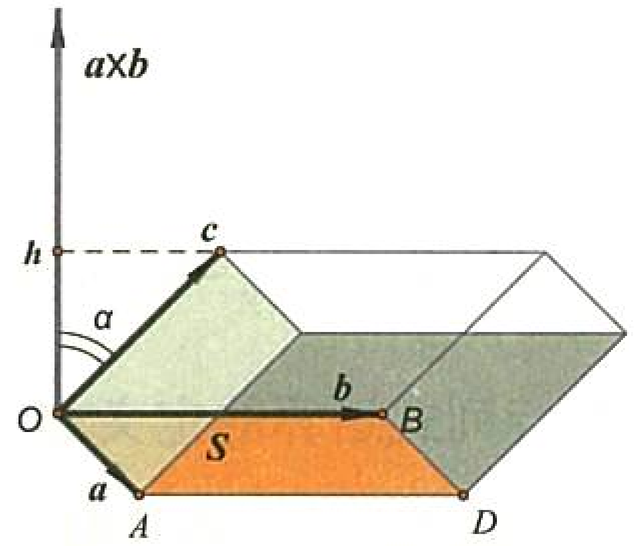


图 2-37 当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 顺序满足右手系则混合积为正

又因为 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{c}|\cos\alpha$，则有当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 组成右手系时，$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 为正值；当向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 组成左手系时，$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 为负值。

因此，以向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 为棱的平行六面体的底（平行四边形 $OADB$）的面积 S 在数值上等于 $|\boldsymbol{a}\times\boldsymbol{b}|$，它的高 h 等于向量 $\boldsymbol{c}$ 在向量 $\boldsymbol{a}\times\boldsymbol{b}$ 上的投影，即 $h=|\boldsymbol{c}|\cos\alpha$，所以平行六面体的体积等于

$$V=Sh=|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{c}|\cos\alpha$$

2.6 向量积和张量之间的关系

从前面的看来，向量的内积定义和外积定义确有意义，但对于我们玩惯实数乘法且还没有得到高等数学训练的人来说，还真有点拿捏不住。为什么不能像两个多项式的乘法一样定义两个向量的乘法呢？

完全可以这样相乘。

2.6.1　二维向量的内积、外积和张量

先看二维空间中的两个向量 $\boldsymbol{a}=(a_x,a_y)$ 和 $\boldsymbol{b}=(b_x,b_y)$，把它们改写成带有 $\boldsymbol{x}$、$\boldsymbol{y}$ 坐标轴的单

位向量的形式，就是 $\boldsymbol{a}=a_x\boldsymbol{i}+a_y\boldsymbol{j}$ 和 $\boldsymbol{b}=b_x\boldsymbol{i}+b_y\boldsymbol{j}$，我们对向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 就像普通的多项式乘法分配律一样展开运算，可得

$$\boldsymbol{ab}=(a_x\boldsymbol{i}+a_y\boldsymbol{j})(b_x\boldsymbol{i}+b_y\boldsymbol{j})=\underline{a_xb_x\boldsymbol{ii}+a_yb_y\boldsymbol{jj}}+\underline{(a_xb_y\boldsymbol{ij}+a_yb_x\boldsymbol{ji})} \tag{2-7}$$

这里有个关键的问题，就是如何定义坐标轴的单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$ 之间的运算？我们发现，不同的规定，就会得到不同的结论：

(1)当定义 $\boldsymbol{i}$、$\boldsymbol{j}$ 之间的运算为内积运算时，即当 $\boldsymbol{ii}=\boldsymbol{jj}=1$，$\boldsymbol{ij}=\boldsymbol{ji}=0$ 时，式（2-7）化简为

$$\boldsymbol{ab}=a_xb_x+a_yb_y=\boldsymbol{a}\cdot\boldsymbol{b}$$

这正是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作内积运算的定义式。

(2)当定义 $\boldsymbol{i}$、$\boldsymbol{j}$ 之间的运算为外积运算时，即当 $\boldsymbol{ii}=\boldsymbol{jj}=0$，$\boldsymbol{ij}=-\boldsymbol{ji}=\boldsymbol{k}$ 时，式（2-7）化简为

$$\boldsymbol{ab}=(a_xb_y-a_yb_x)\boldsymbol{k}=\boldsymbol{a}\times\boldsymbol{b}$$

这正是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作外积运算的定义式，在二维向量空间外又生成了一个第三维向量。

(3)当定义 $\boldsymbol{i}$、$\boldsymbol{j}$ 之间的运算只是作为一个顺序的记号时，即当 $\boldsymbol{ii}=\boldsymbol{jj}=0$，$\boldsymbol{ij}=1$，$\boldsymbol{ji}=-1$ 时，式（2-7）化简为

$$\boldsymbol{ab}=a_xb_y\boldsymbol{ij}+a_yb_x\boldsymbol{ji}=a_xb_y-a_yb_x=\begin{vmatrix}a_x & a_y\\ b_x & b_y\end{vmatrix}=\begin{vmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{vmatrix}$$

这正是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作行向量构成的行列式运算的定义式。

(4)当定义 $\boldsymbol{i}=1$，$\boldsymbol{j}=\sqrt{-1}=\mathrm{i}$，与复数平面进行对应时，式（2-7）化简为

$$\boldsymbol{ab}=\underline{a_xb_x-a_yb_y}+\underline{(a_xb_y-a_yb_x)}\mathrm{i}$$

这正是复数 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作乘法运算的定义式。

总结一下：

式（2-7）的第一部分包含了向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 内积的结果，第二部分包含了向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 外积的结果或者是行列式的结果，即

$$\boldsymbol{ab}=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{b}$$

从这个结论看来，向量的内积运算、外积运算覆盖了二维向量及其复数的所有乘积模式的结果。

实际上，像上述多项式一样的向量乘法叫做向量的直积，向量的直积是向量之间最简单的一种乘法运算，其结果是张量（向量是一阶张量的一种），所以也叫做向量（矢量）的张量积，俗称并矢。

因此，采用较高等一点的说法就是，向量的张量积包含了向量的内积和外积的结果。

2.6.2　三维向量的内积、外积和张量

再看看三维空间中的两个向量 $\boldsymbol{a}=a_x\boldsymbol{i}+a_y\boldsymbol{j}+a_z\boldsymbol{k}$ 和 $\boldsymbol{b}=b_x\boldsymbol{i}+b_y\boldsymbol{j}+b_z\boldsymbol{k}$，同样，我们对向量 $\boldsymbol{a}$

和 $\boldsymbol{b}$ 展开直积运算，可得

$$\begin{aligned}\boldsymbol{ab} &= (a_x\boldsymbol{i}+a_y\boldsymbol{j}+a_z\boldsymbol{k})(b_x\boldsymbol{i}+b_y\boldsymbol{j}+b_z\boldsymbol{k}) \\ &= \underline{a_xb_x\boldsymbol{ii}+a_yb_y\boldsymbol{jj}+a_zb_z\boldsymbol{kk}}+\underline{(a_xb_y\boldsymbol{ij}+a_yb_x\boldsymbol{ji})+(a_xb_z\boldsymbol{ik}+a_zb_x\boldsymbol{ki})+(a_yb_z\boldsymbol{jk}+a_zb_y\boldsymbol{kj})}\end{aligned} \tag{2-8}$$

有点复杂，有没有看出点规律？

与二维向量处理的方式类似，我们定义坐标轴的单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 之间的不同运算，得到了以下不同的结论：

(1)　当定义 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 之间的运算为内积运算时，即 $\boldsymbol{ii}=\boldsymbol{jj}=\boldsymbol{kk}=1$，$\boldsymbol{ij}=\boldsymbol{ji}=\boldsymbol{ik}=\boldsymbol{ki}=\boldsymbol{jk}=\boldsymbol{jk}=0$ 时，式（2-8）化简为

$$\boldsymbol{ab}=a_xb_x+a_yb_y+a_zb_z=\boldsymbol{a}\cdot\boldsymbol{b}$$

这正是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作内积运算的定义式。

(2)　当定义 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 之间的运算为外积运算时，即 $\boldsymbol{ii}=\boldsymbol{jj}=\boldsymbol{kk}=0$，$\boldsymbol{ij}=-\boldsymbol{ji}=\boldsymbol{k}$，$\boldsymbol{ki}=-\boldsymbol{ik}=\boldsymbol{j}$，$\boldsymbol{jk}=-\boldsymbol{kj}=\boldsymbol{i}$ 时，式（2-8）化简为

$$\begin{aligned}\boldsymbol{ab} &= (a_xb_y\boldsymbol{ij}+a_yb_x\boldsymbol{ji})+(a_xb_z\boldsymbol{ik}+a_zb_x\boldsymbol{ki})+(a_yb_z\boldsymbol{jk}+a_zb_y\boldsymbol{kj}) \\ &= (a_xb_y-a_yb_x)\boldsymbol{k}+(a_zb_x-a_xb_z)\boldsymbol{j}+(a_yb_z-a_zb_y)\boldsymbol{i} \\ &= \boldsymbol{a}\times\boldsymbol{b}\end{aligned}$$

这正是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 作外积运算的定义式，在三维向量空间内又生成了第三个同维向量。

总结一下：

式（2-8）第一部分包含了向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 内积的结果，第二部分包含了向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 外积的结果，即：

$$\boldsymbol{ab}=\boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{b}$$

从这个结论看来，向量的内积运算、外积运算覆盖了两个三维向量的所有乘积模式的结果，或者说，向量的张量积包含了向量的内积和外积的结果。

比如，哈密顿算子是一个向量 $\nabla=\frac{\partial}{\partial x}\boldsymbol{i}+\frac{\partial}{\partial y}\boldsymbol{j}+\frac{\partial}{\partial z}\boldsymbol{k}$，当它作用于向量场函数

$$\boldsymbol{v}=v_1\boldsymbol{i}+v_2\boldsymbol{j}+v_3\boldsymbol{k}\text{，（}v_i\text{是 }x\text{、}y\text{、}z\text{ 的函数）}$$

时，分别取内积和外积运算，可得到散度和旋度的结果：

$$\mathrm{div}\boldsymbol{v}=\nabla\cdot\boldsymbol{v},\ \mathrm{rot}\boldsymbol{v}=\nabla\times\boldsymbol{v}$$

内积和外积结合起来就是四元数，显然四元数涵盖了散度和旋度的量，这就是为什么电磁场也可以使用四元数进行分析的原因。

2.7　向量除法的几何意义

向量的乘法有两种：点积和叉积。一般讲除法是乘法的逆运算。那么向量的除法是点积的逆运算呢还是叉积的逆运算？看来比较麻烦，所以大家较少谈论向量的除法。这里我们看看如何麻烦的？先约定一下符号：两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，分别有 $\boldsymbol{a}\cdot\boldsymbol{b}=c$ 和 $\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{c}$。已知 $\boldsymbol{a}$ 和乘积（点乘积或者叉乘积），如何求 $\boldsymbol{b}$。

能不能得到向量 $\boldsymbol{b}$？请看图 2-38。

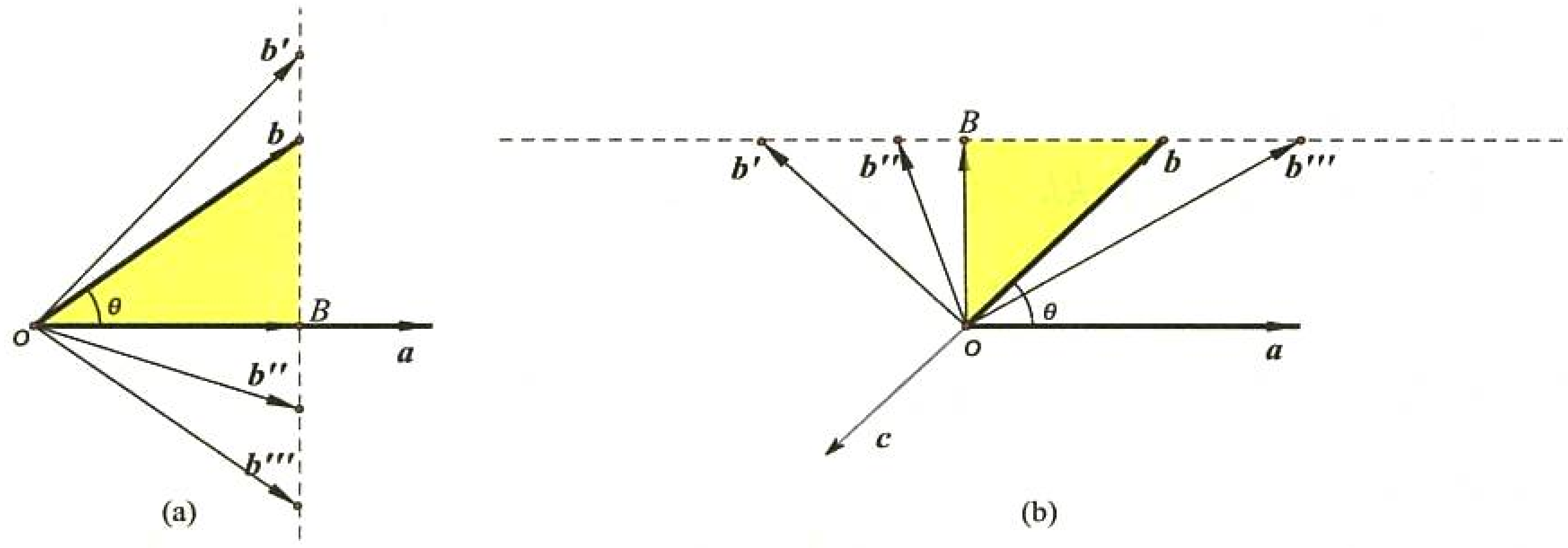


图 2-38 向量内积或者外积不能确定除法

如图 2-38（a）所示，根据点积的公式$\boldsymbol{a}\cdot\boldsymbol{b}=ab\cos\theta$，可得

$$\boldsymbol{a}\cdot\boldsymbol{b}=\boldsymbol{a}\cdot\boldsymbol{b}'=\boldsymbol{a}\cdot\boldsymbol{b}''=\boldsymbol{a}\cdot\boldsymbol{b}'''=\cdots=c$$

无数个向量$\boldsymbol{b},\boldsymbol{b}',\boldsymbol{b}'',\boldsymbol{b}''',\cdots$皆能得到同样的点积值。嘿，数$c$除以向量$\boldsymbol{a}$的结果是$\boldsymbol{b}$是$\boldsymbol{b}'$还是$\boldsymbol{b}'''$？不确定！这个意思是说，点积没有除法。

再看图 2-38（b），根据叉积的公式$\boldsymbol{a}\times\boldsymbol{b}=(ab\sin\theta)\boldsymbol{n}_0$，可得

$$\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{a}\times\boldsymbol{b}'=\boldsymbol{a}\times\boldsymbol{b}''=\boldsymbol{a}\times\boldsymbol{b}'''=\cdots=\boldsymbol{c}$$

无数个向量$\boldsymbol{b},\boldsymbol{b}',\boldsymbol{b}'',\boldsymbol{b}''',\cdots$皆能得到同样的叉积值。向量$\boldsymbol{c}$除以$\boldsymbol{a}$的结果是$\boldsymbol{b}$ 是$\boldsymbol{b}'$还是$\boldsymbol{b}'''$？不确定！这个意思是说，叉积也没有除法。

结果真让人失望！莫急，如果我们把点积和叉积联立解方程组，倒可以解出除法的表达式。解方程：

$$\begin{cases}\boldsymbol{a}\cdot\boldsymbol{b}=c\\ \boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{c}\end{cases}\tag{2-9}$$

对于$\boldsymbol{a}\times\boldsymbol{b}=\boldsymbol{c}$，两边左叉乘$\boldsymbol{a}$，并用二重向量叉积公式$\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c})=\boldsymbol{b}(\boldsymbol{a}\cdot\boldsymbol{c})-\boldsymbol{c}(\boldsymbol{a}\cdot\boldsymbol{b})$得

$$\boldsymbol{a}\times(\boldsymbol{a}\times\boldsymbol{b})=\boldsymbol{a}\cdot(\boldsymbol{a}\cdot\boldsymbol{b})-\boldsymbol{b}\cdot(\boldsymbol{a}\cdot\boldsymbol{a})\tag{2-10}$$

把方程组(2-9)的两个等式带入式(2-10)，即得$\boldsymbol{a}\times\boldsymbol{c}=\boldsymbol{a}\cdot c-\boldsymbol{b}\cdot(\boldsymbol{a}\cdot\boldsymbol{a})$，整理可得

$$\boldsymbol{b}=\frac{c\boldsymbol{a}-\boldsymbol{a}\times\boldsymbol{c}}{\boldsymbol{a}\cdot\boldsymbol{a}}$$

这次$\boldsymbol{b}$ 的结果是唯一确定的值。也就是说，同时知道了点乘积和叉乘积的结果，才可以进行向量的除法运算，这就是向量的除法。

2.8 变向量的几何意义

对于用数组$(a_1,a_2,\cdots,a_n)$表示的n维向量$\boldsymbol{a}$，如果数组中的元素部分或者全部是变量，那么这个变向量在n维坐标系下表示的几何图形是什么呢？

2.8.1 二维变向量的几何图形

在二维平面上，二维向量的两个分量全部为可变的未知量，记为(x_1,x_2)，那么变向量(x_1,x_2)可以表示平面上任意一个向量（无数个）。如果x_1、x_2取遍整个实数范围，则可以涵盖了整个向量平面，因此(x_1,x_2)的几何图形表示为一个平面。如果我们固定(x_1,x_2)一个分量，如(a_1,x_2)，

其中 a_1 表示某一确定的实数，x_2 表示为确定的变元，那么变向量 (a_1,x_2) 表示的是无数个向量，这些向量的末端全部在直线 $x_1=a_1$ 上（见图 2-39（a））。

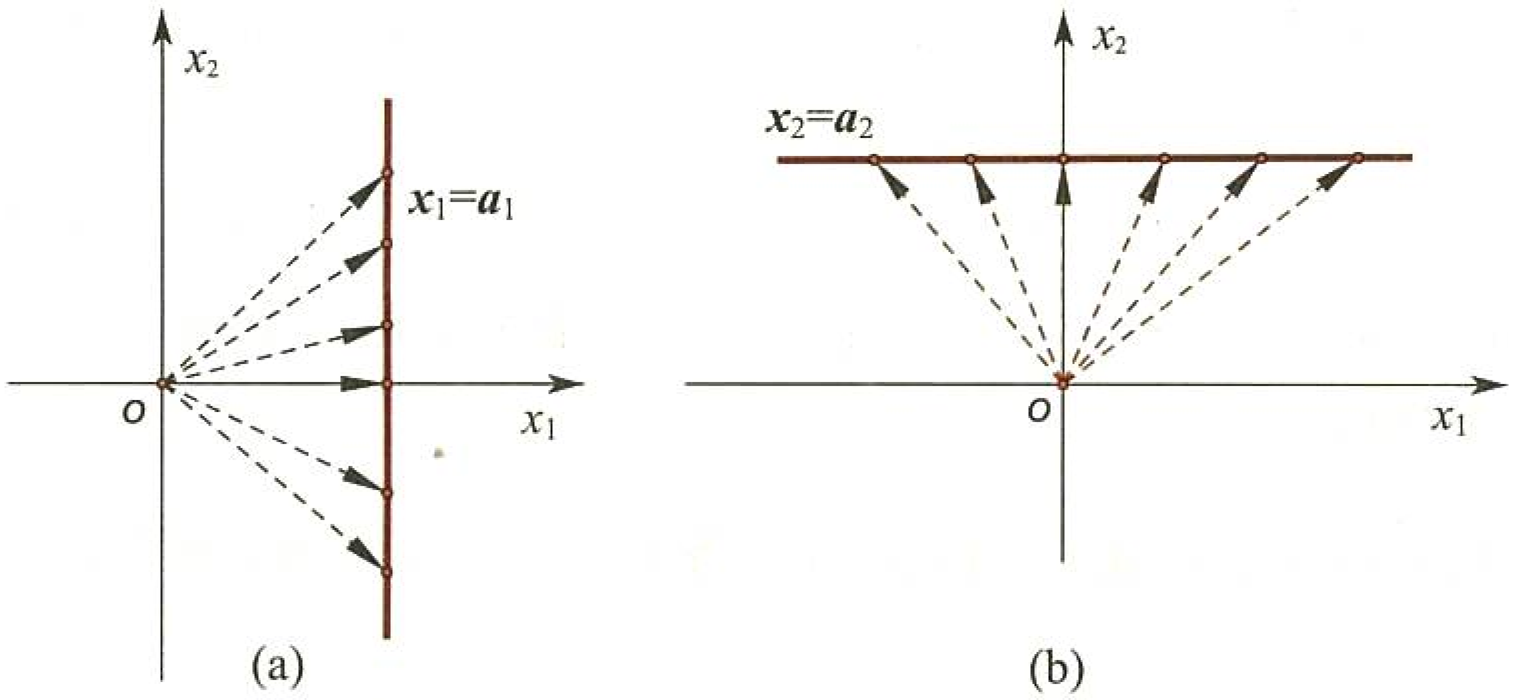


图 2-39 二维变向量的几何图形之一

类似地，变向量 (x_1,a_2) 表示的是直线 $x_2=a_2$，无数个向量的末端全部在直线 $x_2=a_2$ 上（见图 2-39（b））。

进一步，如果向量的分量之间有线性关系的话，比如 $x_2=ax_1+b$，向量可以表示为 (x_1,ax_1+b) 的形式。那么变向量 (x_1,ax_1+b) 就可以表示一个直线，所有向量的末端在直线 $x_2=ax_1+b$ 上（见图 2-40（a））。

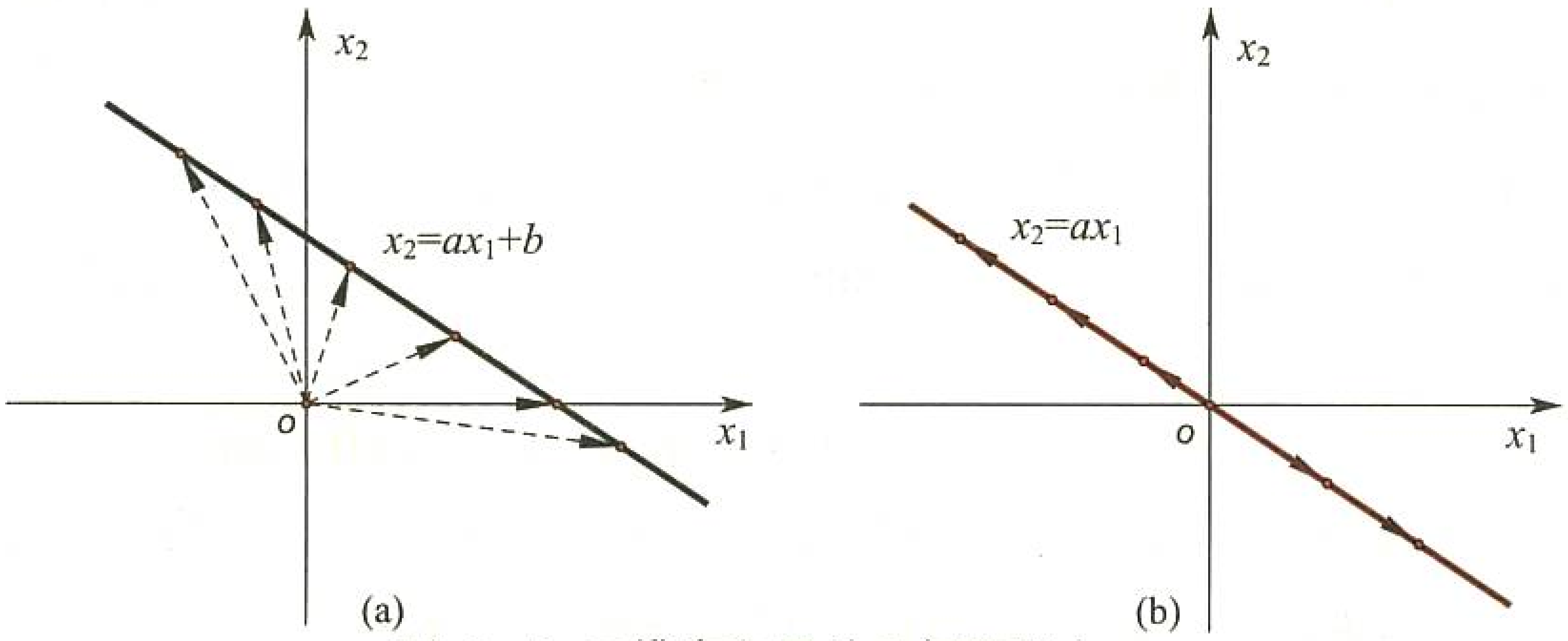


图 2-40 二维变向量的几何图形之二

当然，如果 $b=0$，即 (x_1,ax_1) 就表示为一条过原点的直线，这时所有向量的始端和末端都处于直线上（见图 2-40（b））。

方程 $x_2=ax_1+b$ 和 $x_2=ax_1$ 是平行线，b 是 x_2 轴上的截距。我们用变向量来表示这对平行线的关系就是 $(x_1,ax_1+b)=(x_1,ax_1)+(0,b)$。用向量的观点解释就是，变向量 (x_1,ax_1+b) 是变向量 (x_1,ax_1) 及常向量 $(0,b)$ 的和。从图形上解释就是直线 (x_1,ax_1+b) 是由过原点的直线 (x_1,ax_1) 沿向量 $(0,b)$ 平移 b 得到的。其几何图形如图 2-41所示。

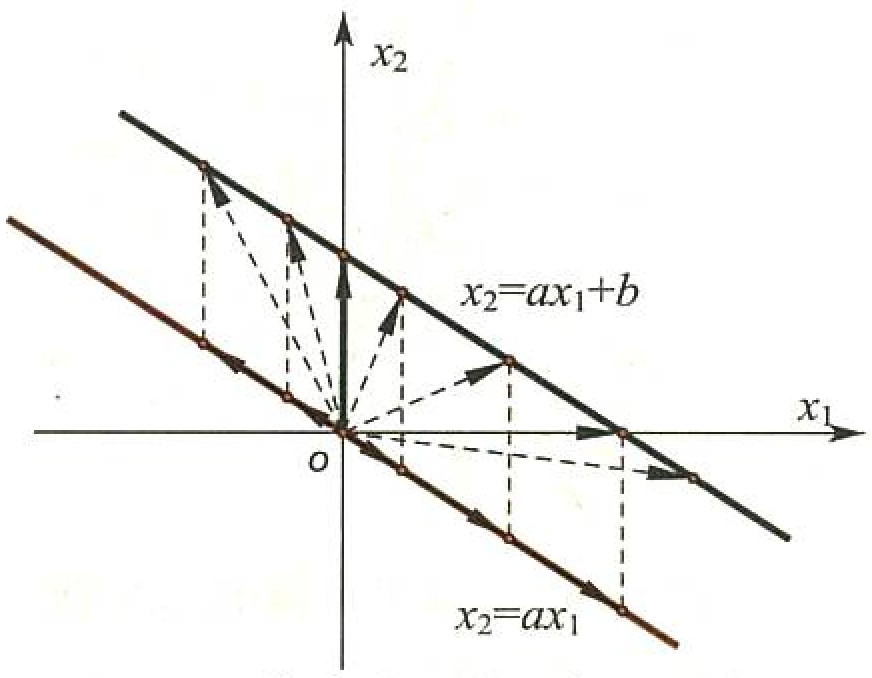


图 2-41 二维变向量的几何图形之三

2.8.2 三维变向量的几何图形

与二维变向量相类似，(x_1,x_2,x_3) 有三个不定变元，表示三维向量空间中的任意一个向量，代表了整个三维空间。

变向量 (a_1,x_2,x_3)、(x_1,a_2,x_3) 和 (x_1,x_2,a_3) 各有两个不定变元，分别表示了平面 $x_1=a_1$，$x_2=a_2$ 和 $x_3=a_3$。每个变向量中任意一个向量的末端都在这个变向量所表示的平面上。这些平面分别垂直于一个坐标轴。

变向量 (a_1,a_2,x_3)、(a_1,x_2,a_3) 和 (x_1,a_2,a_3) 各有一个不定变元，分别表示了三根直线 $\begin{cases}x_1=a_1\\x_2=a_2\end{cases}$，$\begin{cases}x_1=a_1\\x_3=a_3\end{cases}$ 和 $\begin{cases}x_2=a_2\\x_3=a_3\end{cases}$，每个变向量中任意一个向量的末端都在这个变向量所表示的直线上。这些直线分别平行于一个坐标轴。

变向量 (x_1,x_2,ax_1+bx_2) 有两个独立变元 x_1 和 x_2，第三个变元 x_3 与 x_1、x_2 有线性关系：$x_3=ax_1+bx_2$。为了方面看到这个变向量的几何图形，我们对它进行向量分解：

$$(x_1,x_2,ax_1+bx_2)=(x_1,0,ax_1)+(0,x_2,bx_2)=x_1(1,0,a)+x_2(0,1,b)$$

在 x_1 和 x_2 独立地取遍所有不同的实数时，$x_1(1,0,a)+x_2(0,1,b)$ 所形成的无数个向量覆盖了一个平面（见图 2-42）。实际上，在后面的章节中，这个平面是一个向量空间，一个被常向量 $(1,0,a)$ 和 $(0,1,b)$ 所张成的向量平面空间，记为 $\mathrm{Span}\{(1,0,a),(0,1,b)\}$。因此，我们可以有这样的一个等价式：

$$(x_1,x_2,ax_1+bx_2)\cong\mathrm{Span}\{(1,0,a),(0,1,b)\}$$

类似地，变向量 (x_1,x_2,ax_1+bx_2+c) 也有两个独立变元 x_1 和 x_2，第三个变元 x_3 与 x_1、x_2 有线性关系：$x_3=ax_1+bx_2+c$。同样，我们也对它进行向量分解：

$$(x_1,x_2,ax_1+bx_2+c)=(x_1,0,ax_1)+(0,x_2,bx_2)+(0,0,c)=x_1(1,0,a)+x_2(0,1,b)+(0,0,c)$$

同变向量 (x_1,x_2,ax_1+bx_2) 比较起来，变向量 (x_1,x_2,ax_1+bx_2+c) 与它的关系可以表示为

$$(x_1,x_2,ax_1+bx_2+c)=(x_1,x_2,ax_1+bx_2)+(0,0,c)$$

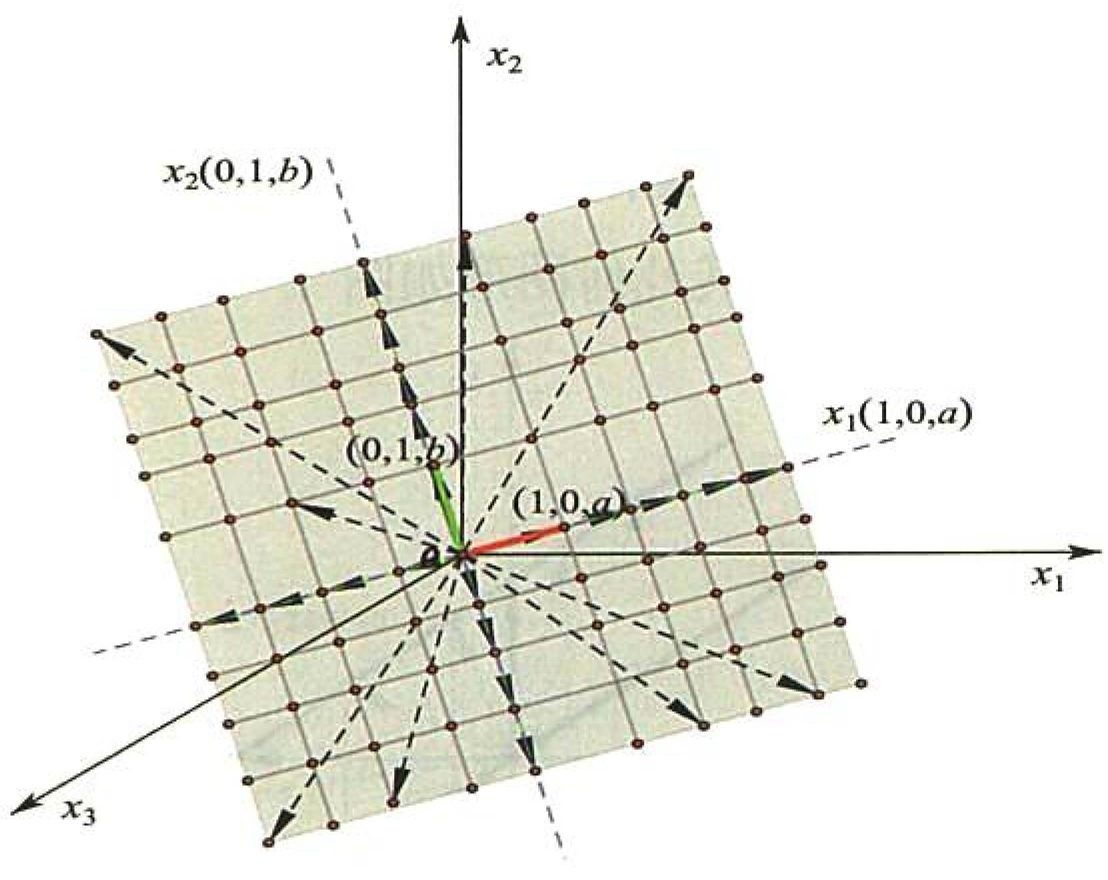


图 2-42 三维变向量的几何图形之一

两个变向量差了一个常向量，因此变向量 (x_1,x_2,ax_1+bx_2+c) 的几何图形是 (x_1,x_2,ax_1+bx_2)

的图形加了一个常向量$(0,0,c)$，得到的新图形仍然是个平面（见图 2-43），只是沿着向量$(0,0,c)$的方向平移了长度为 c 的距离。

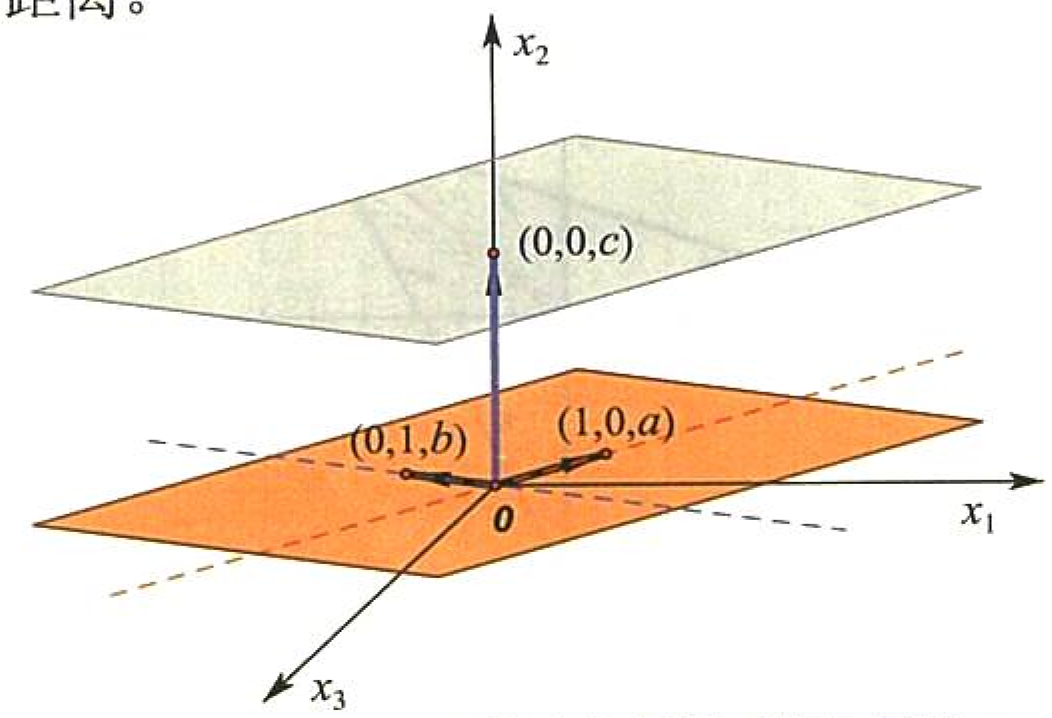


图 2-43 三维变向量的几何图形之二

一个变向量是和一个解析方程或方程组相对应的，因此变向量和方程一样能表示一个几何图形。实际上，变向量也叫向量函数。

2.8.3 变向量的应用

变向量简化了线性方程或方程组，但这种简化与线性方程组的向量方程或者矩阵方程对线性方程的简化又有所不同。下面我们看看如何用变向量的概念去解线性方程组。

对于二阶线性方程组$\begin{cases} x_2 = a_1x_1 + b_1 \\ x_2 = a_2x_1 + b_2 \end{cases}$可以用变向量的交$(x_1, a_1x_1 + b_1) \cap (x_1, a_2x_1 + b_2)$来等价表示。可以想象，两个变向量中相同的向量就是它们的交集。如果两个向量相同，那么向量的夹角就会为零。由行列式的几何意义可知，两个夹角为零的向量构成的平行四边形为零，即行列式为零。可得

$$\begin{vmatrix} x_1, a_1x_1 + b_1 \\ x_1, a_2x_1 + b_2 \end{vmatrix} = 0$$

化简整理得到 $x_1 = 0$ 或者 $x_1 = \dfrac{b_1 - b_2}{a_2 - a_1}$，进而得到向量对$\{(0,b_1),(0,b_2)\}$和$\left\{\left(\dfrac{b_1 - b_2}{a_2 - a_1}, \dfrac{a_2b_1 - a_1b_2}{a_2 - a_1}\right), \left(\dfrac{b_1 - b_2}{a_2 - a_1}, \dfrac{a_2b_1 - a_1b_2}{a_2 - a_1}\right)\right\}$。

实际上，应用行列式为零只是求出所有的相关向量。如上式得到了两对线性相关的向量。第一对向量在 x_2 轴上，方向相反，不是相同的向量，显然不是所求。第二对向量重合，是线性方程组的解$\left(\dfrac{b_1 - b_2}{a_2 - a_1}, \dfrac{a_2b_1 - a_1b_2}{a_2 - a_1}\right)$。

顺便补充一下，从图 2-44中我们看到，还有一对线性相关的向量没有求出来，就是 x_1 轴上的方向相反的一对相关向量。不过如对线性方程组的变向量以 x_2 为自变量时就可以得到这对相关向量。

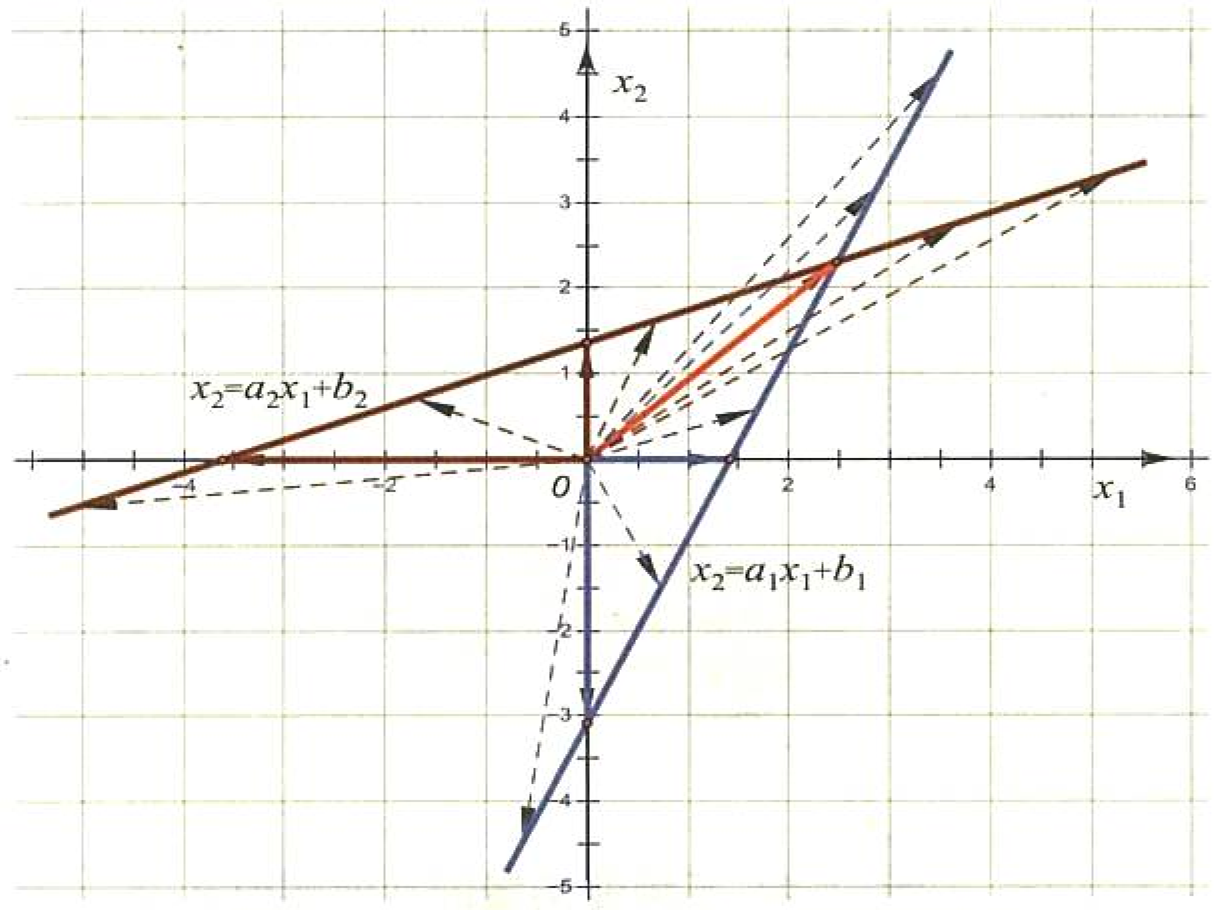


图 2-44 用向量相关求出方程组的解

2.9 复向量的几何意义

我们常常听说，复数可以看做是二维向量，这两者之间的关系到底如何？在讨论复向量之前，我们须先解决这个问题。

2.9.1 向量与复数的关系

据史载，古希腊的亚里士多德已经知道了两个力的合成可以用平行四边形的法则得到。但是，集古希腊数学大成的《几何原本》却没有讨论向量。 以后的一千多年中，经过文艺复兴时期，牛顿创立微积分之后的十七、十八世纪，向量的知识没有什么变化。其中伽利略(1564—1642)清楚地叙述了“平行四边形法则”，仅此而已。这点向量知识稍嫌少了，发展不成一个独立的学科，因而数学家没有把向量当做一回事。进入19世纪，事情开始发生变化。“复数”充当了催化剂。丹麦的魏塞尔(1745—1818)、瑞士的阿工(1768—1822)发现了复数的几何表示，德国的高斯(1777—1855)建立了复平面的概念，从而使平面向量与复数建立起一一对应的关系。这不但为虚数的现实化提供了可能，也为向量的发展开辟了道路。

向量表示为一组有序的实数(a, b)，是一个重大的进步。尽管实平面$\mathbf{R}^2$中的元素和复平面$\mathbf{C}$是一一对应的，并且加法也相同，但逻辑上两者是不同的。在$\mathbf{R}^2$上，我们仅仅能对一个向量作标量乘法（向量的内积和外积都不能看做向量乘法，因为内积得到一个数，而不是向量，它不属于$\mathbf{R}^2$；外积虽然能得到一个向量，但其垂直于而不属于$\mathbf{R}^2$平面），然而在$\mathbf{C}$上，我们可以对任意两个复数z_1、z_2相乘并得到第三个复数z_3。

虽然得到的第三个复数z_3也可以用位于z_1、z_2所在的平面中的向量表示，但这一乘积不是标量也不是用普通的向量分析方法得到的向量。

实际上，我们在2.6.1节“二维向量的内积、外积和张量”中谈到了这个问题：复数z_3如果也看做向量，就是向量z_1、z_2的张量积，是内积和外积的合成。

复数和向量的区别关键点是虚数符号i作为虚轴的单位符号外还参与了乘积运算过程，如

$\mathrm{i}\cdot\mathrm{i}=-1$等。

向量和复数的另外一个区别是向量并不只是平面向量，还有空间向量（三维向量）、四维、五维、六维直到n维向量。但复数只是二元的，如果把复数叫二元数，可惜没有三元数；有四元数，可惜没有五元数、六元数……

虽然复数与向量区别不小，但还是与向量的“线性”没有脱离关系。如果把z_1看做非零复常数，那么$z_3=z_1z_2$就是一个线性变换或线性映射。我们可以用线性函数或线性变换的定义式来验证。当然，可以通过形象的几何运动帮助理解。大家知道（或将会知道），向量的伸缩或旋转是线性变换的几何意义上的典型代表，那么两个复数相乘，就会使其中一个复数旋转和伸缩。如果用指数形式来表示复数z_1、z_2，即令$z_1=r_1e^{\mathrm{i}\theta_1}$，$z_2=r_2e^{\mathrm{i}\theta_2}$，则有

$$z_3=z_1z_2=(r_1r_2)e^{\mathrm{i}(\theta_1+\theta_2)} \tag{2-11}$$

由式(2-11)可知，变换$z_3=z_1z_2$把z_2的半径r_2伸缩了$|z_1|=r_1$倍，并且z_2的辐角围绕原点旋转了$\theta_1=\arg z_1$弧度。

向量没有一般意义上的除法（参见2.7节“向量除法的几何意义”），但复数里是有除法的，两复数相除的结果是一个复数，这个复数的模是前面两复数模的商，辐角是前面两复数辐角的差。

其实，复数与矩阵的关系更加紧密，一个复数可以用一个二阶矩阵来表示，甚至四元数也是如此。复数的矩阵表示将在5.14.9节讨论。

2.9.2　复向量的几何意义

复向量是把复数和向量两个概念揉合在一起，这两个概念有似而不同的几何意义，搀和在一起的几何意义就比较热闹了。

对于复数形式$x+y\mathrm{i}$，其中x、y一般定义为实数。好，我们有两个扩展的观点（因为复数和向量是不同的逻辑）：

(1)　如果把x、y再看做复数，那么可以称之为扩元。令$x=x_1+x_2\mathrm{j}$, $y=y_1+y_2\mathrm{j}$，则有

$$x+y\mathrm{i}=(x_1+x_2\mathrm{j})+(y_1+y_2\mathrm{j})\mathrm{i}=x_1+y_1\mathrm{i}+x_2\mathrm{j}+x_2y_2\mathrm{ji}$$

如果令$\mathrm{ji}=\mathrm{k}$，这正是四元数的形式$x_1+y_1\mathrm{i}+x_2\mathrm{j}+x_2y_2\mathrm{k}$。

继续扩展下去，就有八元数、十六元数等等（先不管它们到底能满足啥运算律哈）。显然这是2的幂2^n为增长速度的扩元。这是一个观点。这个观点逻辑上解释了哈密尔顿为什么不能发明三元数。

(2)　如果把实数x、y看做一维向量，那么它们可以扩展到二维向量。我们可称之为扩维。令$x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$，$y=\begin{pmatrix}y_1\\y_2\end{pmatrix}$，则有

$$x+y\mathrm{i}=\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}y_1\\y_2\end{pmatrix}\mathrm{i}=\begin{pmatrix}x_1+y_1\mathrm{i}\\x_2+y_2\mathrm{i}\end{pmatrix} \tag{2-12}$$

这就是二维复向量。当然可以逐一递增地扩展为三维、四维复向量等。这是一个观点。这个观点实际上告诉了我们复向量的几何意义。

对于复数的几何构成，一个是实数线轴上的值，另一个是扩展的虚数线轴上的值。

对于二维复向量构成形式$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\mathrm{i}$，表示为一个复数；其几何构成，一个是二维实平面$\Pi_1$上的向量值，一个是二维纯虚平面$\Pi_2$上的向量值（见图 2-45（a））。

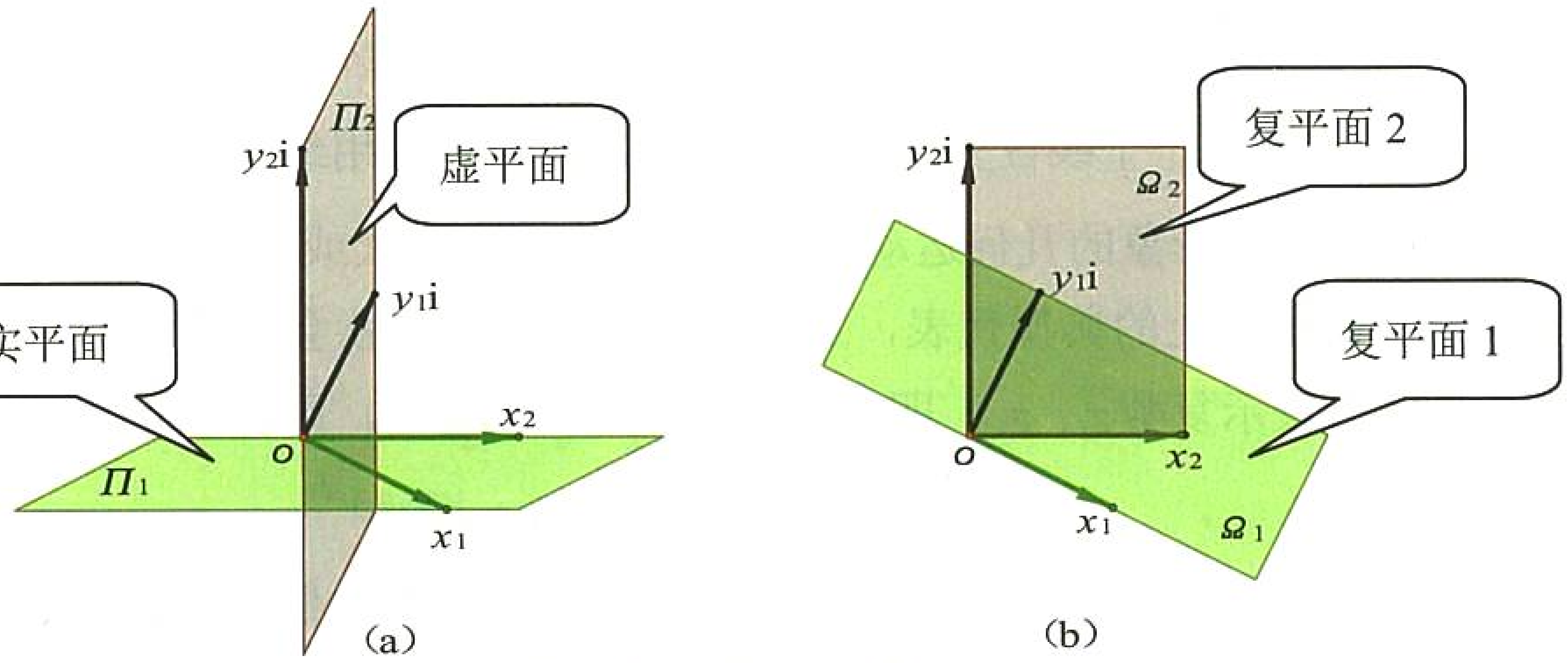


图 2-45 复向量的两种几何意义

而二维复向量的构成形式$\begin{pmatrix} x_1+y_1\mathrm{i} \\ x_2+y_2\mathrm{i} \end{pmatrix}$，表示为一个向量；其几何构成，第一个元素的集合构成一个复平面Ω_1，第二个元素的集合构成另一个复平面Ω_2（见图 2-45（b））。

这两种说法总结一下就是：向量的复数形式和复数的向量形式，两者等价。

全面的说，对于一个复向量可以有三种组合形式（如下式）：

$$\begin{pmatrix} x_1+y_1\mathrm{i} \\ x_2+y_2\mathrm{i} \end{pmatrix}=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\begin{pmatrix} y_1\mathrm{i} \\ y_2\mathrm{i} \end{pmatrix}=\begin{pmatrix} x_1 \\ y_2\mathrm{i} \end{pmatrix}+\begin{pmatrix} y_1\mathrm{i} \\ x_2 \end{pmatrix} \tag{2-13}$$

因此可以看出这些实轴和虚轴有六种平面组合：实–实平面$\{x_1ox_2\}$，虚–虚平面$\{y_1\mathrm{i}\,o\,y_2\mathrm{i}\}$，实–虚平面$\{x_1oy_1\mathrm{i}\}$、$\{x_1oy_2\mathrm{i}\}$、$\{x_2oy_1\mathrm{i}\}$和$\{x_2oy_2\mathrm{i}\}$。图 2-45中实际上给出了其中的四种平面。

关于复向量的进一步讨论，请参考5.8.5节的复特征值及特征向量的内容。

2.10 向量和微积分的关系

我们前面提到过，微积分学的基本思想是“以直代曲”，局部地以切线段代替曲线段；在极限条件下，无数的切线段连接起来完全等同于曲线自身。在这个意义下，微分元$\mathrm{d}x$、$\mathrm{d}y$（包括多元偏微分元∂x、∂y）就是向量。在说明微元是向量之前，我们先回顾一下微分的几何意义。

2.10.1 微分的几何意义

如图 2-46（a）所示，函数$y=f(x)$的图形是一条曲线，对于某一固定的x_0值，曲线上有一定点$M(x_0,y_0)$。当自变量x有微小增量Δx时，得到曲线上另一点$N(x_0+\Delta x,y_0+\Delta y)$，$MP$是曲线在点$M$处的切线，从图中可知 $MQ=\Delta x$，$QN=\Delta y$，$QP=f'(x_0)\Delta x=\mathrm{d}y$。

由此可见，对于可微函数$y=f(x)$而言，Δy是曲线$y=f(x)$上的纵坐标的增量时，$\mathrm{d}y$就是曲线切线上的纵坐标的相应增量：$QP=f'(x_0)\Delta x=\mathrm{d}y$。

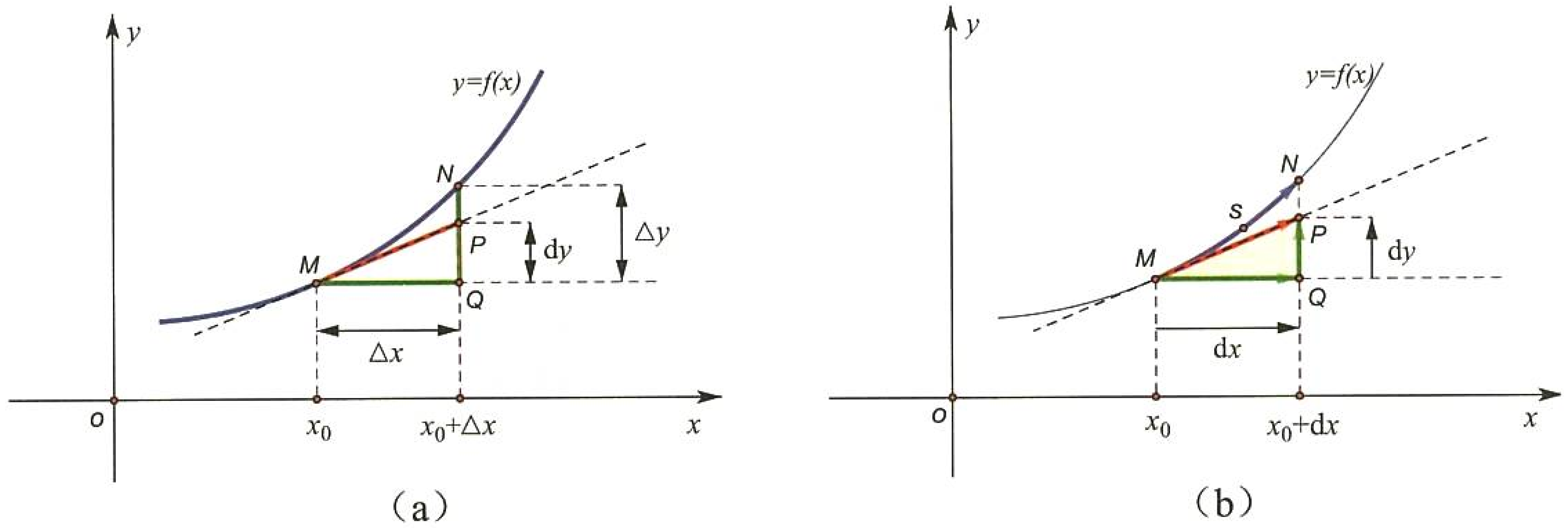


图 2-46 微分的向量分析

由图可见，当Δx趋于无穷小$\mathrm{d}x$时，N点趋向于P点（两点趋于重合），因此有

$$\Delta y = QN = QP = \mathrm{d}y$$

这就是说“曲线”$y = f(x)$的改变量Δy可以用“直线”（即切线）的改变量$\mathrm{d}y$来近似地代替；也就是说，在局部上可以“以直代曲”。

如果我们以向量的观点来看微分（见图 2-46（b）），$\mathrm{d}x$是沿着x轴方向的微分向量，$\mathrm{d}y$是沿着y轴方向的微分向量，那么向量三角形的斜边$\overrightarrow{MP}$就是向量之和$\mathrm{d}x+\mathrm{d}y$。因此，曲线上的一段有向弧$\overset{\frown}{MSN}$就等同于斜边向量$\mathrm{d}x+\mathrm{d}y$。

这就是微分的几何意义。

2.10.2 微元就是向量

根据前面的分析，可微函数曲线上的每一段微小的曲线段都可以分解为微元向量之和$\mathrm{d}x+\mathrm{d}y$。根据这个结论，我们可以得到求导的几何解释。下面举个例子。

对于三角正弦函数$y=\sin x$，周期为2π。我们将x轴分为n个$\mathrm{d}x$区间，由此对函数图形分割出一段段的曲线段出来，这些线段具有方向（见图 2-47（a））。然后把切取的每一个有向线段近似看做切线向量，再将切线向量分解成$\mathrm{d}x$（x轴向向量）和$\mathrm{d}y$（y轴向量），其几何图形如图 2-47（b）。这里$\mathrm{d}x$是等分的固定值。

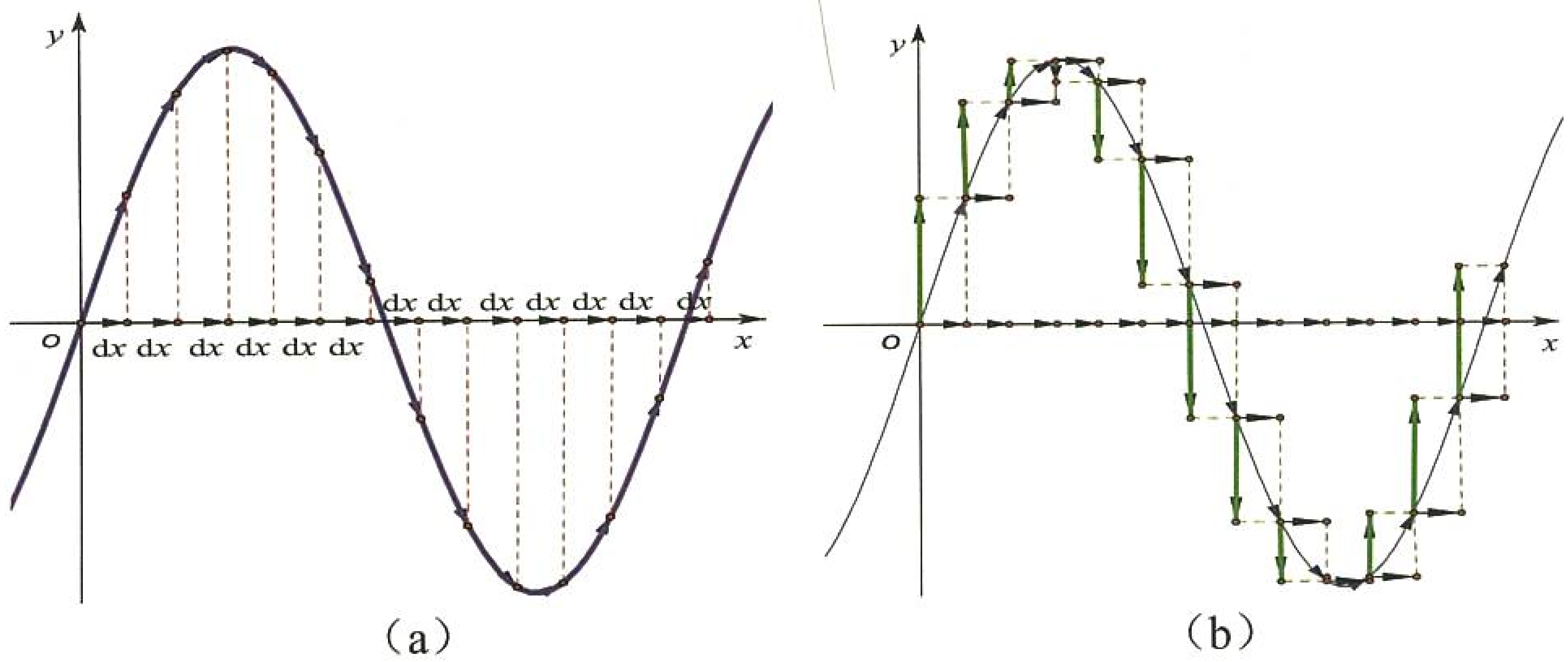


图 2-47 将正弦函数的有向曲线段分解为向量$\mathrm{d}x$和$\mathrm{d}y$

从图 2-47（b）中，只取$\mathrm{d}y$的值（y轴分量），或者把$\mathrm{d}y$向量的线段始端依次排放在x轴上

（见图 2-48），那么向量 dy 的末端构成了一个曲线图形的轮廓，这是与 $y=\cos x$ 的图形同比例的一个图形。

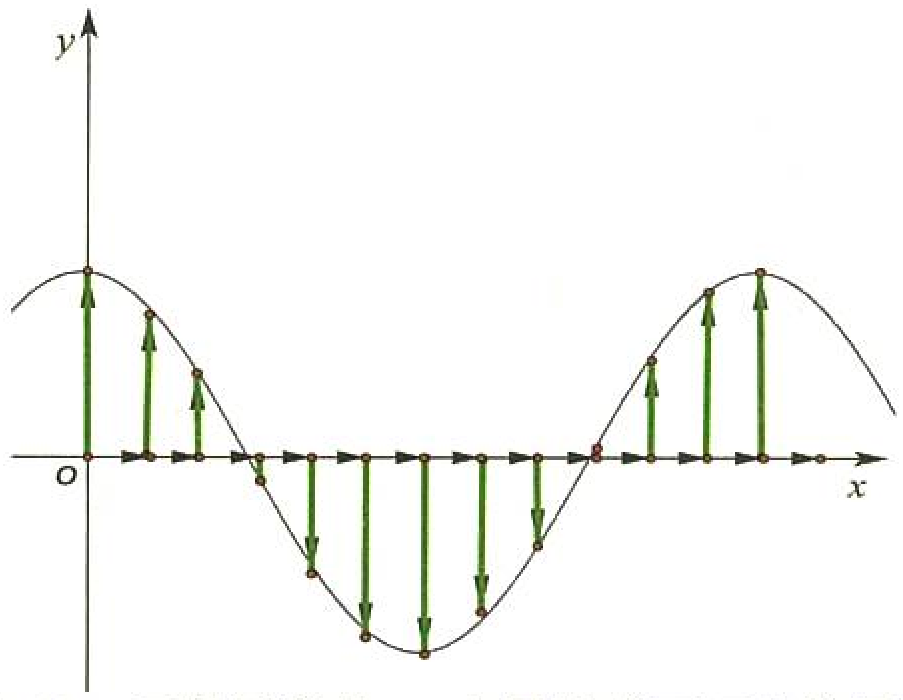


图 2-48　正弦函数的 dy 向量构成了了余弦函数

此例中，我们实际上取了 dx 为 $\pi/6$ 的固定值，因此由正弦函数的导数公式 $dy/dx=\cos x$ 得到

$$dy=\cos x dx=\frac{\pi}{6}\cos x$$

这就是图 2-48所示的几何图形的函数式。

2.11　向量与解析几何的关系

在中学，解析几何课程主要研究平面上一些点、线的几何位置和几何性质，所涉及的点、直线、曲线均共面，它们都落在坐标系所在的平面内。在大学，解析几何课程所讨论的主要是空间图形，所涉及的点、线、面的位置关系复杂。教科书从空间向量与坐标入手，主要以向量为工具，研究空间中的一些几何图形及性质，不论从知识上还是从思维上，都是对中学内容的加深和提高。平面上的几何图形比较直观，点、线之间的关系比较简单，主要以坐标为工具进行研究；而空间上的几何图形大多比较复杂，构成图形的点、线、面及其之间的位置关系较复杂，仅用坐标工具远远不能达到研究问题的最佳境界，只有使用向量这一工具，才能比较好而深刻地探讨空间图形问题。实际上，向量综合了图形（直线段或者说方向）和数字（向量终点用一组数表示）两种几何分析方法。

无论是直线、平面方程的讨论，还是空间曲面、曲线的参数方程以及通过方程去讨论图形，都需要向量的参与。它就像木匠的尺子、石匠的凿子，在整个学科的展开中处处需要。向量工具灵活、好用。它有时是三角形的边，三角形的边的关系经过向量的参与就变成简单的和的关系。它有时又是一个方向，用它可以决定直线的方向、平面的倾斜度，而一族直线、一族平面之间的关系有时通过两个小小的向量之间的关系就可以解决了。在讨论空间曲线、空间曲面的方程时，复杂、多变的轨迹问题又变成了有公共起点的变向量问题，寻找到变向量的变化规律后，问题就迎刃而解了。

几何学家项武义认为“向量几何在本质上是解析几何的返璞归真。”即向量几何揭示了解析（坐标）几何的本质，是解析几何的向前发展。向量几何使用“向量的数量积”，使之成为超越解析几何的有力工具。两条直线的夹角，当然也可以从两直线方程的系数求得。但是，在

向量几何里，它可以用两直线的方向向量的数量积加以表示。本来很费事的夹角问题，通过一次运算就解决了。利用向量，许多几何命题迎刃而解。至于利用向量讨论直线与直线的垂直与平行，空间线面、面面之间的位置关系，比起综合方法需要“个别处理”的技巧，它是一个“一揽子”解决的手段，显得十分轻松。

因此说，向量思想是解析几何学的灵魂，要学好“解析几何”这一课程就必须掌握向量的作用，并要牢牢把握好向量这一有效工具。

第3章 行列式的几何意义

在中国古代，用筹算表示联立一次方程未知量的系数时，就有了行列式的萌芽——排列的方式。日本吸收了这种思想，在 1683 年，日本学者关孝和（Seki Takakusu）对行列式的概念和它的展开已有了清楚的叙述。到 18 世纪，瑞士数学家克莱姆（G.Gramer）和法国数学家拉普拉斯（P.S.Laplace）建立了行列式理论。

行列式先于矩阵的概念出现，是因为行列式属于算术（中国及日本古代数学主要是算术类），是作为解线性方程组的工具，而解方程组可以不需要矩阵的概念。虽然把它放在矩阵后面看做矩阵的一个运算更合乎逻辑，但行列式显然有其独特自立的数学性质，行列式是矩阵大哥这是历史的选择，先讨论其几何意义一点也不突兀。

行列式的几何意义具有深刻含义。它是指行列式的行向量或列向量所构成的平行多面体的有向体积。这个有向体积是许多块更小的有向面积或有向体积的累加。在我们逐步地讨论这个几何意义之前，先来回顾一下行列式的定义。

3.1 行列式的定义

行列式是由一些数据排列成的方阵经过规定的计算方法而得到的一个数。当然，如果行列式中含有未知数，那么行列式就是一个多项式。它本质上代表一个数值，这点请与矩阵区别开来。矩阵只是一个数表（当然这个数表更不简单，参见第第 5 章章），行列式还要对这个数表按照规则进一步计算，最终得到一个实数、复数或者多项式。

行列式分阶，比如二阶行列式、三阶行列式直至 n 阶行列式。下面我们罗列了各阶行列式的定义（这里以拉普拉斯展开定理的形式给出了行列式定义，这样可以使各阶行列式看起来有规律些），以方便后面的论述。

(1) 一阶行列式：

$$|a_1| = a_1$$

一阶行列式就等于元素 a_1 自己。各位看官注意啊，a_1 值可正可负，行列式两条竖线的记号不要当成绝对值的符号。

(2) 二阶行列式：

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 \cdot |b_2| - a_2 \cdot |b_1| = a_1 b_2 - a_2 b_1$$

这个结果是不是很面熟？有点像三维向量叉积的第三个元素。你看，

$$\boldsymbol{a} \times \boldsymbol{b} = (a_1, a_2, a_3) \times (b_1, b_2, b_3) = (a_2 b_3 - a_3 b_2\ ,\ a_3 b_1 - a_1 b_3\ ,\ a_1 b_2 - a_2 b_1)$$

什么有点像啊，本来就是一家人！注意叉积是和有向面积联系在一起的。

(3) 三阶行列式：

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \cdot \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \cdot \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \cdot \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_1 b_3 c_2 - a_2 b_1 c_3 - a_3 b_2 c_1$$

三阶行列式的计算比较常用。上式帮助记忆之经典的展开运算法是对角线法，这个大家都知道，不再多讲。但为使对角线法看起来更有规律更好玩，这里稍稍改变了一下，如图 3-1 所示。

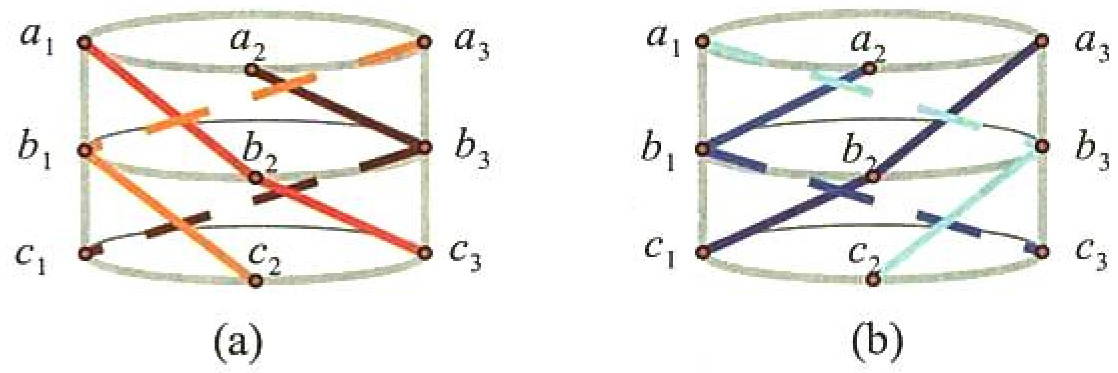


图 3-1　三阶行列式计算的圆柱规律图

把行列式的元素依序排在圆柱体的外圆上，沿**右下**方向的乘积顺序得到正符号（见图 3-1（a）），沿**左下**方向的乘积顺序得到负符号（见图 3-1（b））。如果把圆柱体拓扑变形为圆锥体，它的顶视图如图 3-2 所示，运算顺序很有规律和美感。图中，行向量沿圆的逆时针方向排列元素，列向量沿圆的半径方向排列。乘积项的元素分布在各个半圆弧上。

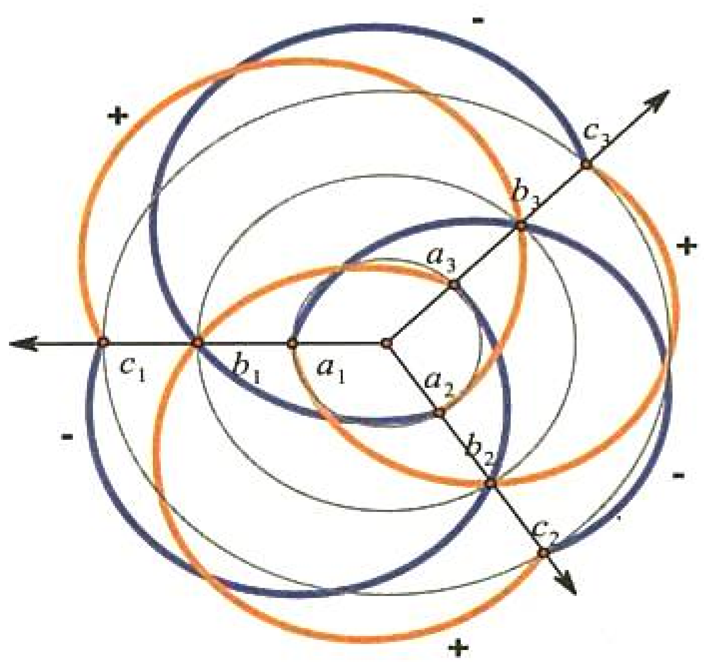


图 3-2 三阶行列式计算的圆锥规律图

(4)　四阶行列式：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{11}\cdot\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12}\cdot\begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{13}\cdot\begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14}\cdot\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix}$$

$$\begin{aligned} = &+a_{11}a_{22}a_{33}a_{44} - a_{11}a_{22}a_{34}a_{43} + a_{11}a_{23}a_{34}a_{42} \\ &-a_{11}a_{23}a_{32}a_{44} + a_{11}a_{24}a_{32}a_{43} - a_{11}a_{24}a_{33}a_{42} \\ &-a_{12}a_{21}a_{33}a_{44} + a_{12}a_{21}a_{34}a_{43} - a_{12}a_{23}a_{34}a_{41} \\ &+a_{12}a_{23}a_{31}a_{44} - a_{12}a_{24}a_{31}a_{43} + a_{12}a_{24}a_{33}a_{41} \\ &+a_{13}a_{21}a_{32}a_{44} - a_{13}a_{21}a_{34}a_{42} + a_{13}a_{22}a_{34}a_{41} \\ &-a_{13}a_{22}a_{31}a_{44} + a_{13}a_{24}a_{31}a_{42} - a_{13}a_{24}a_{32}a_{41} \\ &-a_{14}a_{21}a_{32}a_{43} + a_{14}a_{21}a_{33}a_{42} - a_{14}a_{22}a_{33}a_{41} \\ &+a_{14}a_{22}a_{31}a_{43} - a_{14}a_{23}a_{31}a_{42} + a_{14}a_{23}a_{32}a_{41} \end{aligned}$$

可惜，在复杂的四阶以上的行列式中，没有了以上三阶行列式的特有美感的运算图解。这说明行列式的几何本质远没有如三阶行列式的图解一样简单，这时逆序数的排列概念就显得必要起来。再者，手算四阶行列式很是恐怖，在 MATLAB 横行的年代也没此必要。把四阶行列式的展开算式罗列在此，只是给诸位一个感性认识。

(5) n 阶行列式：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13}\dots a_{1n} \\ a_{21} & a_{22} & a_{23}\dots a_{2n} \\ \vdots & \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n2} & a_{n3}\dots a_{nn} \end{vmatrix} = a_{11}\cdot\begin{vmatrix} a_{22} & a_{23}\dots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{n2} & a_{n3}\dots a_{nn} \end{vmatrix} - a_{12}\cdot\begin{vmatrix} a_{21} & a_{23}\dots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n3}\dots a_{nn} \end{vmatrix} + \dots + (-1)^{n+1}a_{1n}\cdot\begin{vmatrix} a_{21} & a_{22}\dots a_{2n-1} \\ \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n2}\dots a_{nn-1} \end{vmatrix}$$

$$= \sum_{(j_1 j_2 \cdots j_n)} (-1)^t a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

式中：$j_1 j_2 \cdots j_n$ 是 $1,2,\cdots,n$ 的一个排列，$\sum\limits_{(j_1 j_2 \cdots j_n)}$ 表示对 $j_1 j_2 \cdots j_n$ 取遍 $1,2,\cdots,n$ 的一切 $(j_1 j_2 \cdots j_n)$ 排列求和，t 为排列 $j_1 j_2 \cdots j_n$ 的逆序数。

n 阶行列式的展开有 $n!$ 个乘积项，这些乘积项具有统一的表达式，因而具有统一的规律。通过观察以上各阶行列式的定义式，我们至少看到有两个小规律。一个规律是 n 阶行列式可以化为更低一阶的 n-1 阶行列式的和，这将会帮助我们简化行列式的计算。另一个规律是 n 阶行列式实际上是不同行不同列的 n 个元素的乘积项的代数和（和或差，每项符号由置换的逆序数的奇偶性决定）。在前面的向量一章中，我们介绍了向量的张量积，现在看来，行列式实际上是 n 个行向量的张量积中的部分结果。这个部分项的和具有独立的代数及几何意义，因而在线性代数中占有一席之地。

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超平行多面体的**有向面积或有向体积**；另一个解释是矩阵 $\boldsymbol{A}$ 的行列式 $\det\boldsymbol{A}$ 就是线性变换 $\boldsymbol{A}$ 下的图形面积或体积的**伸缩因子**。

一个 2×2 矩阵 $\boldsymbol{A}$ 的行列式，是 $\boldsymbol{A}$ 的行向量（或列向量）决定的平行四边形的有向面积。

用几何观点来看，二阶行列式 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ 是 xoy 平面上以行向量 $\boldsymbol{a}=(a_1,a_2)$，$\boldsymbol{b}=(b_1,b_2)$ 为邻边的平行四边形的有向面积：若这个平行四边形是由向量 $\boldsymbol{a}$ 沿逆时针方向转到 $\boldsymbol{b}$ 而得到的，则面积取正值；若这个平行四边形是由向量 $\boldsymbol{a}$ 沿顺时针方向转到 $\boldsymbol{b}$ 而得到的，则面积取负值。若 $\boldsymbol{a}$、$\boldsymbol{b}$ 与 $\boldsymbol{a}'$、$\boldsymbol{b}'$ 张成的平行四边形的有向面积符号相同，则称它们有相同定向。平面上平行四边形有两种定向。

类似地，三阶行列式的值就是它的三个向量在 $oxyz$ 空间上张成的平行六面体的有向体积。这里空间平行六面体也有两种定向：当 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 构成右手系时，体积取正值；当 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 构成左手系时，体积取负值。这同时启发我们可以把 n 阶行列式定义为一个 n 维平行多面体的有向容积。

关于伸缩因子的几何解释，需要引入矩阵和线性变换的概念。行列式被看做对矩阵的某种运算，并反映了矩阵就是某一线性变换的性质。实际上也是这样。

本章主要从向量及其张成的面积和体积的几何图像角度分别就二阶和三阶行列式给出其几何意义。为了较深入地理解其几何含义，我们逐个解释行列式的主要性质（从矩阵的线性变换的角度讲解行列式的伸缩因子的几何意义的内容放在5.10节“矩阵的行列式的几何意义”里面）。

3.2 二阶行列式的几何意义

3.2.1 二阶行列式的几何意义

首先简述一阶行列式的几何意义。

一阶行列式$|a_1|=a_1$的意思就是a_1的一阶行列式是数a_1或者是向量a_1本身。这个数a_1的本身是一维坐标轴上的有向长度（见图 3-3）。这里我强调的是有向的，长度是有向的，是个向量，这一直是个很重要的概念。

图 3-3 一阶行列式是数轴上线段的有向长度

再接着讨论二阶行列式。

上面说，二阶行列式$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$的几何意义是xoy平面上以行向量$\boldsymbol{a}=(a_1,a_2)$，$\boldsymbol{b}=(b_1,b_2)$为邻边的平行四边形的有向面积（见图 3-4）。为什么？

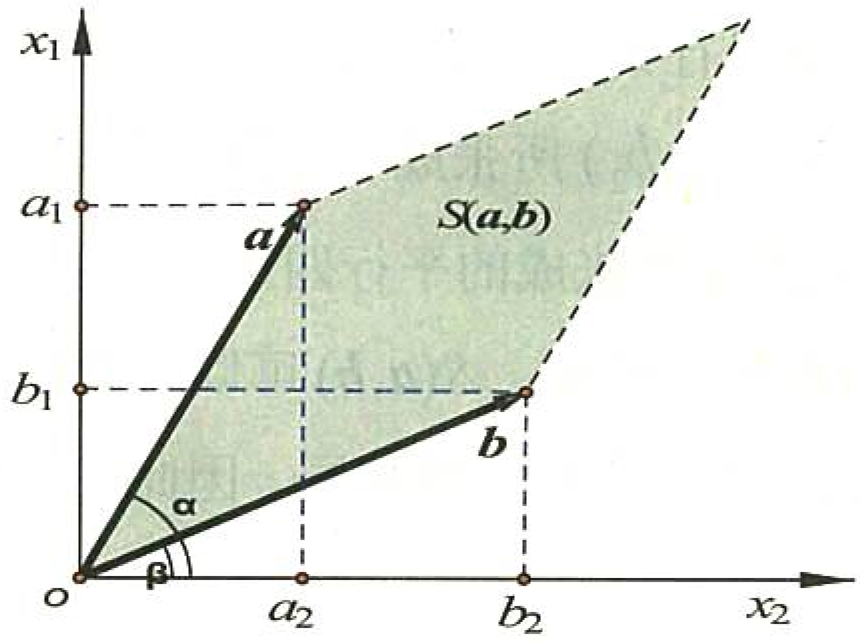


图 3-4 二阶行列式是平面上平行四边形的有向面积

我们来考察这个平行四边形与构成它的两个向量之间的关系。

在二维几何空间$\mathbf{R}^2$中取定一个直角坐标系$\{\mathbf{0};\boldsymbol{e}_1,\boldsymbol{e}_2\}$，设$\boldsymbol{a}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2$，$\boldsymbol{b}=b_1\boldsymbol{e}_1+b_2\boldsymbol{e}_2$，则以$\boldsymbol{a}$、$\boldsymbol{b}$为边的平行四边形的面积为

$$S(\boldsymbol{a},\boldsymbol{b})=ab\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle$$

这里：$a=\sqrt{a_1^2+a_2^2}$，$b=\sqrt{b_1^2+b_2^2}$，$\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle$为向量$\boldsymbol{a}$、$\boldsymbol{b}$之间的夹角正弦，即

$$\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle=\sin(\alpha-\beta)=\sin\alpha\cos\beta-\cos\alpha\sin\beta$$

参照图 3-4中的关系把三角式用坐标值表示出来：

$$\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle=\frac{a_1}{a}\cdot\frac{b_2}{b}-\frac{a_2}{a}\cdot\frac{b_1}{b}=\frac{a_1b_2-a_2b_1}{ab}$$

整理得

$$ab\sin\langle\boldsymbol{a},\boldsymbol{b}\rangle=a_1b_2-a_2b_1$$

又

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}=a_1b_2-a_2b_1$$

因此

$$S(\boldsymbol{a},\boldsymbol{b})=\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

至此可以得到，二阶行列式的几何意义就是由行列式的向量所张成的平行四边形的面积。另外，两个向量的叉积也是这个公式。在向量一章中，向量叉积的一个公式就是

$$\boldsymbol{a}\times\boldsymbol{b}=(ab\sin\theta)\boldsymbol{n}_0$$

这里，$\boldsymbol{n}_0$是右手系下垂直于$\boldsymbol{a}$和$\boldsymbol{b}$ 展成的平面的单位向量。

因此，二阶行列式的另一个意义就是两个行向量或列向量的叉积的数值。如果数值是正值，则与右手系的x_3坐标轴同向；如果数值是负值，则与x_3坐标反向。那么二阶行列式就与两个向量的叉积完全等价了。

3.2.2 二阶行列式性质的几何解释

下面我们仍然从向量的角度来解释或证明二阶行列式的几个主要性质。

性质1 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}=\begin{vmatrix} ka_1 & ka_2 \\ b_1 & b_2 \end{vmatrix}$，$k$为实数。

这个性质是说，一个实数乘以行列式等于一个行向量乘以这个实数的行列式。几何解释就是：两个行向量$\boldsymbol{a}=(a_1,a_2)$，$\boldsymbol{b}=(b_1,b_2)$所张成平行四边形之有向面积的k倍面积等于这样两个向量$k\boldsymbol{a}=(ka_1,ka_2)$，$\boldsymbol{b}=(b_1,b_2)$所张成的平行四边形的面积，也就是$S(k\boldsymbol{a},\boldsymbol{b})=kS(\boldsymbol{a},\boldsymbol{b})$。

通过图 3-5容易得到几何解释。图中，$S(\boldsymbol{a},\boldsymbol{b})$可以看做以$\boldsymbol{a}$为底的平行四边形的面积，$S(k\boldsymbol{a},\boldsymbol{b})$是以$k\boldsymbol{a}$为底的平行四边形的面积，高相同。因此底边向量$\boldsymbol{a}$变化了$k$倍，面积也变化了$k$倍。

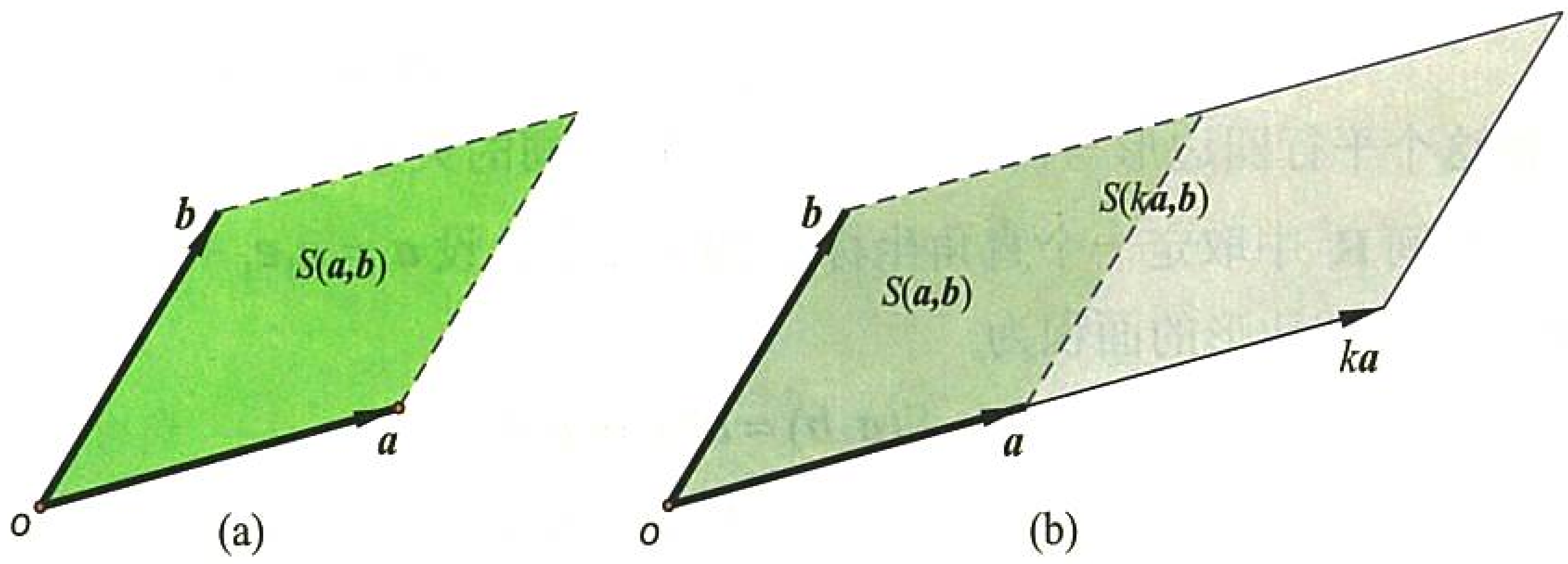


图 3-5 行列式比例的几何解释

性质 2 $\begin{vmatrix} a_1 & a_2 \\ b_1+c_1 & b_2+c_2 \end{vmatrix}=\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}+\begin{vmatrix} a_1 & a_2 \\ c_1 & c_2 \end{vmatrix}$。

这个性质是说，一个行列式可以通过拆分某一行向量而得到两个行列式之和。对于三个向量$\boldsymbol{a}=(a_1,a_2)$，$\boldsymbol{b}=(b_1,b_2)$，$\boldsymbol{c}=(c_1,c_2)$，那么向量$\boldsymbol{a}$和$\boldsymbol{b}$ 张成的平行四边形有向面积与向量$\boldsymbol{a}$和$\boldsymbol{c}$张成的平行四边形有向面积之和等于向量$\boldsymbol{a}$和$\boldsymbol{b}+\boldsymbol{c}$ 张成的平行四边形有向面积，即

$$S(\boldsymbol{a},\boldsymbol{b}+\boldsymbol{c})=S(\boldsymbol{a},\boldsymbol{b})+S(\boldsymbol{a},\boldsymbol{c})$$

图 3-6 和图 3-7 对此进行了图解说明。

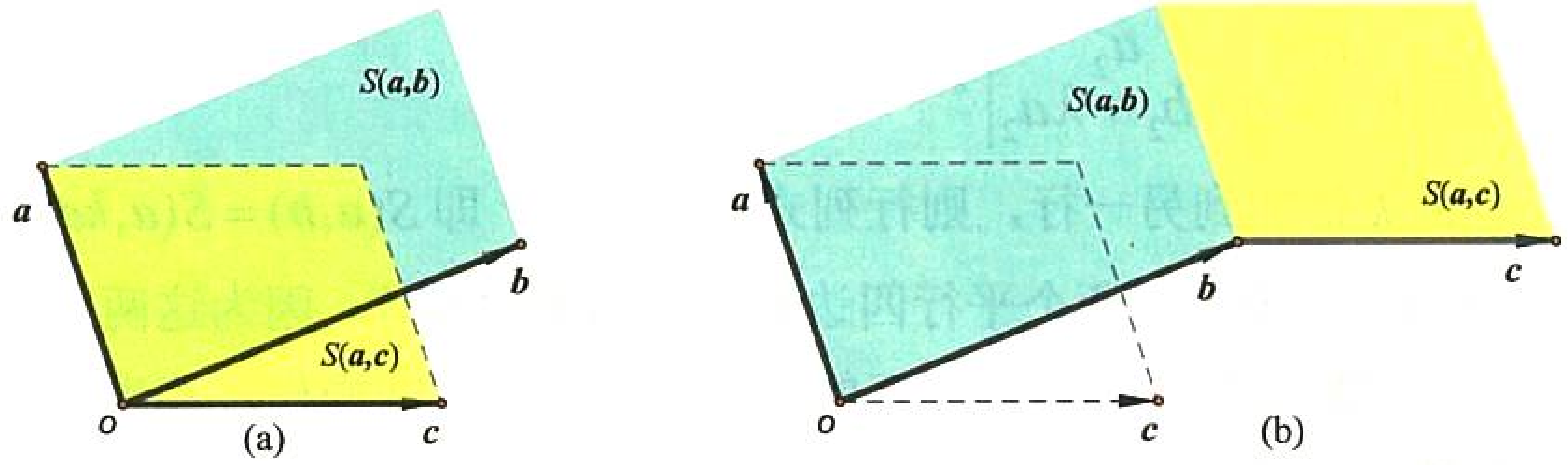


图 3-6 行列式加法的几何解释图一

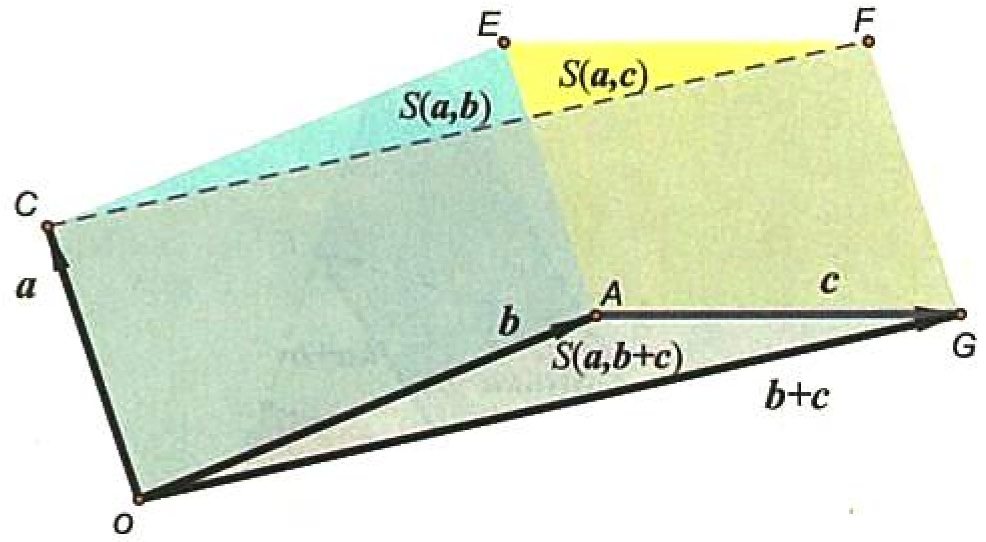


图 3-7 行列式加法的几何解释图二

：我们在第 2.5.4 节中应用有向面积的概念解释了叉积的分配律。因为二阶行列式等同于叉积运算，因此具有类同的几何意义，它们都是一个意义：有向面积满足三角形法则。

性质 3 $\begin{vmatrix} a_1 & a_2 \\ ka_1 & ka_2 \end{vmatrix} = 0$。

此行列式是说两行对应元素成比例，则行列式为零。对于两个向量 $\boldsymbol{a}=(a_1,a_2)$，$\boldsymbol{b}=(ka_1,ka_2)=k(a_1,a_2)$，显然成比例，比例系数为 k。如果把成比列的两个向量的始端都移动到原点，则两向量会在同一条直线上（见图 3-8），显然围成的四边形的面积为零，即 $S(\boldsymbol{a},\boldsymbol{b})=0$，因此行列式为零。

当然如果两个向量相同，同样道理，行列式的值也为零。

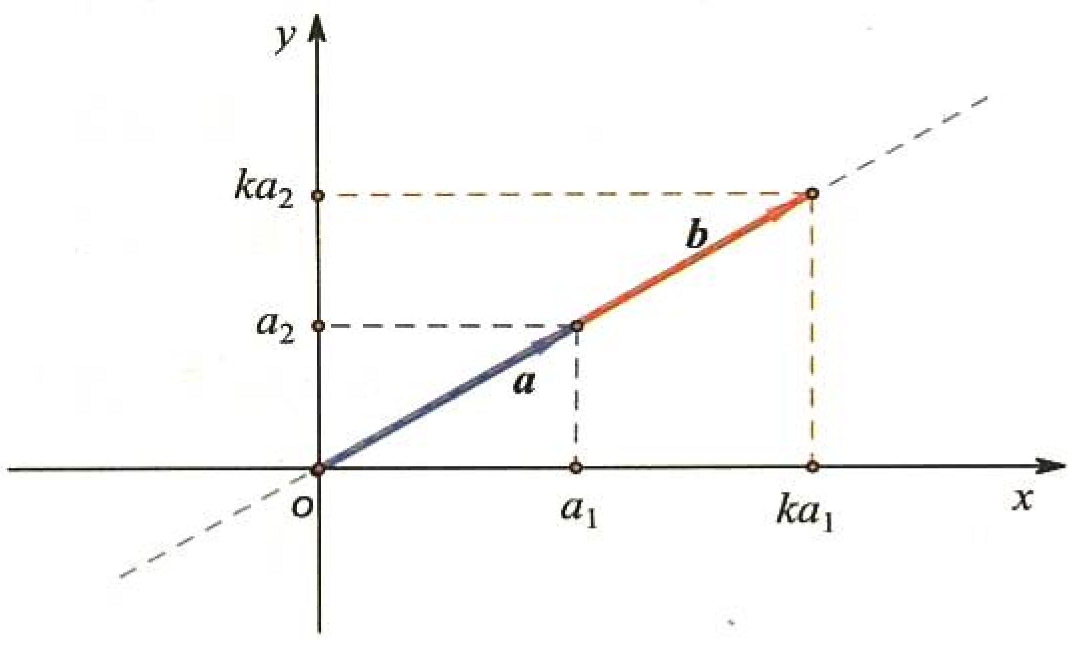


图 3-8 行向量成比例则行列式为零

性质 4 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = -\begin{vmatrix} b_1 & b_2 \\ a_1 & a_2 \end{vmatrix}$。

交换行列式的两行则行列式换号。这个性质由行列式的叉积特性得到。交换行列式的两行，就是改变了向量 $\boldsymbol{a}$ 和向量 $\boldsymbol{b}$ 的叉积顺序，根据 $\boldsymbol{a}\times\boldsymbol{b}=-\boldsymbol{b}\times\boldsymbol{a}$，因此行列式换号。由此我们得到一个印象：就是一个给定的行列式，它的行向量顺序也给定了，不能随意改变其顺序。

性质 5 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1+\lambda a_1 & b_2+\lambda a_2 \end{vmatrix}$。

把行列式的一行的 k 倍加到另一行，则行列式的值不变，即 $S(\boldsymbol{a},\boldsymbol{b})=S(\boldsymbol{a},k\boldsymbol{a}+\boldsymbol{b})$。

由图 3-9 知，在同一平面上，两个平行四边形阴影的面积相等，因为这两个平行四边形同底（都等于 $\boldsymbol{a}$）同高。显然，把行列式的一行的 k 倍加到另一行的操作，相当于把原平行四边形在保持同底同高的情况下发生了切变。

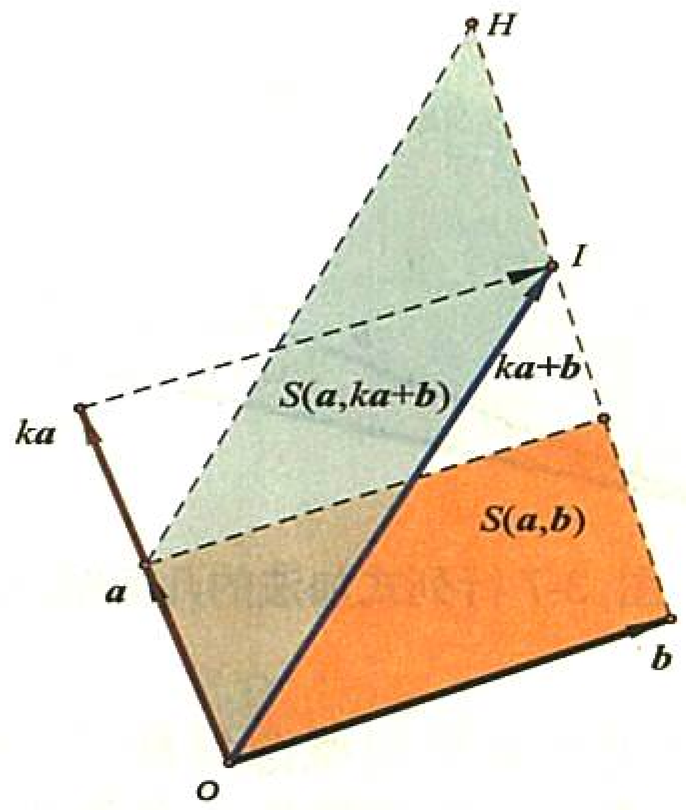


图 3-9　k 倍行向量加到另一行不改变行列式的值

性质 6 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$。

行列式转置，其值不变。

如果要讲清楚转置行列式的几何意义，必须再一次使用行列式叉积的定义。另外，我们要回顾向量一章中两个向量的叉积的解析定义及意义。从前面的分析知道，两个向量的叉积等于这两个向量的各个分量（各坐标轴方向的分量）分别进行叉积的求和。

各个分量互相垂直，因而进行叉积运算张成的四边形是方形的面积。

如图 3-10（a）所示，对于二阶矩阵的行列式$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$，$a_1\boldsymbol{i}$ 和 $b_2\boldsymbol{j}$ 叉乘得到的四方形 oa_1Bb_2 的有向面积是 a_1b_2，$a_2\boldsymbol{i}$ 和 $b_1\boldsymbol{j}$ 叉乘得到的 oa_2Ab_1 的有向面积是-a_2b_1（注意是负值）。所以矩阵的行列式$\begin{vmatrix} a_1, a_2 \\ b_1, b_2 \end{vmatrix}$=“$oa_1Bb_2$ 的有向面积”+“oa_2Ab_1 的有向面积”$=a_1b_2-a_2b_1$。从几何图形上（见图 3-10(a)）看，行列式等于大四方形的面积减去小四方形的面积（因为小四方形是负向面积）。

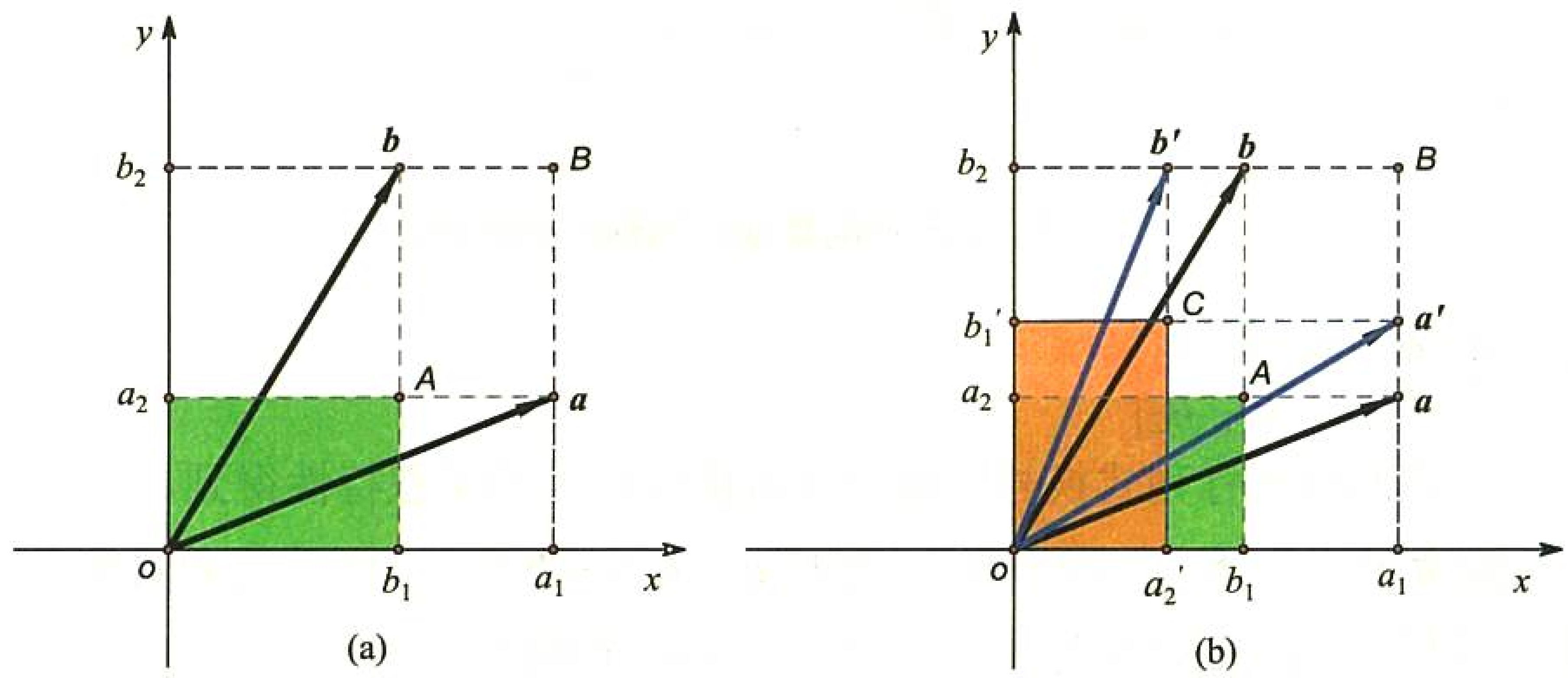


图 3-10 转置行列式的几何解释

行列式$\begin{vmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{vmatrix}=\begin{vmatrix}a_1 & a_2\\b_1 & b_2\end{vmatrix}$转置后得到$|\boldsymbol{a}'\ \boldsymbol{b}'|=\begin{vmatrix}a_1 & b_1\\a_2 & b_2\end{vmatrix}$。重新对这个新的行列式进行分量叉积运算。

我们会发现（见图 3-10（b））大四方形没有变化，而小四方形进行了基于 xy 角平分线的镜像变化。有向面积的绝对值和方向都没有变化。因而

$$\begin{vmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{vmatrix}=|\boldsymbol{a}'\ \boldsymbol{b}'|=a_1b_2-a_2b_1$$

性质得证。

下面是上述转置行列式的具体数据，用于帮助读者得到具体感知。如图 3-11 所示，数据如下：

$$\begin{vmatrix}a_1 & a_2\\b_1 & b_2\end{vmatrix}=\begin{vmatrix}5 & 2\\3 & 7\end{vmatrix}=5\times7-2\times3=29$$

大长方形的面积是 35，小长方形的面积是 6，差为行列式的值。行列式转置后，

$$\begin{vmatrix}a_1 & b_1\\a_2 & b_2\end{vmatrix}=\begin{vmatrix}5 & 3\\2 & 7\end{vmatrix}=5\times7-3\times2=29$$

大长方形的面积仍然是 35，小长方形经过基于 y=x 直线的镜像或者对折，面积仍然是 6，差为行列式的值不变。

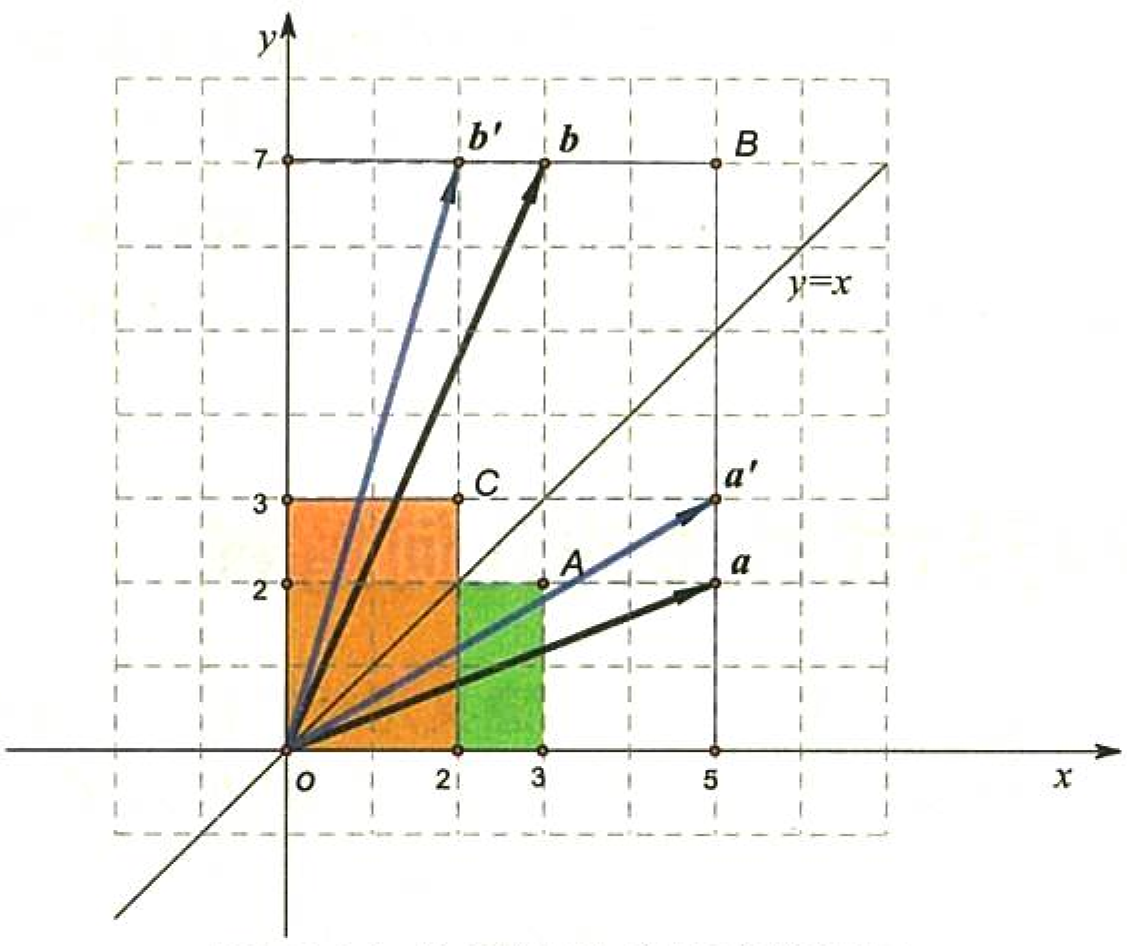


图 3-11　转置行列式的计算例子

矩阵的行列式等于其转置矩阵的行列式这条性质同时揭示了一个认识就是：按矩阵行向量构成的平行四边形的有向面积等于列向量构成的平行四边形的有向面积；换句话讲，对于矩阵的行列式几何意义，处理成行向量的图形与处理成列向量的图形是等效的。

总之，前面行列式的性质在变换时可以归结到下面的三个变换性质：

（1） 用一个数 k 乘以向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 中之一的 $\boldsymbol{a}$，则平行四边形的面积就相应地增大了 k 倍；

（2） 把向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 中的一个乘以数 k 之后加到另一个上，则平行四边形的面积不变；

（3） 以单位向量(1，0)，(0，1)构成的平行四边形(即单位正方形)的面积为 1。

这三条性质，便是我们用公理化的方法定义行列式的几何背景。

3.3 三阶行列式的几何意义

3.3.1 三阶行列式的几何意义

一个 3×3 阶的行列式是其行向量或列向量所张成的平行六面体的有向体积。这个结论可以从两个向量所张成的平行四边形推知。如图 3-12 所示，由两个向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 张成的平行四边形为 $0aPb$，面积 S 为 $\boldsymbol{a}$、$\boldsymbol{b}$ 构成的行列式。那么沿着第三个向量 $\boldsymbol{c}$ 方向生长出无数个平行于原四边形的新的平行四边形来，直至到向量 $\boldsymbol{c}$ 的末端为止。显然，所有的这些平行四边形构成一个以向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 为棱的平行六面体，这些四边形的面积叠加起来正是平行六面体的体积。

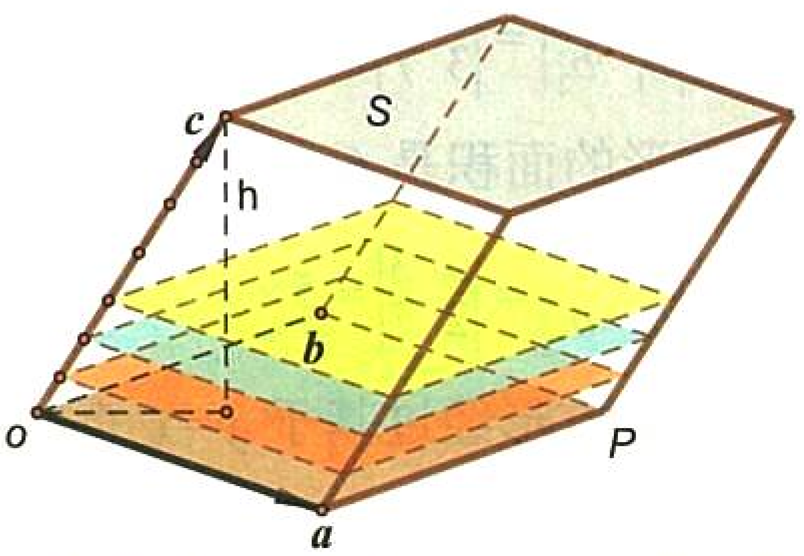


图 3-12 三阶行列式是平行六面体的有向体积

上面的这个平面叠加的过程其实就是向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 混合积的过程：

- 向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 张成的平行四边形为 $oaPb$ 就是向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 叉积的结果 $\boldsymbol{a}\times\boldsymbol{b}$；
- 平行四边形 $oaPb$ 沿着向量 $\boldsymbol{c}$ 张成的平行六面体体积就是向量 $\boldsymbol{c}$ 与 $\boldsymbol{a}$、$\boldsymbol{b}$ 混合积 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 的结果。

3.3.2 三阶行列式性质的几何解释

下面对 3×3 阵的行列式的基本性质的几何意义进行解释。为了书写及描述方便，我们以 det（列向量）的方式表述三阶行列式，列向量用黑体的 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$、$\boldsymbol{d}$ 来表示。则

$$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

性质 1 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}+\boldsymbol{d})=\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})+\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{d})$。

一个行列式可以通过拆分某一个列向量得到两个行列式的和。

看看这个性质的数学表达式，把 det 看做算子，有点像分配律的公式什么的，几何解释如图 3-13 所示。

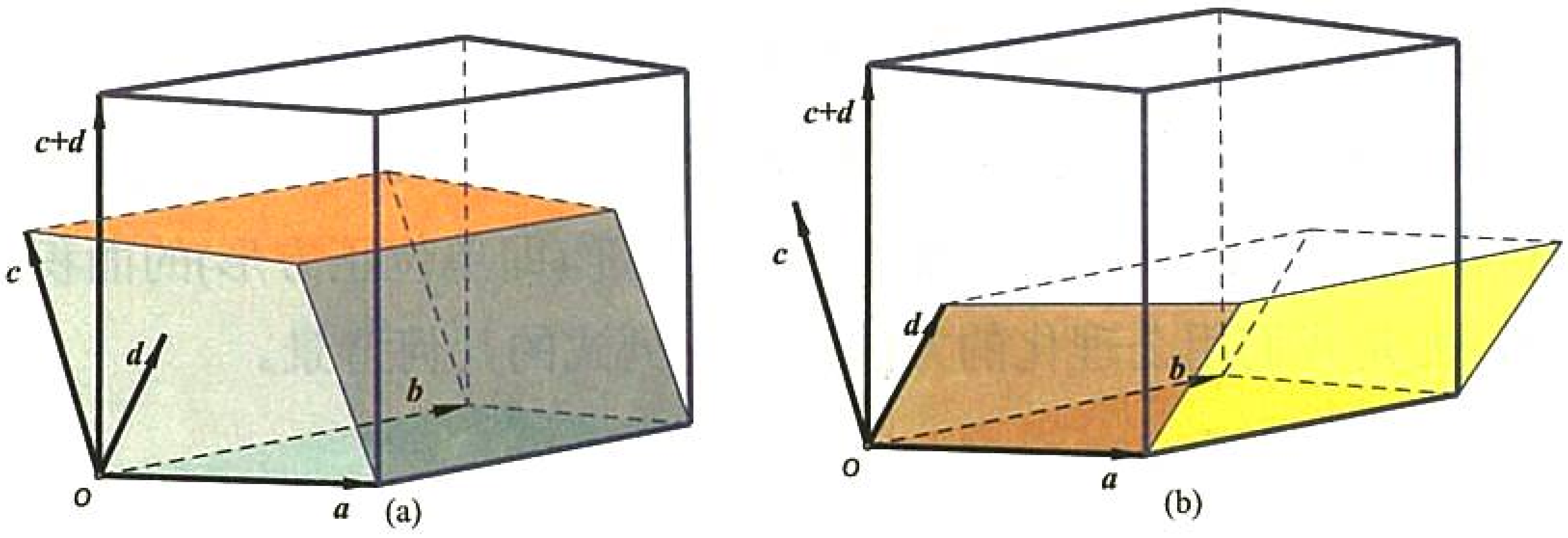


图 3-13 行列式加法的几何解释图一

这里，把行列式 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}+\boldsymbol{d})$ 的第三列拆分，变成两个行列式 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ 和 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{d})$。

三阶行列式可以看做平行六面体的有向体积。图 3-13（a）阴影体积表示由向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$张成的平行六面体，代表行列式$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$；图 3-13（b）阴影体积表示由向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{d}$张成的平行六面体，代表行列式$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{d})$。这两个平行六面体共有一个底面积$\det(\boldsymbol{a},\boldsymbol{b})$。这两个平行六面体的和就是由向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}+\boldsymbol{d}$张成的粗实线平行六面体$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}+\boldsymbol{d})$，图形参见图 3-14。

形象地说，我们把右倾的一摞书——六面体$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{d})$抱起来，摞在左倾的一摞书——六面体$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$上面，再把两摞书向右推整齐，刚好得到六面体$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}+\boldsymbol{d})$。显然，它们的棱向量$\boldsymbol{c}$、$\boldsymbol{d}$、$\boldsymbol{c}+\boldsymbol{d}$满足向量和的三角形法则。

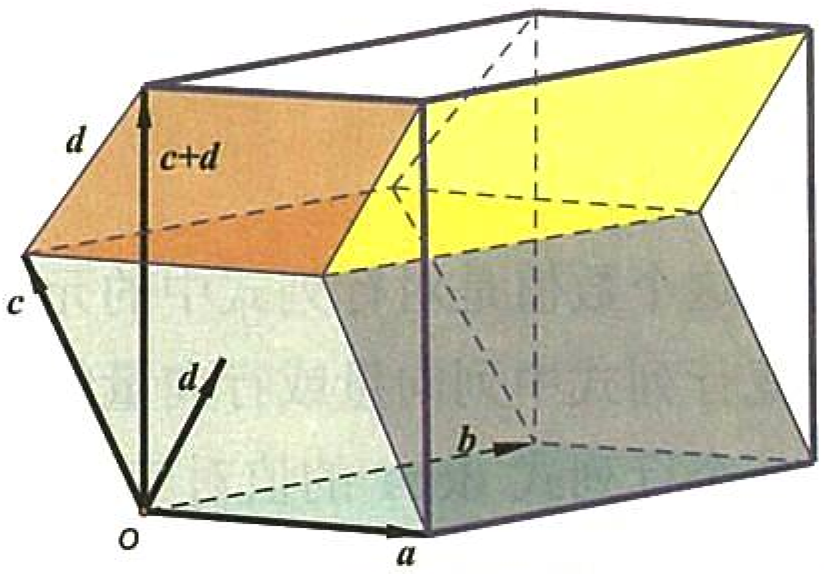


图 3-14 行列式加法的几何解释图二

和前面的有向面积满足三角形法则类似，这里的有向体积同样满足三角形法则。为了更清楚地表达有向体积的三角形法则，把图 3-14 变形，得到如图 3-15 所示的立方三角形图形。

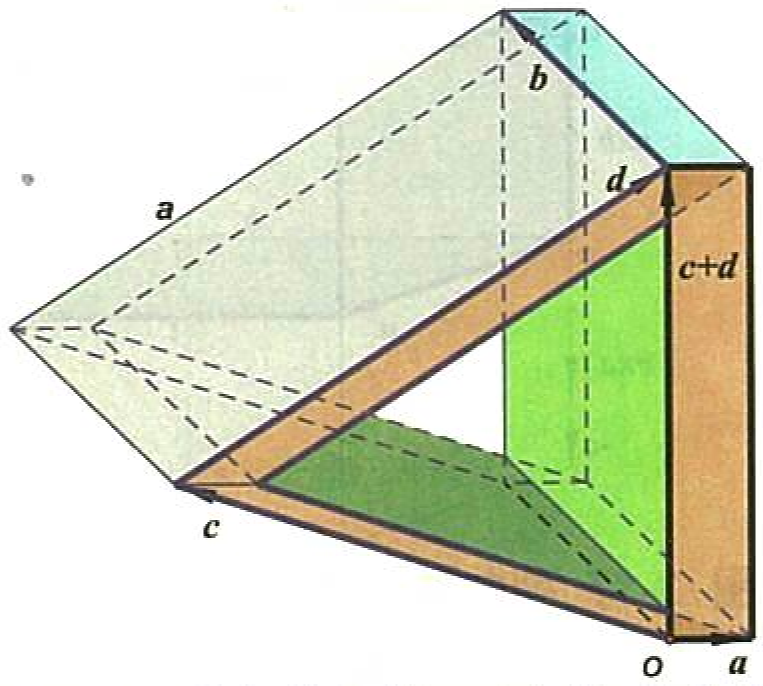


图 3-15 有向体积的三角形加法法则

性质 2　$\det(\boldsymbol{a},\boldsymbol{a},\boldsymbol{c})=0$。

行列式的有两行或者两列元素相同，它对应的空间平行六面体的两条邻边重合，相当于三维空间中六面体被压成了高度为零的二维平面，显然， 这个平面的三维体积$\det(\boldsymbol{a},\boldsymbol{a},\boldsymbol{c})$为零。如图 3-16 所示，我们把平行六面体$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$中的棱向量$\boldsymbol{b}$ 以坐标原点为轴沿一弧线下压（六面保持对应面平行）切变为$\det(\boldsymbol{a},\boldsymbol{b}',\boldsymbol{c})$，显然六面体的高度变小了，行列式值变小了。继续下压，直到$\boldsymbol{b}'$与$\boldsymbol{a}$ 重合，即变成了$\det(\boldsymbol{a},\boldsymbol{a},\boldsymbol{c})$，平行六面体变成了一个平行四边形平面。

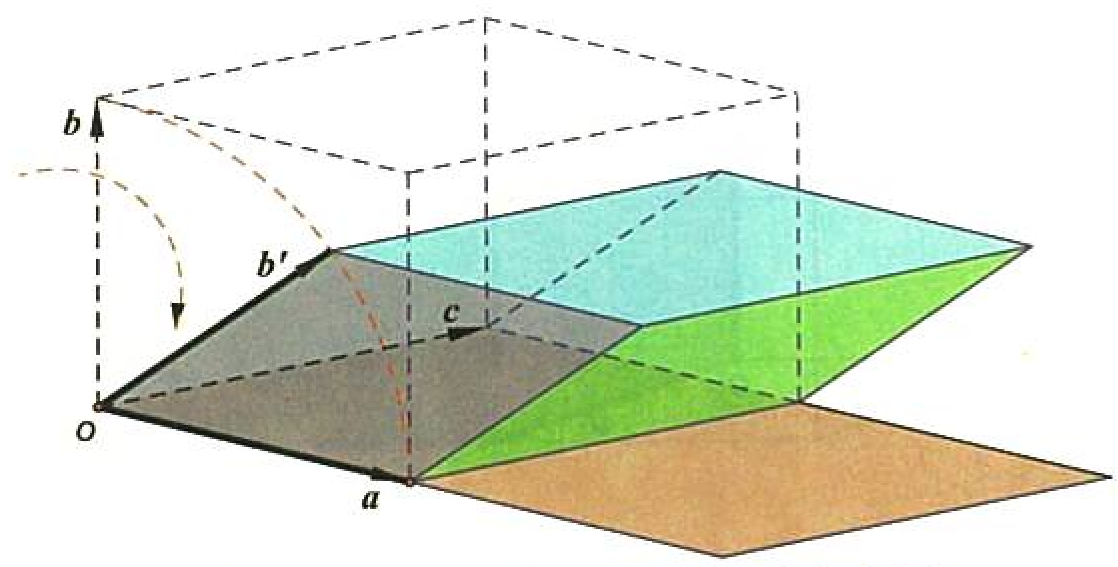


图 3-16 行向量相同的行列式为零

行列式的这些性质可以不经意地应用在实际生活中，如有人形容公共汽车的拥挤，开玩笑大喊“别挤了，再挤把人都挤成照片了”。这里面有物体维数的变化。人体是三维物体，体积不为零，挤成二维照片后，体积就变成了零。行列式也是这样：三阶行列式表示平行六面体的有向体积，如果其中有某两列相等或成比例，就是说平行六面体的三条相邻的棱中有两条重合或同向，平行六面体退化成平面图形，也就是被“挤成照片”了，体积变成零。类似地，二阶行列式表示平行四边形的有向面积，如果两列相等，“平行四边形”的相邻两边重合，平行四边形退化为一条线段，面积为零。一般地，n 阶行列式可以想象成一个 n 维超立方体的 n 维“体积”，如果它有某两列相等或成比例，则“n 维立方体”退化为 n-1 维或者更低维数的图形，“n 维体积”当然就等于零。

性质 3 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=-\det(\boldsymbol{b},\boldsymbol{a},\boldsymbol{c})$。

一个行列式对应着一个数值，这个数值是对行列式中的元素经过运算得到的。这个运算是与元素的位置有关系的，因此改变行列式中列向量或行向量的位置就会改变行列式的结果。幸而只改变结果的符号。一般地，一个行列式 $\det\boldsymbol{A}$ 的值对应矩阵 $\boldsymbol{A}$ 的列向量的一个固定顺序。当 $\det\boldsymbol{A}$ 为负值时，它确定原像的一个反射。所以，这种变换改变了原像的定向。

实际上，$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=-(\boldsymbol{b}\times\boldsymbol{a})\cdot\boldsymbol{c}=-\det(\boldsymbol{b},\boldsymbol{a},\boldsymbol{c})=\det(\boldsymbol{b},\boldsymbol{a},-\boldsymbol{c})$，因此如果交换了向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的位置，也就是改变了叉乘的顺序，因此叉乘结果改变了方向，也即改变了符号。

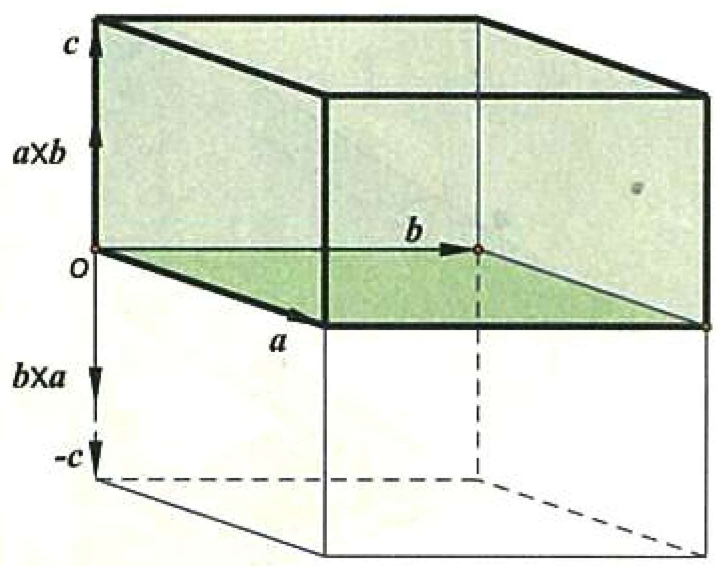


图 3-17 列向量交换位置引起行列式变号

如图 3-17 所示，阴影的平行六面体 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$、$\boldsymbol{a}\times\boldsymbol{b}$ 的方向和向量 $\boldsymbol{c}$ 同向(据右手法则)；当向量 $\boldsymbol{a},\boldsymbol{b}$ 的位置交换后，$\boldsymbol{b}\times\boldsymbol{a}$ 的方向与 $\boldsymbol{a}\times\boldsymbol{b}$ 相反，因而与向量 $\boldsymbol{c}$ 点乘后得到向下的平行六面体。所以平行六面体 $\det(\boldsymbol{b},\boldsymbol{a},\boldsymbol{c})$ 和 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ 以 $\boldsymbol{a}$、$\boldsymbol{b}$ 张成的平面为镜面互为反射。

性质 4 $k\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\det(k\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\det(\boldsymbol{a},k\boldsymbol{b},\boldsymbol{c})=\det(\boldsymbol{a},\boldsymbol{b},k\boldsymbol{c})$。

这就是说，平行六面体体积的 k 倍等于六面体的三条棱中一条棱长的 k 倍。这是显然的。因为立方体的体积增大可以沿着立方体某一棱方向增大相同的倍数，如图 3-18 所示。

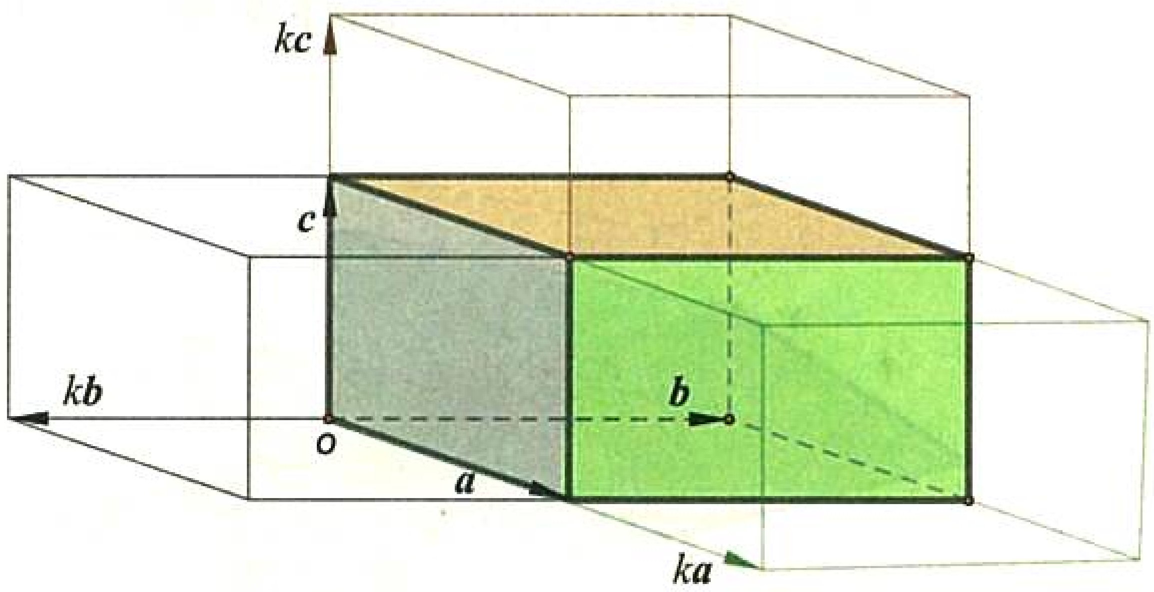


图 3-18 任一个列向量长度变化 k 倍，行列式值也增大 k 倍

性质 5　$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\det(\boldsymbol{a},\boldsymbol{b},k\boldsymbol{a}+\boldsymbol{c})$。

此性质表述了以 $\boldsymbol{a}\times\boldsymbol{b}$ 为底面积的平行六面体在 $\boldsymbol{a}$ 方向上进行了切向变换，变换后的六面体因为底面积不变，高也不变，所以体积不变。

如图 3-19（a）所示，原有六面体 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ 是由向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 张成的。向量 $\boldsymbol{a}$ 乘以一个负的 k 值后与向量 $\boldsymbol{c}$ 相加得到新的向量 $k\boldsymbol{a}+\boldsymbol{c}$，三个向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$k\boldsymbol{a}+\boldsymbol{c}$ 构成了一个新的平行六面体 $\det(\boldsymbol{a},\boldsymbol{b},k\boldsymbol{a}+\boldsymbol{c})$，这个六面体与原六面体 $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$ 同底等高，因而体积相同。

如图 3-19（b）所示，向量 $\boldsymbol{a}$ 乘以一个正的 k 值后与向量 $\boldsymbol{c}$ 相加的结果，图形类似。

通过观察，我们发现，切变后的平行六面体与 k 值无关。k 值不同，向量 $k\boldsymbol{a}+\boldsymbol{c}$ 的终端始终在一条与向量 $\boldsymbol{a}$ 平行的直线上滑动，因而保持了六面体的等高。

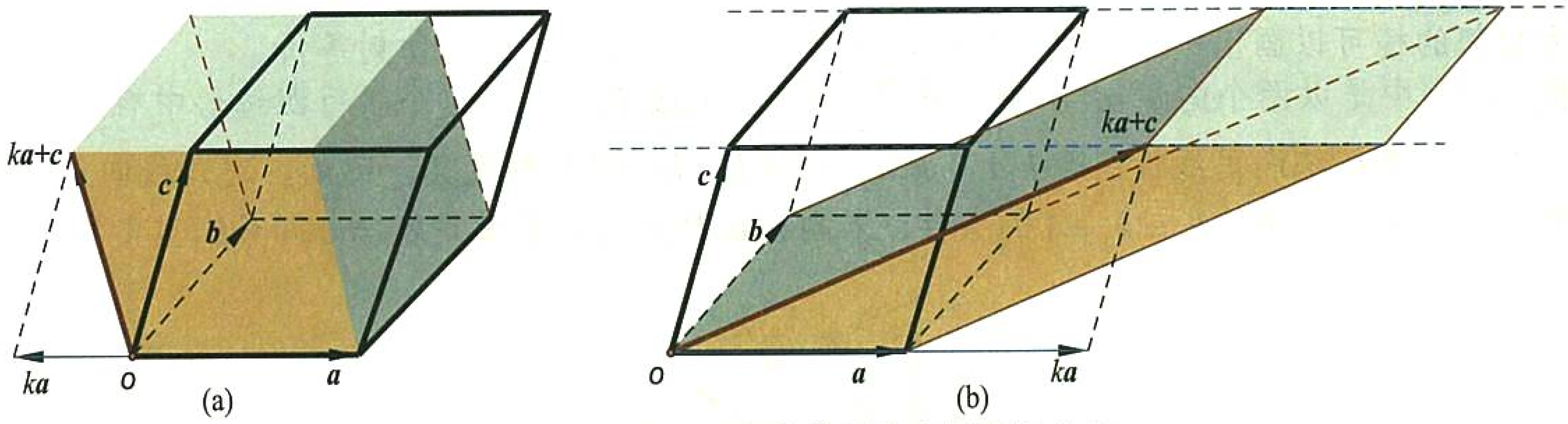


图 3-19 切向变换不改变行列式的值

性质 6　$\det\boldsymbol{A}=\det\boldsymbol{A}^{\mathrm{T}}$。

矩阵 $\boldsymbol{A}$ 的行列式等于矩阵 $\boldsymbol{A}$ 转置的行列式。也就是说，矩阵的行列式既可以看做 n 个行向量的行列式，也可以看做 n 个列向量的行列式。为了便于讨论，我们把三阶行列式及其转置行列式按照第一行元素展开式列举在这里：

$$\det\boldsymbol{A}=\begin{vmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\\c_1&c_2&c_3\end{vmatrix}=a_1\cdot\begin{vmatrix}b_2&b_3\\c_2&c_3\end{vmatrix}-a_2\cdot\begin{vmatrix}b_1&b_3\\c_1&c_3\end{vmatrix}+a_3\cdot\begin{vmatrix}b_1&b_2\\c_1&c_2\end{vmatrix}=a_1b_2c_3+a_2b_3c_1+a_3b_1c_2-a_1b_3c_2-a_2b_1c_3-a_3b_2c_1$$

$$\det\boldsymbol{A}^{\mathrm{T}}=\begin{vmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{vmatrix}=a_1\cdot\begin{vmatrix}b_2&c_2\\b_3&c_3\end{vmatrix}-b_1\cdot\begin{vmatrix}a_2&c_2\\a_3&c_3\end{vmatrix}+c_1\cdot\begin{vmatrix}a_2&b_2\\a_3&b_3\end{vmatrix}=a_1b_2c_3+b_1c_2a_3+c_1a_2b_3-a_1c_2b_3-b_1a_2c_3-c_1b_2a_3$$

上述两式显然是相等的。这个相等有里外两层几何意义可以解释。先说外层的几何解释。对照一下图 3-20（a）、（b）两图形，图 3-20（a）图是行列式展开时元素相乘的计算图，图 3-20（b）是行列式转置时以 $a_1b_2c_3$ 为反射轴的元素交换图。行列式转置顺序与计算顺序多么一致，怪不得不改变结果呢。

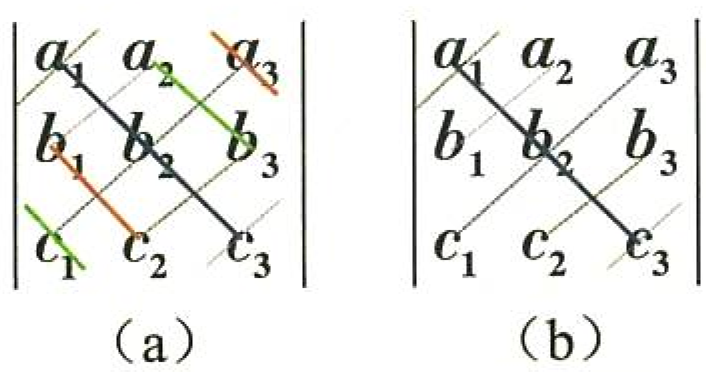


图 3-20 行列式计算顺序与转置的顺序一致

直接从行列式的定义“n 阶行列式实际上是不同行不同列的 n 个元素的乘积项的代数和”

的含义出发，大家是否有这个直觉：从行列式的每一行中取一个元素，使得这些元素又分别来自不同的列，这和行列式转置后从每一列中取一个元素，使得这些元素又分别来自不同的行所构成的乘积项是相同的。既然这样，行列式转置前后的结果也是相同的。

这个直觉引发了新的几何解释。因此行列式和它的转置相等还有一个深层次的几何意义，可以在下面逆序数的几何意义中得到一些体现。行列式的乘积项及其逆序数的几何意义实际上是行列式最根本的几何意义，因而可以解释所有的行列式的定义及其性质。后面的章节我们会逐步探讨到这个重要的课题。

什么是向量张成的平行四边形及平行六面体

以两个从原点出发的向量为邻边可以画出一个平行四边形；以三个从原点出发的不同方向向量为相邻的棱可以画出一个平行六面体，这个平行六面体与立方体仿射等价。

图 3-21 中是以两个向量 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 张成的平行四边形 $o\boldsymbol{\alpha}_1P\boldsymbol{\alpha}_2$，在这个平行四边形中有无数的向量，这些向量统统满足线性组合 $k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2$（$0\leqslant k_1,k_2\leqslant 1$）的定义。比如，对角线向量 $\overrightarrow{0P}$ 是长度最大的向量，满足 $k_1=k_2=1$ 的条件；末端在 $\boldsymbol{\alpha}_1P$ 上的向量满足 $k_1=1, 0\leqslant k_2\leqslant 1$ 的条件等等。

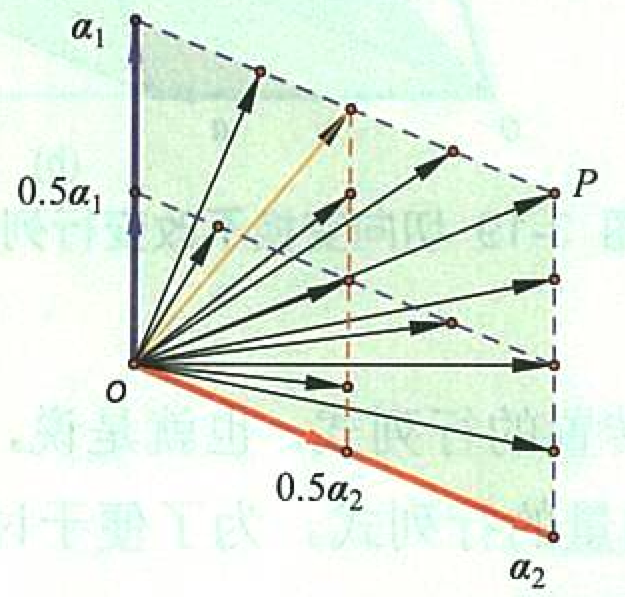


图 3-21 两个向量张成一个以向量为邻边的平行四边形

我们可以推广到 n 维欧几里德空间中的 m 个向量张成的 m 维超平行多面体。如果这 m 个向量表示为 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$，那么这个 m 维超平行多面体是由以下无穷个向量

$$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_m\boldsymbol{\alpha}_m \quad (0\leqslant k_1,k_2,\cdots,k_m\leqslant 1)$$

组成的。

m 维超平行多面体仿佛是一个枝繁叶茂的大树所构成的一个物理空间，主枝干就是 m 个向量的向量组，其它的大大小小的枝杈叶芽是由向量组这个主枝干扩张生长出来的无穷个向量。这就像，一棵大树围占的实体空间的体积主要是由这棵树的主枝干所决定的。

在后面的章节中，我们还要讨论由向量张成的空间的概念。这时候，实数 k_i 的范围不是在区间 $[0,1]$ 之上了，而是整个实数域 $\mathbf{R}$，因此空间的范围也将变得无穷大了。

3.4 行列式化为对角形的几何解释

一个行列式第 i 行加上 j 行的 k 倍，可以使第 i 行的某一个元素变为 0，而这个行列式的值不变。这个性质在化简行列式时非常有用。在前面的二阶和三阶行列式的性质中我们也已看到了它的几何意义。不过前面对它的几何解释一般都是使用向量的“箭头”的图形形式给出的，没有考虑向量的元素的变化的几何意义。本节我们就此性质的应用来探究行列式中每一个元素的几何意义。

把一般二阶行列式化为上三角形式的行列式，只要把行列式左下角元素即第二行第一列化为 0 就行了。因此，第一行向量元素(a_1, a_2)乘以$-\dfrac{b_1}{a_1}$（设$a_1 \neq 0$）后加到第二行上，即可得到如下的三角式：

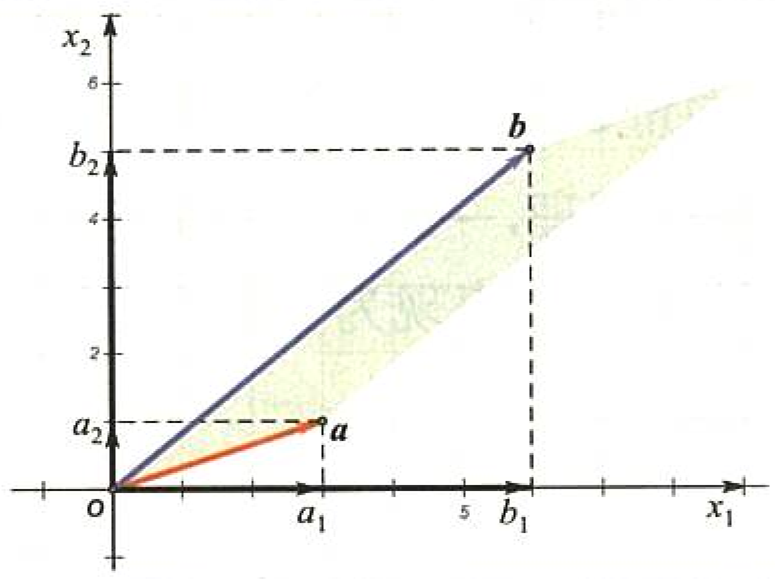
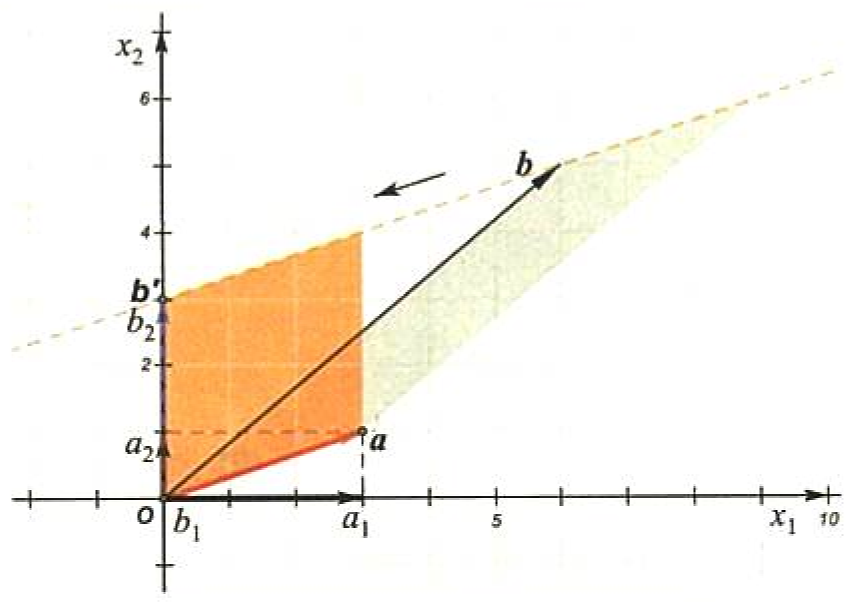
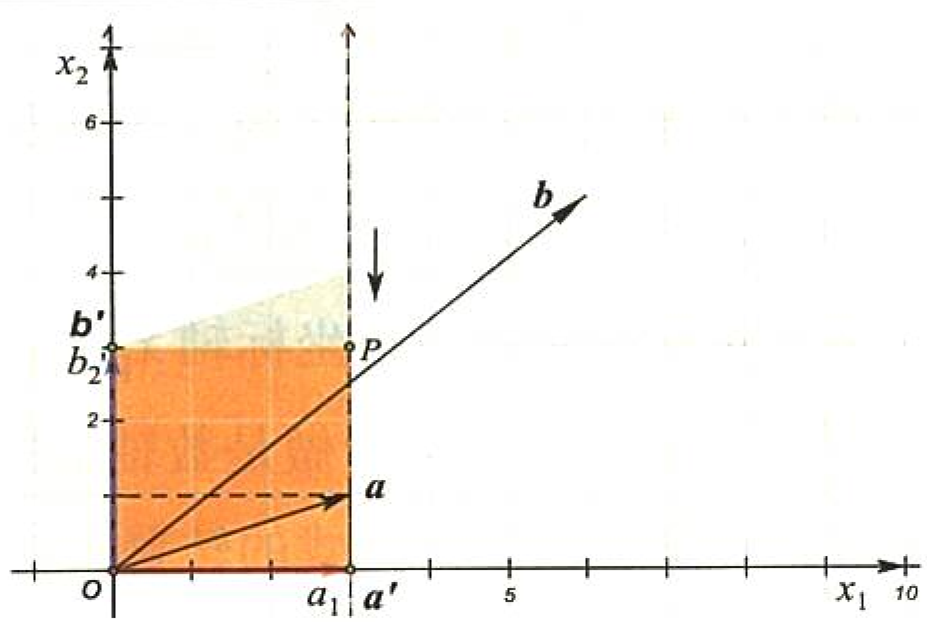
$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ 0 & \dfrac{a_1b_2 - a_2b_1}{a_1} \end{vmatrix}$$

继续化简，把第二行的元素$\left(0, \dfrac{a_1b_2 - a_2b_1}{a_1}\right)$乘以$-\dfrac{a_1a_2}{a_1b_2 - a_2b_1}$后加到第一行上，得到如下的对角形行列式：

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & 0 \\ 0 & \dfrac{a_1b_2 - a_2b_1}{a_1} \end{vmatrix}$$

这个化简的过程我们可以有一个很形象的几何变化过程，如表 3-1 所示。

表 3-1 二阶行列式化为对角形过程的几何解释

顺序	描述	对应的几何图形
1	一个二阶行列式 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ 其值等于右图阴影平行四边形的有向面积。它是由行向量$\boldsymbol{a}$、$\boldsymbol{b}$张成的。	
2	化简到如下的上三角式：$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ 0 & \dfrac{a_1b_2 - a_2b_1}{a_1} \end{vmatrix}$ 其几何变化过程是向量$\boldsymbol{b}$的末端沿着$o\boldsymbol{a}$的平行线$\boldsymbol{b}'\boldsymbol{b}$滑行到$x_2$坐标轴上（切变），此时$b_1$化为 0。显然，两个阴影平行四边形的面积不变。	
3	继续化简到对角形：$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & 0 \\ 0 & \dfrac{a_1b_2 - a_2b_1}{a_1} \end{vmatrix}$ 其几何变化过程是向量$\boldsymbol{a}$的末端沿着$o\boldsymbol{b}'$的平行线$\boldsymbol{a}'\boldsymbol{a}$滑行到$x_1$坐标轴上，此时$a_2$化为 0。显然，三个阴影平行四边形的面积相同。	
4	至此，一个二阶行列式所表示的平行四边形被变成了一个对角行列式所表示的长方形（本例为正方形oa_1Pb_2'，其中$b_2' = \dfrac{a_1b_2 - a_2b_1}{a_1}$）。	

三阶行列式有类似的变换情形，对角化的过程会把一个平行六面体变化为一个等体积的立方体或长方体。具体的图形不再绘出。

那么 n 阶行列式我们亦不怀疑地认为也可以被表示成一个 n 维的长方体的几何图形。这个结论对于我们理解行列式的展开式很有帮助。

3.5行列式乘积项的几何意义

n 阶行列式的展开式是乘积项 $(-1)^t a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 的和，这些乘积项其实也可以有几何解释的。下面我们详细地了解一下各阶行列式的情况。

3.5.1 二阶行列式乘积项的几何意义

对于二阶行列式而言，既然二阶行列式的几何图形是一个有方向的面积，那么从二阶行列式公理化定义 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$ 看，又是如何构成这个面积的呢？显然，式中 a_1b_2 项和 $-a_2b_1$ 项的和构成了这个面积。a_1b_2 项和 $-a_2b_1$ 项又是什么呢？由图 3-22 易知，a_1 是向量 $\boldsymbol{a}$ 在 x_1 轴上的投影，b_2 是向量 $\boldsymbol{b}$ 在 x_2 轴上的投影，a_1b_2 项是 $\boldsymbol{a}$ 在 x_1 轴上的投影和 $\boldsymbol{b}$ 在 x_2 轴上的投影所围成的长方形面积。同理，$-a_2b_1$ 项是 $\boldsymbol{a}$ 在 x_2 轴上的投影和 $\boldsymbol{b}$ 在 x_1 轴上的投影所围成的长方形面积，不过这个面积表达式表现为负数的表达式。

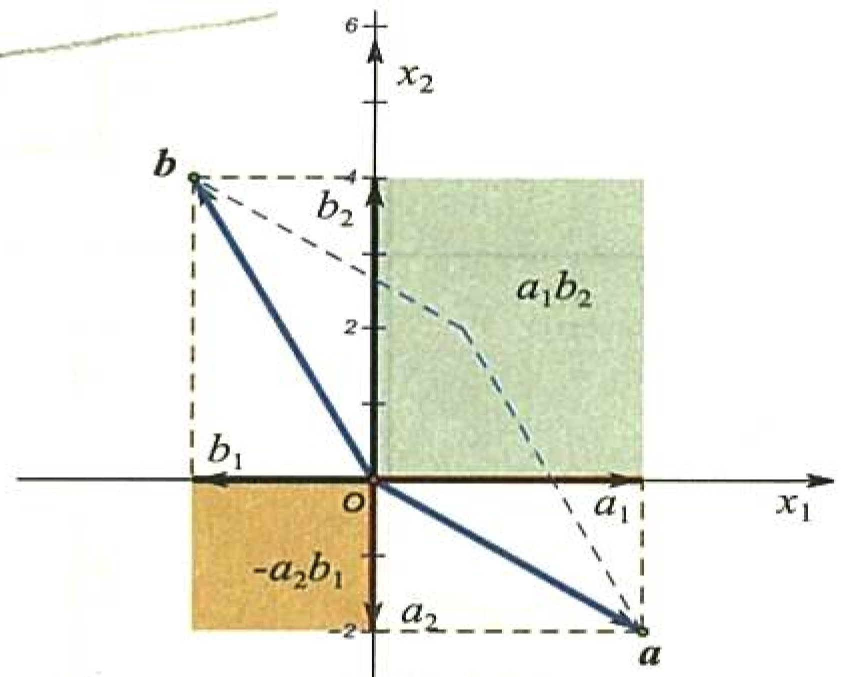


图 3-22 二阶行列式乘积项是有向面积块

这里有两个问题要注意：

1) 因为二阶行列式表达的是向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 所围成的面积，因此向量 $\boldsymbol{a}$ 自己的分向量 a_1 和 a_2 围成的面积 a_1a_2 不在统计范围之内，向量 $\boldsymbol{b}$ 自己的分向量围成的面积 b_1b_2 也不在统计范围之内。

2) 向量 $\boldsymbol{a}$ 的分向量 a_1 和向量 $\boldsymbol{b}$ 的分向量 b_1 围成的面积 a_1b_1 要统计，但可惜的是 $a_1b_1=0$，因为两个分向量投影到同一根坐标轴 x_1 上。同理，$a_2b_2=0$。（另外，向量 $\boldsymbol{a}$ 在坐标轴 x_1 上的投影 a_1 既是数轴上的数，又可看做是数轴上的分向量 $\boldsymbol{a}_1$，这里无需分开）。

前面我们在解释二阶行列式的转置时讲过，两个向量的叉积等于这两个向量的各个分量（各坐标轴方向的分量）分别进行叉积的求和。各个分量互相垂直，因而进行叉积运算张成的四边形是长方形的面积。向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的叉积（有向面积）等于每个向量的分向量所有可能的叉积之和。因为在直角坐标系下，向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 落在同一坐标轴上的分量叉积显然为零，因此进行叉积的分量必然是不在同一坐标轴上的分量。

a_1b_2 项和 $-a_2b_1$ 项都是有向面积，a_1b_2 项的面积和此二阶行列式的面积方向相同，因此符号为正，$-a_2b_1$ 项的面积和此二阶行列式的面积方向相反，因此符号为负。

那么二阶行列式的面积及其乘积项的面积的方向如何确定呢？对二阶行列式来说，有不同的方法可以确定，这里使用叉积的右手法则来判断。我们把行列式$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$看做由两个行向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 所组成，其中排在第一行的元素组成第一个向量 $\boldsymbol{a}=(a_1,a_2)$，排在第二行的元素组成第二个向量 $\boldsymbol{b}=(b_1,b_2)$（看成两个列向量也有类似的结果，因为转置行列式值不变）。这是一个顺序，不要错了。那么行列式$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$的面积方向就是由第一向量转向第二向量时右手大拇指所确定的方向。

如图 3-23（a）所示，行列式的方向为指向读者；如图 3-23（b）所示，行列式的方向指向页面内，背向读者。

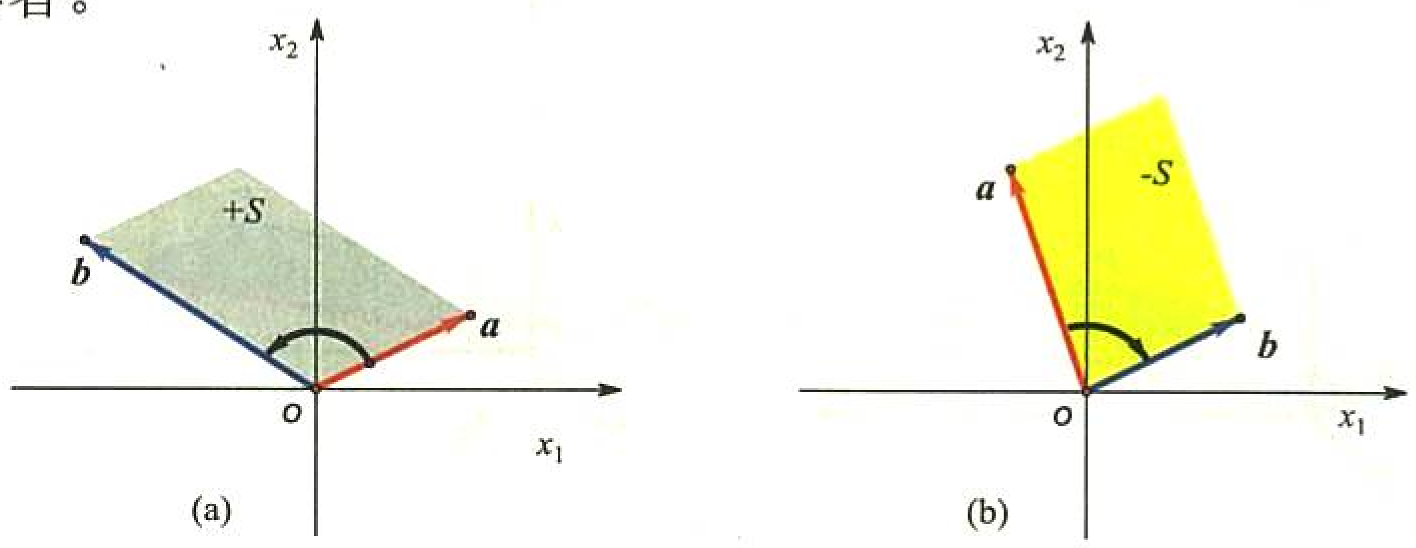


图 3-23 行列式是其行向量的叉积

图 3-24 中，行列式的方向是指向读者，那么，分向量 $\boldsymbol{a}_1$ 和 $\boldsymbol{b}_2$ 张成的长方形（右上角图块）面积的方向也是指向读者的，因此面积为正，故记为 a_1b_2；分向量 $\boldsymbol{a}_2$ 和 $\boldsymbol{b}_1$ 张成的长方形（左下角图块）面积的方向指向页面内，背向读者，与行列式方向相反，因此面积为负，故记为 $-a_2b_1$。

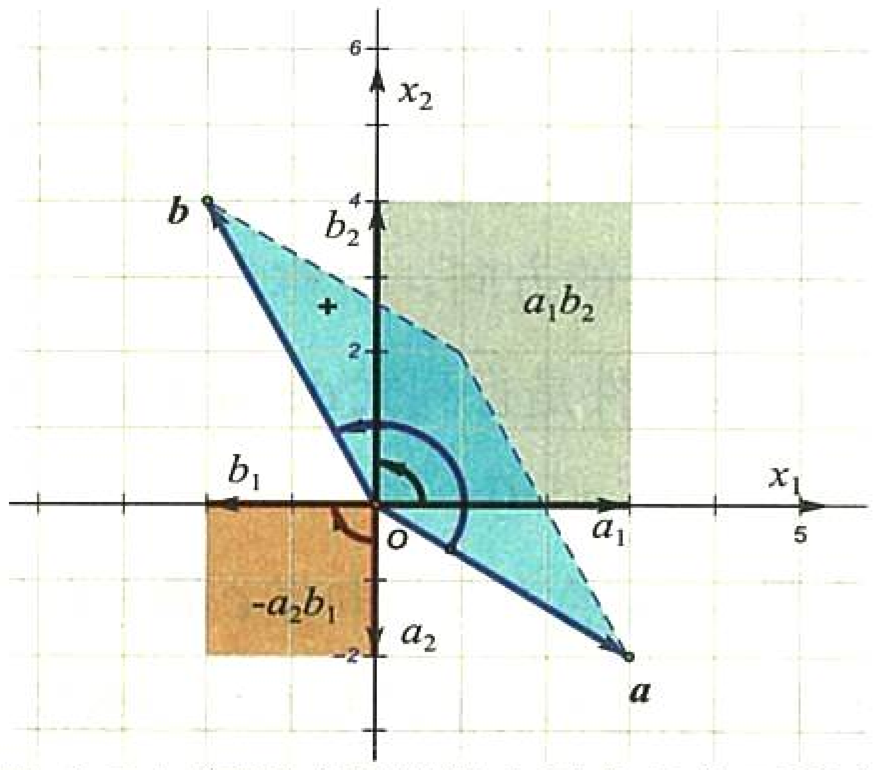


图 3-24 行列式是其行向量分量的叉积之和

3.5.2 三阶行列式乘积项的几何意义

依据三阶行列式的公式$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$，与二阶行列式的乘积项的几何解释类似，三阶行列式的乘积项如 $a_1b_2c_3$、$a_2b_3c_1$ 等，可以看成具有有方向

的小长方体的体积。也就是说，由行列式的三个向量张成的三维平行六面体可以分解为一个个由各坐标分量混合积构成的小长方体。这些小长方体共有六块，其体积具有方向。

既然平行六面体和小长方体都是具有方向的三维几何图形，那么这些方向是如何确定呢？

与上一节的论述类似，平行六面体的方向是由张成这些三维几何图形的有序的三元向量组所决定的。三个无关向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$的有序组根据右手法则可以确定一个方向。这个确定的方法类同于叉积定义的方法：

如图 3-25 所示，从向量$\boldsymbol{a}$的方向转过一个$0\sim\pi$之间的角度到达向量$\boldsymbol{b}$的方向，右手拇指方向是向量$\boldsymbol{a}$、$\boldsymbol{b}$所确定平面的正向。如果向量$\boldsymbol{c}$在此平面正向的一侧，则$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$的有序组就确定了一个正的方向（见图 3-25（a））；如果向量$\boldsymbol{c}$在正向平面的另一侧，则$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$的有序组就确定了一个负的方向（见图 3-25（b））。和向量的叉积联系起来，第三个向量$\boldsymbol{c}$与向量$\boldsymbol{a}\times\boldsymbol{b}$在$\boldsymbol{a}$、$\boldsymbol{b}$所确定平面的同一侧时确定一个正的方向。

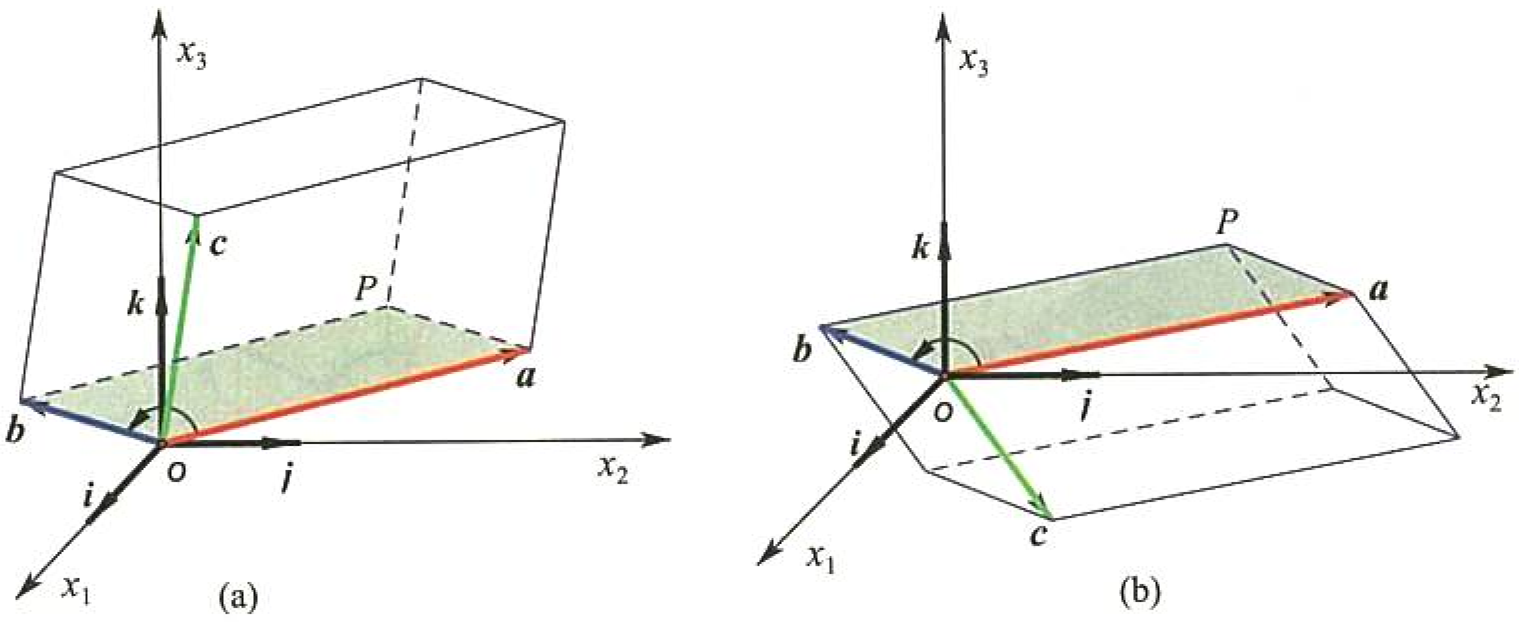


图 3-25 三阶行列式正负值的几何意义

换一种说法就是，从向量$\boldsymbol{c}$的逆方向看过去，向量$\boldsymbol{a}$旋转过一个$0\sim\pi$之间的角度到达向量$\boldsymbol{b}$的方向符合逆时针方向。因此，我们有一些结论：

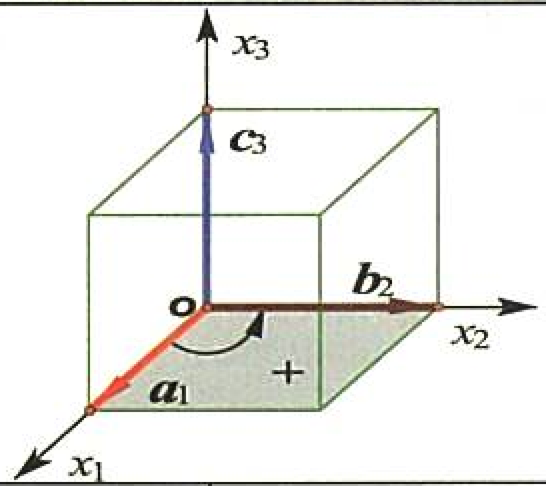
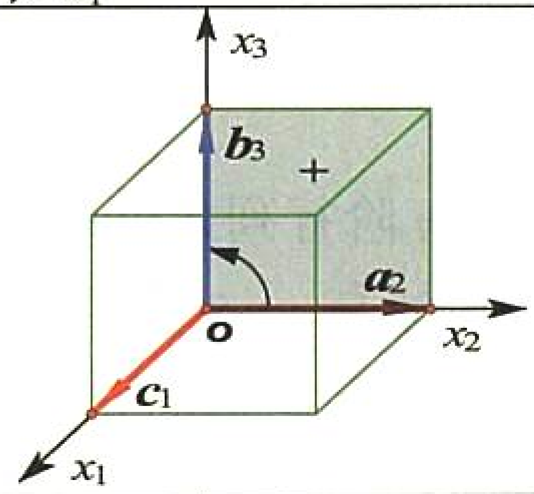
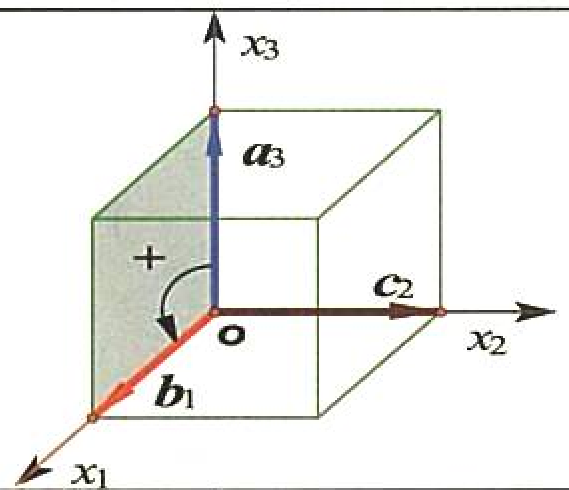
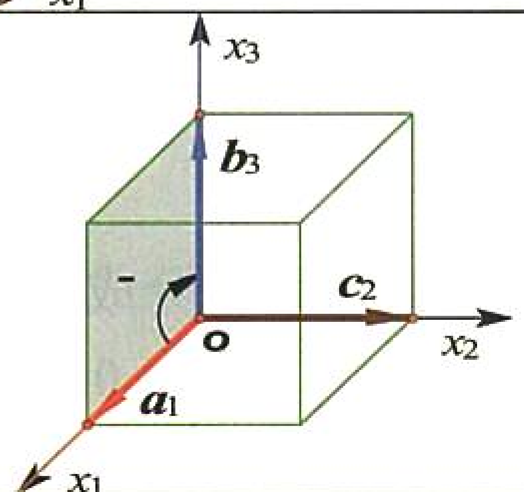
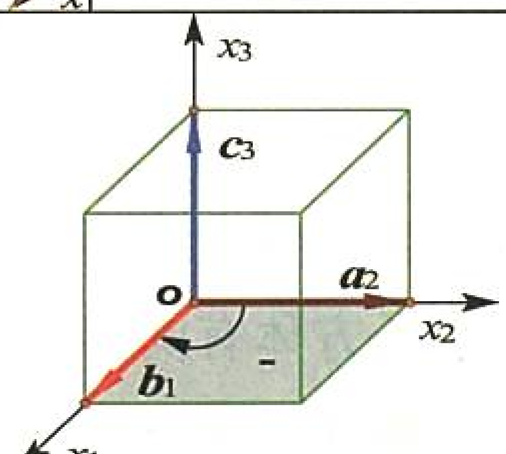
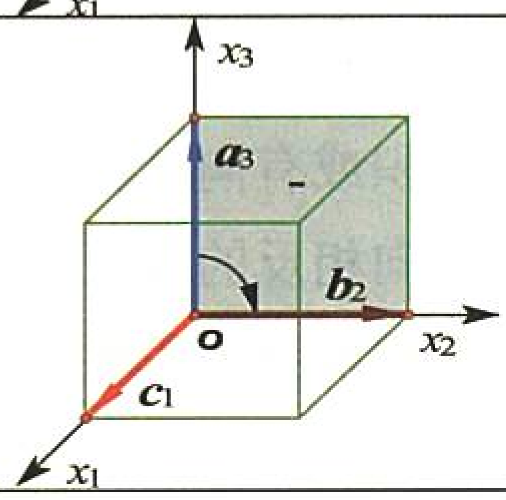
- $\boldsymbol{b}$、$\boldsymbol{a}$、$\boldsymbol{c}$有序组与$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$有序组具有相反的方向。
- $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{a}\times\boldsymbol{b}$有序组与右手系三维坐标轴向量的有序组$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$的方向相同。
- 如果$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$有序组与坐标轴向量有序组$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$有同样的定向，我们就称$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$有序组相对于坐标系(x_1,x_2,x_3)有正向；如果有相反的方向，就称为负向。
- $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$有序组有正向的充分必要条件是行列式$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})>0$。
- 行列式$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})>0$意味着$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}>0$。

$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}>0$，也就是说，向量$\boldsymbol{a}\times\boldsymbol{b}$的方向与向量$\boldsymbol{c}$的方向作成一个锐角，也就是向量$\boldsymbol{c}$与向量$\boldsymbol{a}\times\boldsymbol{b}$在$\boldsymbol{a}$、$\boldsymbol{b}$所确定平面的同一侧,这和前面的解释是一致的。

同样，这些小长方体的方向是由三个坐标轴上的向量投影所确定：与坐标轴向量有序组$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$有同样的定向，我们就称$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$有序组相对于坐标系(x_1,x_2,x_3)有正向；如果有相反的方向，就称为负向。因为小长方体是由三个坐标轴上的三个不同向量的投影向量所张成的，三个投影分量互相垂直，因此我们可以方便地使用右手法则确定其方向。如果是正向，则长方体体积的乘积项大于 0，否则小于 0。

对于行列式的展开式$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$，表 3-2 中画出了此公式六个乘积项$\pm a_ib_jc_k$的几何图形，六个小长方体的方向确定了乘积项的符号。

表 3-2 三阶行列式乘积项的几何意义

顺序	乘积项及其解释	对应的几何图形
1	$+a_1b_2c_3$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_1,\boldsymbol{b}_2,\boldsymbol{c}_3\}$与坐标系方向一致，乘积项为正	
2	$+a_2b_3c_1$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_2,\boldsymbol{b}_3,\boldsymbol{c}_1\}$与坐标系方向一致，乘积项为正	
3	$+a_3b_1c_2$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_3,\boldsymbol{b}_1,\boldsymbol{c}_2\}$与坐标系方向一致，乘积项为正	
4	$-a_1b_3c_2$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_1,\boldsymbol{b}_3,\boldsymbol{c}_2\}$与坐标系方向相反，乘积项为负	
5	$-a_2b_1c_3$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_2,\boldsymbol{b}_1,\boldsymbol{c}_3\}$与坐标系方向相反，乘积项为负	
6	$-a_3b_2c_1$ 项： 以右手法则，分向量有序组$\{\boldsymbol{a}_3,\boldsymbol{b}_2,\boldsymbol{c}_1\}$与坐标系方向相反，乘积项为负	

3.5.3 n阶行列式乘积项的几何意义

n阶行列式的展开式是乘积项 $(-1)^t a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 的和，t为排列 $j_1j_2\cdots j_n$ 的逆序数。行列式的行或列向量在坐标系中向各个坐标轴的投影就是行列式中的元素 $a_{1j_1},a_{2j_2},\cdots,a_{nj_n}$，这些元素当然也是行或列向量在各个坐标轴上的分向量。这些坐标轴上的分向量也可以张成一个个小小的 n 维超长方体（共 $n!$个）。我们刚好知道，n 阶行列式的超平行多面体的几何图形是由行(或列)向量张成的，而且这个 n 维超平行多面体与一个 n 维超长方体等体积。那么大家就会容易得到以下的认识：

- $n!$个小的 n 维长方体叠加成了大的表示 n 阶行列式的 n 维超长方体；
- $n!$个小的 n 维长方体就是 $n!$个乘积项 $(-1)^t a_{1j_1}a_{2j_2}\dots a_{nj_n}$ 的几何图形。

实际上，一个 n 阶行列式可以分拆成 $n!$个只有坐标轴分向量组成的 n 阶对角行列式，而 $n!$个 n 阶对角行列式就是 $n!$个乘积项 $(-1)^t a_{1j_1}a_{2j_2}\dots a_{nj_n}$。

比如，一个二阶行列式可以分拆成两个这样的二阶对角行列式（参见 3.2.2 节中的行列式加法）：

$$
\begin{aligned}
\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} &= \begin{vmatrix} a_1 & 0 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ b_1 & b_2 \end{vmatrix} \\
&= \begin{vmatrix} a_1 & 0 \\ b_1 & 0 \end{vmatrix} + \begin{vmatrix} a_1 & 0 \\ 0 & b_2 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ b_1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ 0 & b_2 \end{vmatrix} \\
&= \begin{vmatrix} a_1 & 0 \\ 0 & b_2 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ b_1 & 0 \end{vmatrix} \\
&= a_1b_2 - a_2b_1
\end{aligned}
$$

一个三阶行列式可以拆分成六个（其余的行列式值等于零）三阶对角行列式：

$$
\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & 0 & 0 \\ 0_1 & b_2 & 0 \\ 0 & 0 & c_3 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & 0 \\ 0 & 0 & b_3 \\ c_1 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_3 \\ b_1 & 0 & 0 \\ 0 & c_2 & 0 \end{vmatrix} + \begin{vmatrix} a_1 & 0 & 0 \\ 0 & 0 & b_3 \\ 0 & c_2 & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & 0 \\ b_1 & 0 & 0 \\ 0 & 0 & c_3 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_3 \\ 0 & b_2 & 0 \\ c_1 & 0 & 0 \end{vmatrix}
$$

$$
= a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1
$$

因此，二阶行列式的有向面积是由两块小长方形有向面积累加的代数和。三阶行列式的有向体积是由六块小长方体有向体积累加的代数和。

一个行列式的整体几何意义是有向线段（一阶行列式）或有向面积（二阶行列式）或有向体积（三阶行列式及以上）。因此，行列式最基本的几何意义是由各个坐标轴上的有向线段所围起来的所有有向面积或有向体积的累加和。这个累加要注意每个面积或体积的方向或符号，方向相同的要加，方向相反的要减，因而，这个累加的和是代数和。

决定每一个面积块和体积块的符号或方向的因素是乘积项中元素的排列顺序，这个排列顺序决定了逆序数。通过上面行列式的拆解式，我们看到，每一个乘积项对应一个同样阶数的行列式，乘积项中每一个元素对应着一个向量（在坐标轴上的分量），向量在作外积运算后的方向决定了乘积项的符号。

3.6 拉普拉斯展开定理及代数余子式的几何解释

由前面的分析知道，三阶行列式有如下的关系（即拉普拉斯展开式）：

$$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\begin{vmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{vmatrix}=(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}=c_1\begin{vmatrix}a_2 & b_2\\ a_3 & b_3\end{vmatrix}-c_2\begin{vmatrix}a_1 & b_1\\ a_3 & b_3\end{vmatrix}+c_3\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}$$

其中

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{pmatrix}a_1\\ a_2\\ a_3\end{pmatrix}\times\begin{pmatrix}b_1\\ b_2\\ b_3\end{pmatrix}=\left(a_2b_3-a_3b_2,\ -(a_1b_3-a_3b_1),\ a_1b_2-a_2b_1\right)=\left(\begin{vmatrix}a_2 & b_2\\ a_3 & b_3\end{vmatrix},-\begin{vmatrix}a_1 & b_1\\ a_3 & b_3\end{vmatrix},\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}\right)$$

也就是说，拉普拉斯展开式就是最后一个向量 $\boldsymbol{c}$ 与另外两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的叉积向量的内积公式。

那么，我们怎样理解这个行列式降阶展开式的几何意义呢？

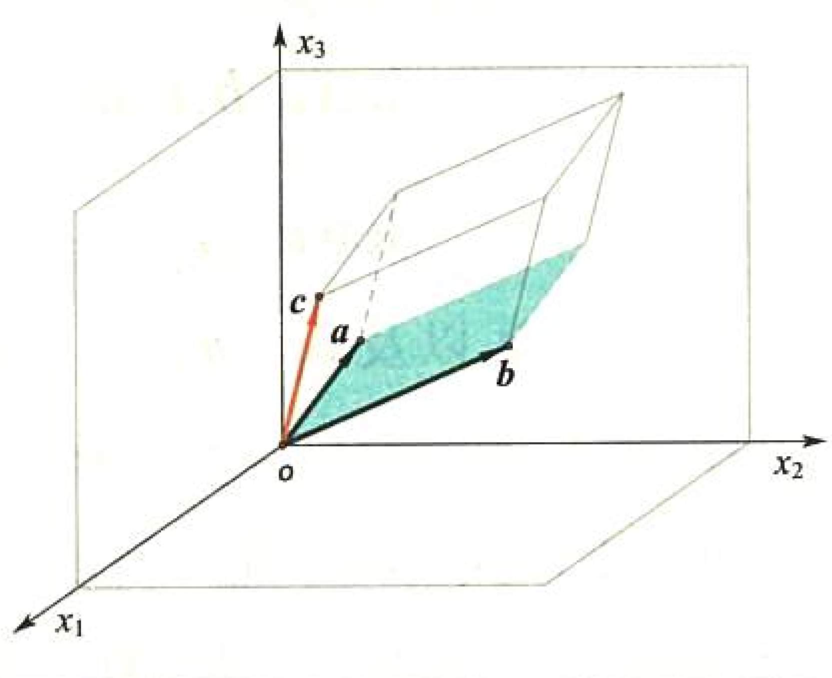
回想到三阶行列式的几何意义是三维空间中的一个平行六面体的体积，向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 的混合积 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 的混合过程告诉我们这个平行六面体是如何张成的：把三维空间中的一个平行四边形进一步叠加成（或张成）空间中的一个平行六面体。

好，两个方面的思路结合起来，拉普拉斯的展开定理就是把上述混合积的过程进行了投影。详细说来：先把 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 张成的三维空间中的平行四边形（$\boldsymbol{a}\times\boldsymbol{b}$ 的结果）分别向三个坐标平面进行投影（三个投影即 $\boldsymbol{a}\times\boldsymbol{b}$ 的三个分量，详见 2.4.1 节），然后再把向量 $\boldsymbol{c}$ 向三个坐标轴进行投影（就是向量 $\boldsymbol{c}$ 的坐标），最后三个投影平面分别乘以与之垂直的三个向量 $\boldsymbol{c}$ 的投影，得到了三个小平行六面体之体积。这三个小平行六面体就是三阶行列式的拉普拉斯展开的三个乘式。

$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 换一种说法就是“面”向量（$\boldsymbol{a}\times\boldsymbol{b}$）点积于“线”向量 $\boldsymbol{c}$，得到体积值。所以，三阶行列式的拉普拉斯展开过程就是混合积 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 的过程。

表 3-3 中的系列图形给出了拉普拉斯展开式的几何解释：三阶行列式所表述的平行六面体的体积可以分解为三个小的平行六面体的体积的和。

表 3-3 拉普拉斯展开式的几何解释

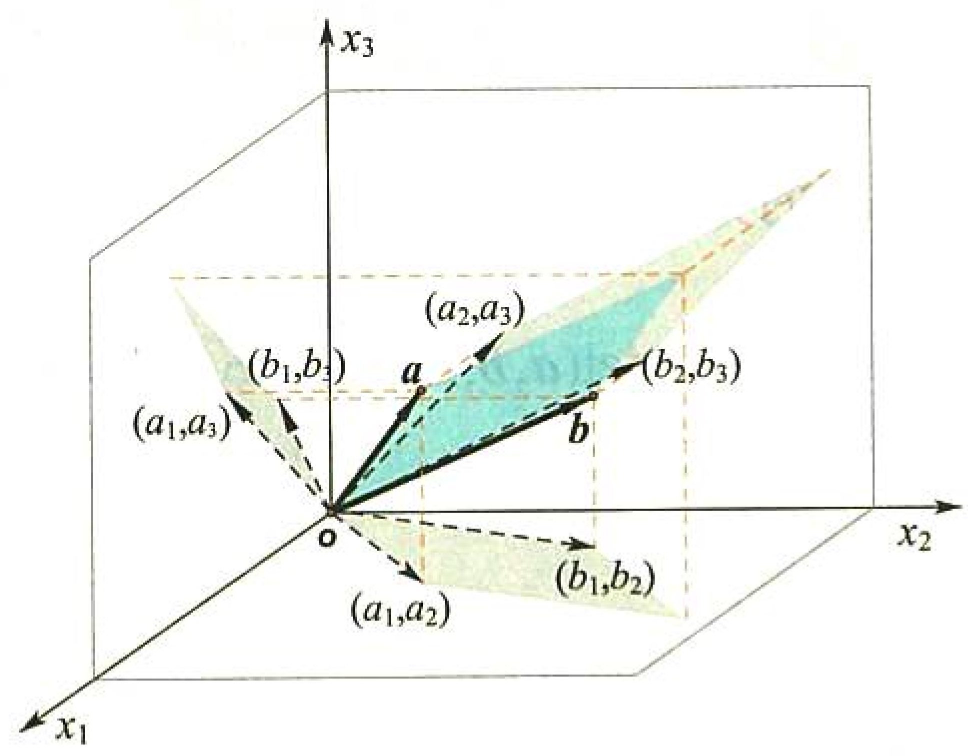
顺序	图形解释	对应的几何图形
1	列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 构成的三阶行列式： $\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\begin{vmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{vmatrix}$ 其几何意义是三个向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 张成的空间平行六面体。行列式展开的过程是混合积 $(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 的过程。 下面就混合积的过程分别给出其几何图形的解释	

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

$\boldsymbol{a}\times\boldsymbol{b}$ 的几何图解：

由向量$\boldsymbol{a}$、$\boldsymbol{b}$张成的空间平行四边形，其面积为$\boldsymbol{a}\times\boldsymbol{b}$ 。这个空间平行四边形向三个坐标平面同时进行投影，得到了三个小平行四边形，各位于坐标平面x_2ox_3、x_3ox_1、x_1ox_2上，各个小平行四边形的面积分别为
$\begin{vmatrix}a_2 & b_2\\ a_3 & b_3\end{vmatrix}, -\begin{vmatrix}a_1 & b_1\\ a_3 & b_3\end{vmatrix}, \begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}$，这三个行列式又分别被称为$c_1$、$c_2$、$c_3$的代数余子式

</td>
<td>

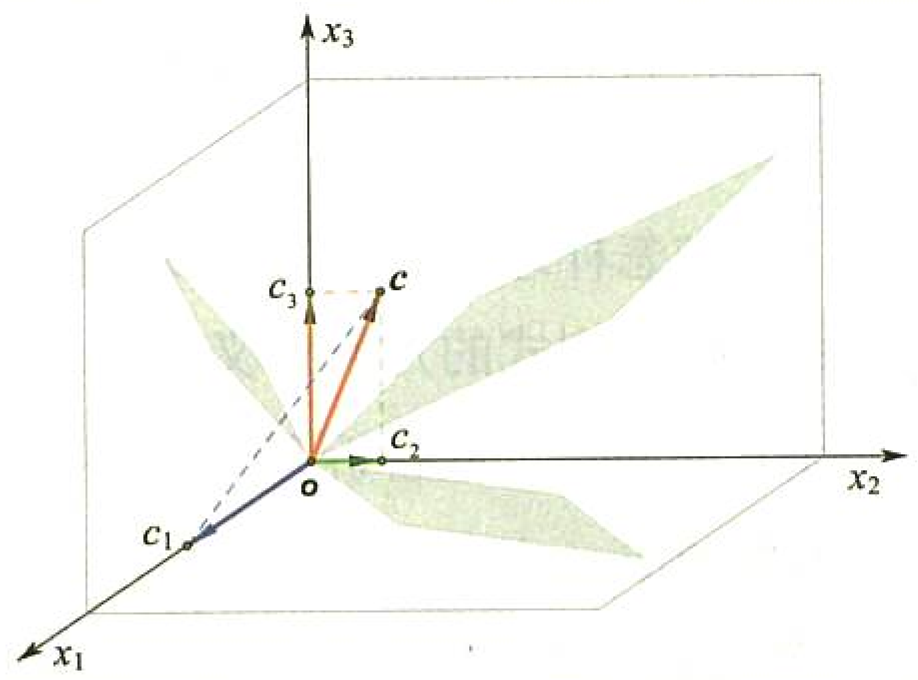


</td>
</tr>
<tr>
<td>3</td>
<td>

再说向量$\boldsymbol{c}$的图解：

第三个向量$\boldsymbol{c}$也同时向三个坐标轴投影，x_1、x_2、x_3轴上的投影分量分别是c_1、c_2、c_3

</td>
<td>



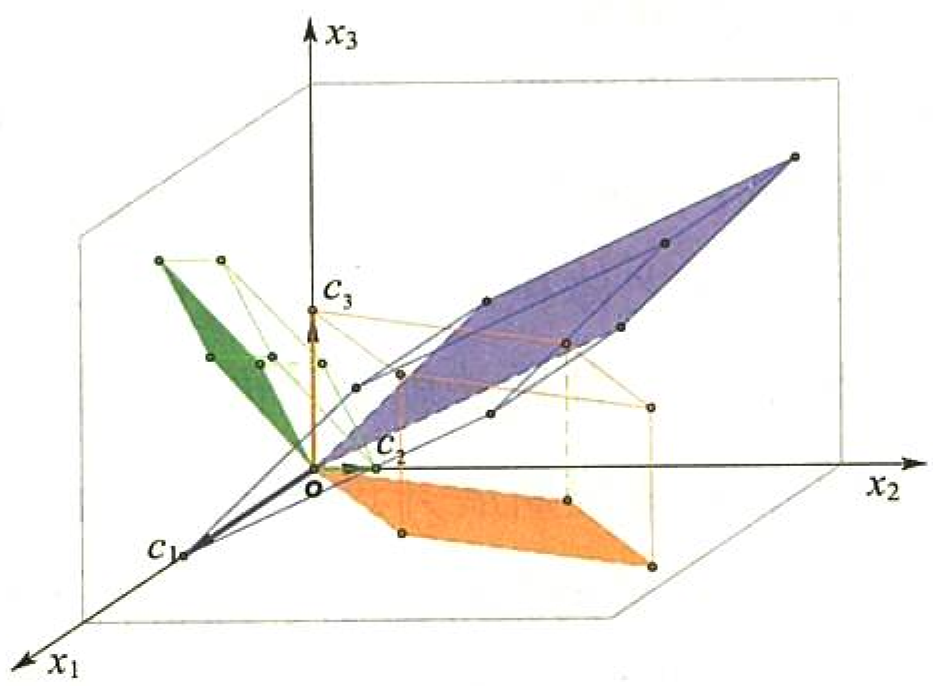
</td>
</tr>
<tr>
<td>4</td>
<td>

最后说混合积$(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 的过程：

分量$\boldsymbol{c}_1$ 与x_2ox_3上的小平行四边形重新张成一个小平行六面体。因分量$\boldsymbol{c}_1$ 垂直于这个平行四边形底面积，故其体积恰好等于
$c_1\begin{vmatrix}a_2 & b_2\\ a_3 & b_3\end{vmatrix}$。

同理，有另外两个小平行六面体，体积分别是$-c_2\begin{vmatrix}a_1 & b_1\\ a_3 & b_3\end{vmatrix}$ 和$c_3\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}$

</td>
<td>



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

从中，我们也看到了代数余子式的几何意义。三阶行列式的代数余子式是两个行向量叉积的平行四边形在每个坐标平面上的投影。

进一步地，把拉普拉斯展开式扩展到四阶及 n 阶，有以下的定义式：

$$\det(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})=(\boldsymbol{a}\times\boldsymbol{b}\times\boldsymbol{c})\cdot\boldsymbol{d}$$
$$\vdots$$
$$\det(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_{n-1},\boldsymbol{a}_n)=(\boldsymbol{a}_1\times\boldsymbol{a}_2\times\cdots\times\boldsymbol{a}_{n-1})\cdot\boldsymbol{a}_n$$

式中，$\boldsymbol{a}\times\boldsymbol{b}\times\boldsymbol{c}$ 以及$\boldsymbol{a}_1\times\boldsymbol{a}_2\times\cdots\times\boldsymbol{a}_{n-1}$ 被称为三个以至 $n-1$ 个向量的向量积或外积。$\boldsymbol{a}\times\boldsymbol{b}\times\boldsymbol{c}$ 与向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$分别正交；$\boldsymbol{a}_1\times\boldsymbol{a}_2\times\cdots\times\boldsymbol{a}_{n-1}$ 也分别与$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_{n-1}$ 正交。前面说过，超过三维以上，我们将不能几何地定义叉积的方向，但我们仍然可以知道，三个以上的 n 维空间的向量积的长度仍然可以几何地解释为是 $n-1$ 个向量$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_{n-1}$ 所张成的 $n-1$ 维平行多面体的体积。

3.7 克莱姆法则的几何意义

1750 年，瑞士的克莱姆发现了用行列式求解线性方程组的克莱姆（Cramer）法则。这个法则在表述上简洁自然，思想深刻，包含了对多重行列式的计算，是对行列式与线性方程组之间关系的深刻理解。如果我们不能从几何上解释这个法则，就难以领会向量、行列式和线性方程组之间的真正关系。

3.7.1　二阶克莱姆法则的几何解释

对于二阶线性方程组：$\begin{cases} a_1x+b_1y=c_1 \\ a_2x+b_2y=c_2 \end{cases}$，其克莱姆法则的解为

$$x=\frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}},\ y=\frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_1 & b_2 \end{vmatrix}}$$

这里为解释其几何意义，同样从行列式的列或行向量的角度入手。2 阶列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$、$\boldsymbol{x}$ 分别表示为

$$\boldsymbol{a}=\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \boldsymbol{b}=\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \boldsymbol{c}=\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \boldsymbol{x}=\begin{pmatrix} x \\ y \end{pmatrix}$$

上述二阶线性方程组使用向量的形式表示为 $(\boldsymbol{a},\boldsymbol{b})\boldsymbol{x}=\boldsymbol{c}$。下面推导出二阶克莱姆法则。

对于二阶线性方程组的系数行列式 $|\boldsymbol{a},\boldsymbol{b}|$，我们构造出：

$$x|\boldsymbol{a},\boldsymbol{b}|=|x\boldsymbol{a},\boldsymbol{b}|=|x\boldsymbol{a}+y\boldsymbol{b},\boldsymbol{b}|=|\boldsymbol{c},\boldsymbol{b}|$$

因此
$$x=\frac{|\boldsymbol{c},\boldsymbol{b}|}{|\boldsymbol{a},\boldsymbol{b}|}$$

据此构造的推理过程，绘出构造过程中的几何图形，如图 3-26 所示。

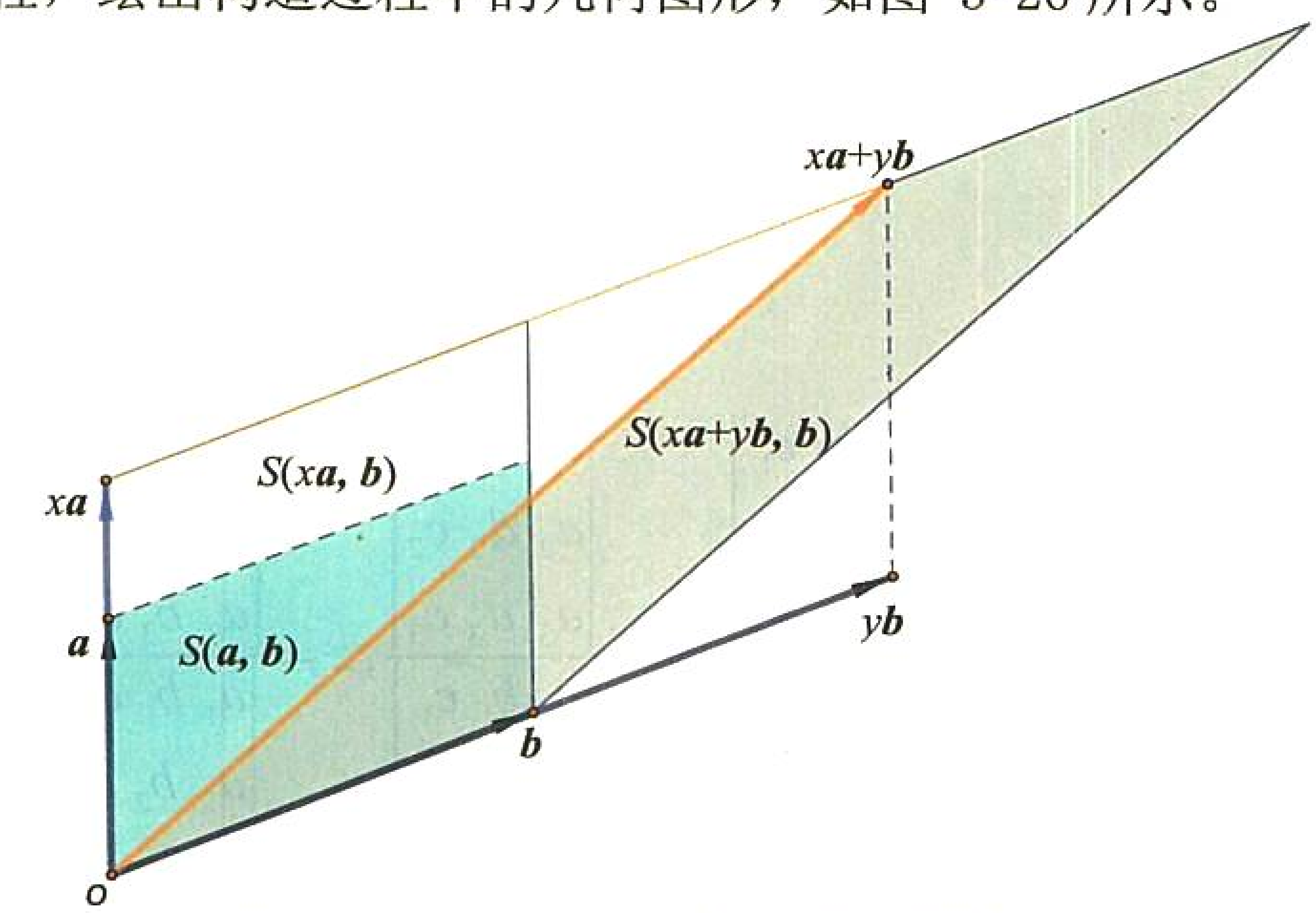


图 3-26 二阶克莱姆法则的几何解释

在图 3-26 中，向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的叉积是平行四边形，其面积是 $S(\boldsymbol{a},\ \boldsymbol{b})=|\boldsymbol{a},\ \boldsymbol{b}|$。这个面积

$S(\boldsymbol{a},\ \boldsymbol{b})$乘以一个数值 x（图里 x 大于 1），等同于向量 $\boldsymbol{b}$ 不变，同时与向量 $\boldsymbol{a}$ 伸长得到的新向量 $x\boldsymbol{a}$ 叉积张成大平行四边形。结果是平行四边形 $S(\boldsymbol{a},\ \boldsymbol{b})$ 向上伸展为大平行四边形 $S(x\boldsymbol{a},\ \boldsymbol{b})$。

接着，大平行四边形进行右向切变换，切变面积不变。这个切变等同于向量 $\boldsymbol{b}$ 不变，同时向量 $x\boldsymbol{a}$ 和 $y\boldsymbol{b}$ 相加得到新向量 $x\boldsymbol{a}+y\boldsymbol{b}$，向量 $\boldsymbol{b}$ 与新向量 $x\boldsymbol{a}+y\boldsymbol{b}$ 叉积张成细长形的平行四边形，结果是大平行四边形右向切变换为细长形的平行四边形，面积表达式为 $S(x\boldsymbol{a}+y\boldsymbol{b},\ \boldsymbol{b})$，即 $S(\boldsymbol{c},\ \boldsymbol{b})$。

因此各个平行四边形的面积有等式

$$xS(\boldsymbol{a},\ \boldsymbol{b})=S(x\boldsymbol{a},\ \boldsymbol{b})=S(x\boldsymbol{a}+y\boldsymbol{b},\ \boldsymbol{b})=S(\boldsymbol{c},\ \boldsymbol{b})$$

从而 x 的表达式为

$$x=\frac{S(x\boldsymbol{a},\ \boldsymbol{b})}{S(\boldsymbol{a},\ \boldsymbol{b})}=\frac{S(x\boldsymbol{a}+y\boldsymbol{b},\ \boldsymbol{b})}{S(\boldsymbol{a},\ \boldsymbol{b})}=\frac{S(\boldsymbol{c},\ \boldsymbol{b})}{S(\boldsymbol{a},\ \boldsymbol{b})}$$

把面积用行列式表示，即可得克莱姆法则的表达式

$$x=\frac{S(\boldsymbol{c},\ \boldsymbol{b})}{S(\boldsymbol{a},\ \boldsymbol{b})}=\frac{|\boldsymbol{c},\ \boldsymbol{b}|}{|\boldsymbol{a},\ \boldsymbol{b}|}$$

因此我们可以归纳出：

x 为由面积 $S(a,\ b)$ 伸缩和切变到 $S(c,\ b)$ 的面积之比例，是变化前后的面积之比。

同理，我们构造出 $y|\boldsymbol{a},\boldsymbol{b}|=|\boldsymbol{a},y\boldsymbol{b}|=|\boldsymbol{a},x\boldsymbol{a}+y\boldsymbol{b}|=|\boldsymbol{a},\boldsymbol{c}|$，得到 $y=\dfrac{|\boldsymbol{a},\boldsymbol{c}|}{|\boldsymbol{a},\boldsymbol{b}|}$。它的几何图形类同于 $x=\dfrac{|\boldsymbol{c},\boldsymbol{b}|}{|\boldsymbol{a},\boldsymbol{b}|}$，咱不再画了。

3.7.2　三阶克莱姆法则的几何解释

三阶线性方程组如下：

$$\begin{cases}a_1x+b_1y+c_1z=d_1\\a_2x+b_2y+c_2z=d_2\\a_3x+b_3y+c_3z=d_3\end{cases}$$

其克莱姆法则的解为

$$x=\frac{\begin{vmatrix}d_1&b_1&c_1\\d_2&b_2&c_2\\d_3&b_3&c_3\end{vmatrix}}{\begin{vmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{vmatrix}},y=\frac{\begin{vmatrix}a_1&d_1&c_1\\a_2&d_2&c_2\\a_3&d_3&c_3\end{vmatrix}}{\begin{vmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{vmatrix}},z=\frac{\begin{vmatrix}a_1&b_1&d_1\\a_2&b_2&d_2\\a_3&b_3&d_3\end{vmatrix}}{\begin{vmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{vmatrix}}$$

同样地，三阶列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$、$\boldsymbol{d}$、$\boldsymbol{x}$ 分别表示为

$$\boldsymbol{a}=\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix},\boldsymbol{b}=\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix},\boldsymbol{c}=\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix},\boldsymbol{d}=\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix},\boldsymbol{x}=\begin{pmatrix}x\\y\\z\end{pmatrix}$$

因此三阶线性方程组表示为$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\boldsymbol{x}=\boldsymbol{d}$。与二阶克莱姆法则的推导类似，下面推导三阶克莱姆法则。

对于三阶线性方程组的系数行列式$|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|$，我们构造出：

$$x|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|=|x\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|=|x\boldsymbol{a}+y\boldsymbol{b},\boldsymbol{b},\boldsymbol{c}|=|x\boldsymbol{a}+y\boldsymbol{b}+z\boldsymbol{c},\boldsymbol{b},\boldsymbol{c}|=|(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})\boldsymbol{x},\boldsymbol{b},\boldsymbol{c}|=|\boldsymbol{d},\boldsymbol{b},\boldsymbol{c}|$$

因此
$$x=\frac{|\boldsymbol{d},\boldsymbol{b},\boldsymbol{c}|}{|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|}$$

类似地，我们可以得到$y=\dfrac{|\boldsymbol{a},\boldsymbol{d},\boldsymbol{c}|}{|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|}$，$z=\dfrac{|\boldsymbol{a},\boldsymbol{b},\boldsymbol{d}|}{|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|}$。据上述推理过程，绘出$x=\dfrac{|\boldsymbol{d},\boldsymbol{b},\boldsymbol{c}|}{|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|}$的几何图形如图 3-27 所示。

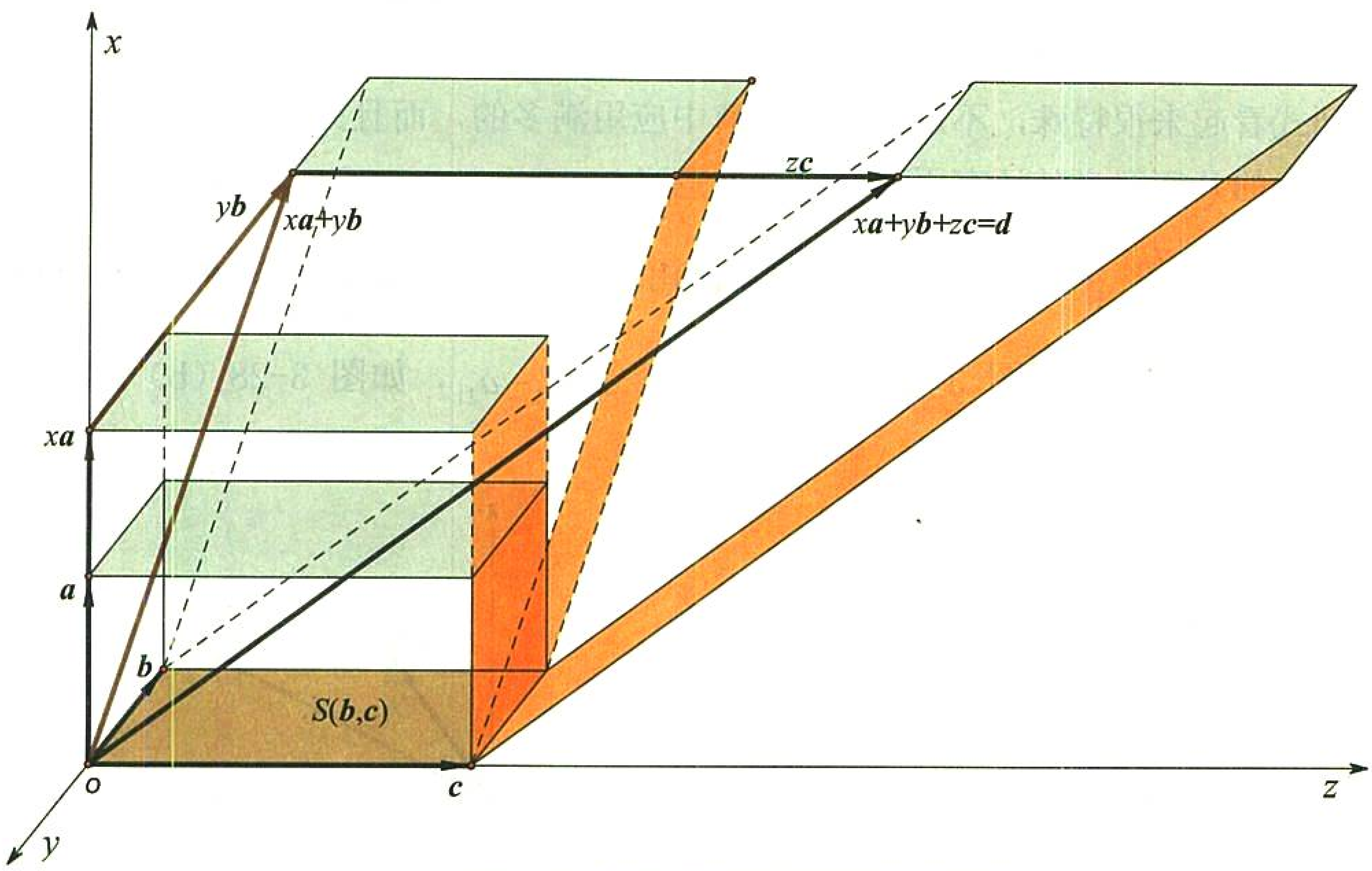


图 3-27 三阶克莱姆法则的几何解释

由图 3-27 看到，由向量$\boldsymbol{d}$、$\boldsymbol{b}$、$\boldsymbol{c}$张成的平行六面体的体积$|\boldsymbol{d},\boldsymbol{b},\boldsymbol{c}|$除以由向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$张成的平行六面体的体积$|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|$就是向量$\boldsymbol{a}$的伸缩量$x$值。因为我们看到，在图所示的变换中，以$\boldsymbol{b}$、$\boldsymbol{c}$两向量张成的平行四边形面积$S(\boldsymbol{b},\boldsymbol{c})$为底的平行六面体的变化过程中，只有向量$\boldsymbol{a}$进行了伸缩变化，其余的平行六面体的变化只是先后沿着向量$\boldsymbol{b}$和$\boldsymbol{c}$的方向进行了等体积的切向变化，因此向量$\boldsymbol{a}$方向的高度没有变化。

克莱姆法则的意义是可以用方程组的系数和常数项的行列式把方程组的解简洁地表达出来。但在实际工程应用中由于计算量较大，常常采用高斯消元法来解大型的线性方程组。在后面的线性方程组一章（详见第 6 章）中，我们将探讨高斯消元法的几何意义。

在以上的行列式的各类性质的几何解释中，除了伸缩、旋转、镜像就是切变，向量的大小

和方向在变化，但没有对向量进行弯曲、扭曲变化，全部是直线段的变化。所有的变化保持直线性。因此行列式的计算中包含线性变换的内涵。 这个意思到了矩阵一章就比较清楚了。

3.8 一类行列式的几何意义

3.8.1 最后一列为 1 的行列式

某一类行列式的形式如下：

$$\begin{vmatrix} a_{11} & a_{12}\cdots a_{1n-1} & 1 \\ a_{21} & a_{22}\cdots a_{2n-1} & 1 \\ \vdots & \vdots \quad\quad \vdots & \vdots \\ a_{n1} & a_{n2}\cdots a_{nn-1} & 1 \end{vmatrix}$$

这类行列式有个特点，就是行列式的最后一列全部是数字 1。在前面章节中我们给出了一般行列式的几何意义，比如三阶行列式是由三个三维向量所张成的平行多面体的有向体积，那么对于最后一列的元素全部为 1 的三阶行列式至少也有同样的几何意义。

这类行列式看起来很特殊，不过在解析几何中应用满多的，而且此类行列式除了前面所讲述的一般情况下的几何意义外还有更简洁的几何意义。

对于二阶行列式$\begin{vmatrix} a_{11} & 1 \\ a_{21} & 1 \end{vmatrix}$，其一般的几何意义是由二维向量$\boldsymbol{a}=(a_{11},1)$和 $\boldsymbol{b}=(a_{21},1)$所张成的平行四边形的有向面积（见图 3-28（a）），其值等于 $a_{11}-a_{21}$，如图 3-28（b）所示。显然，两个向量的末端都在一条直线上。

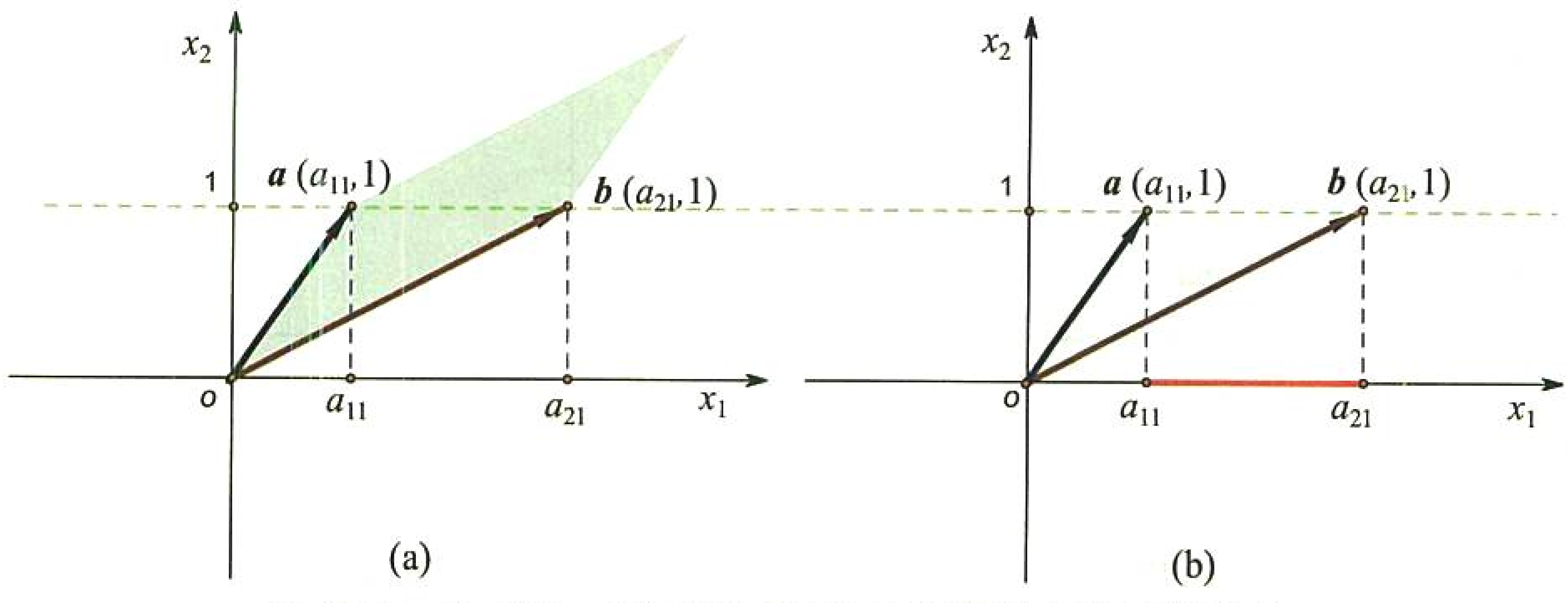


图 3-28 含“1”二阶行列式的有向面积等于有向线段长

另外，从行列式的值的表达式$a_{11}-a_{21}$，直接得到另外一个几何意义，就是坐标 x 轴上的有向线段 $\overrightarrow{a_{21}a_{11}}$ 的有向长度。显然，如果此行列式为零，那么向量$\boldsymbol{a}$、$\boldsymbol{b}$重合；在解析几何上可以判断点 a、b 重合。

对于三阶行列式$\begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 1 \end{vmatrix}$，其一般的几何意义是由三个三维行向量$(a_{11}, a_{12}, 1)$、$(a_{21}, a_{22}, 1)$、$(a_{31}, a_{32}, 1)$所张成的平行六面体的有向体积，如图 3-29 所示。

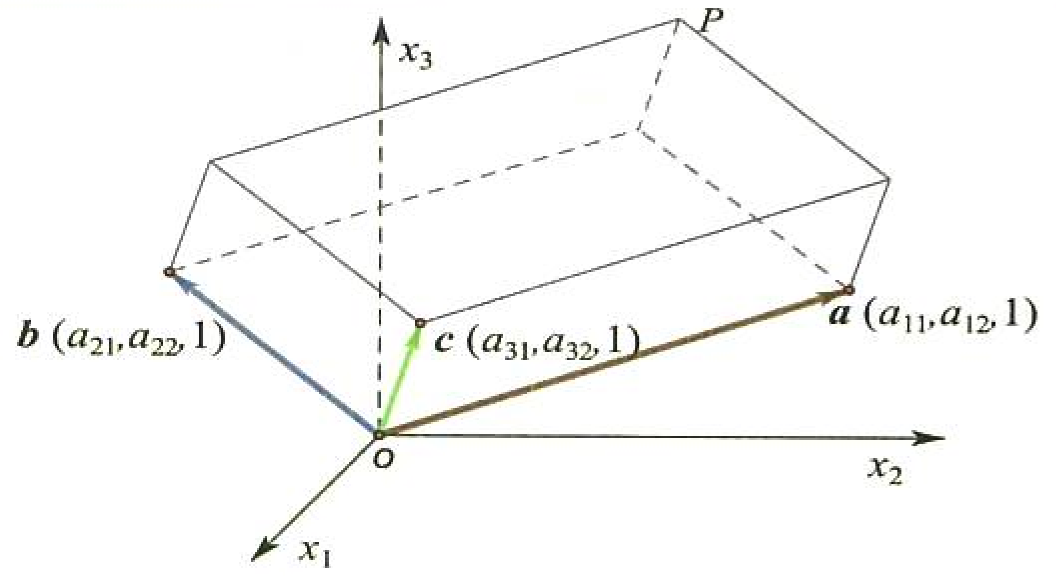


图 3-29　含“1”列的三阶行列式等于平行六面体的有向体积

显然也可以看出，三个向量的末端都在一个平面 $x_3=1$ 上，这一点下面马上提到。

对于三阶行列式，我们可以利用行列式的性质对其进行变换，第 2 行和第 3 行元素分别减去第 1 行元素，进而按照第 3 列进行展开，得到一个二阶行列式：

$$\begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 1 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21}-a_{11} & a_{22}-a_{12} & 0 \\ a_{31}-a_{11} & a_{32}-a_{12} & 0 \end{vmatrix} = \begin{vmatrix} a_{21}-a_{11} & a_{22}-a_{12} \\ a_{31}-a_{11} & a_{32}-a_{12} \end{vmatrix}$$

二阶行列式 $\begin{vmatrix} a_{21}-a_{11} & a_{22}-a_{12} \\ a_{31}-a_{11} & a_{32}-a_{12} \end{vmatrix}$ 的向量张成的面实际上是在二维平面 $x_3=1$ 上，即由行向量 $\overrightarrow{A_1A_2}=(a_{21}\text{-}a_{11},\ a_{22}\text{-}a_{12})$ 和 $\overrightarrow{A_1A_3}=(a_{31}\text{-}a_{11},\ a_{32}\text{-}a_{12})$ 所张成的平行四边形的有向面积，如图 3-30 所示的阴影平行四边形。

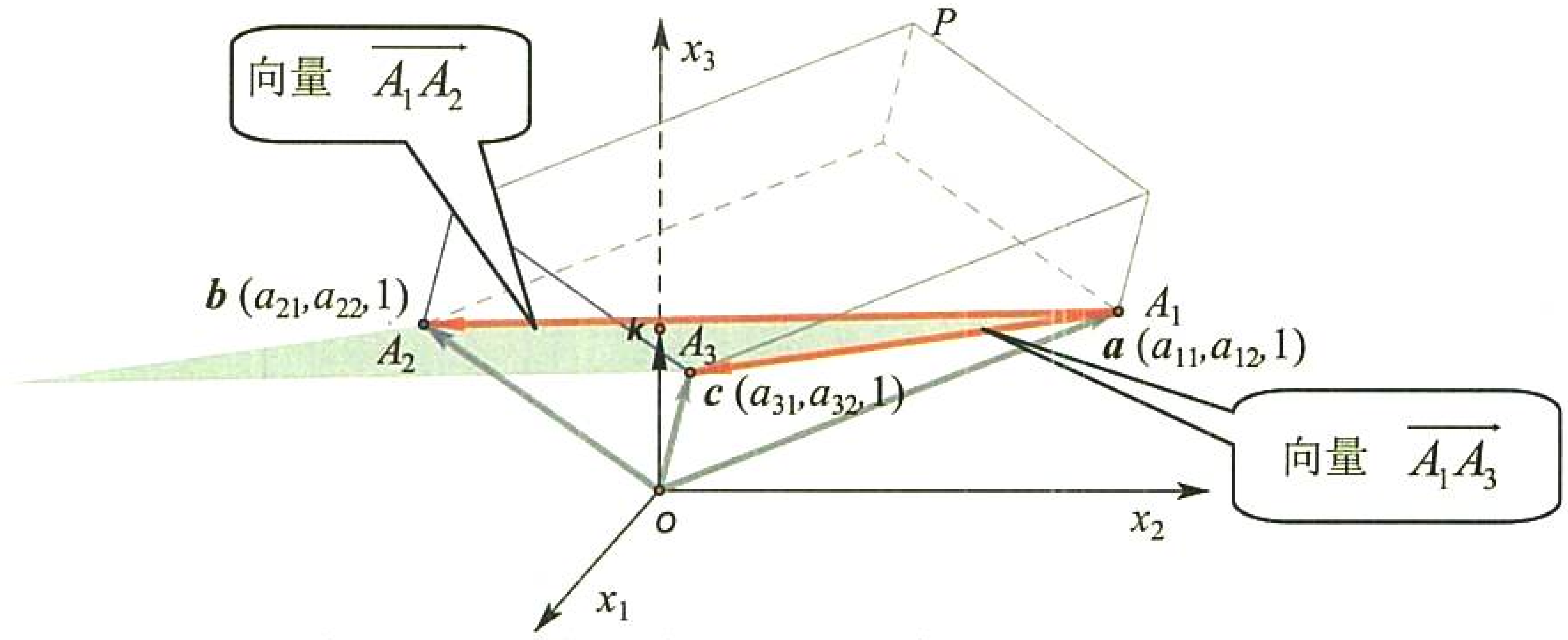


图 3-30 含“1”三阶行列式的有向体积等于有向面积

稍进一步，更具有实用意义的是此行列式还等于在二维平面 $x_3=1$ 上以三点 $A_1(a_{11},a_{12})$、$A_2(a_{21},a_{22})$ 和 $A_3(a_{31},a_{32})$ 为顶点的三角形 $\Delta A_1A_2A_3$ 面积的 2 倍，见图 3-31 中的阴影三角形。

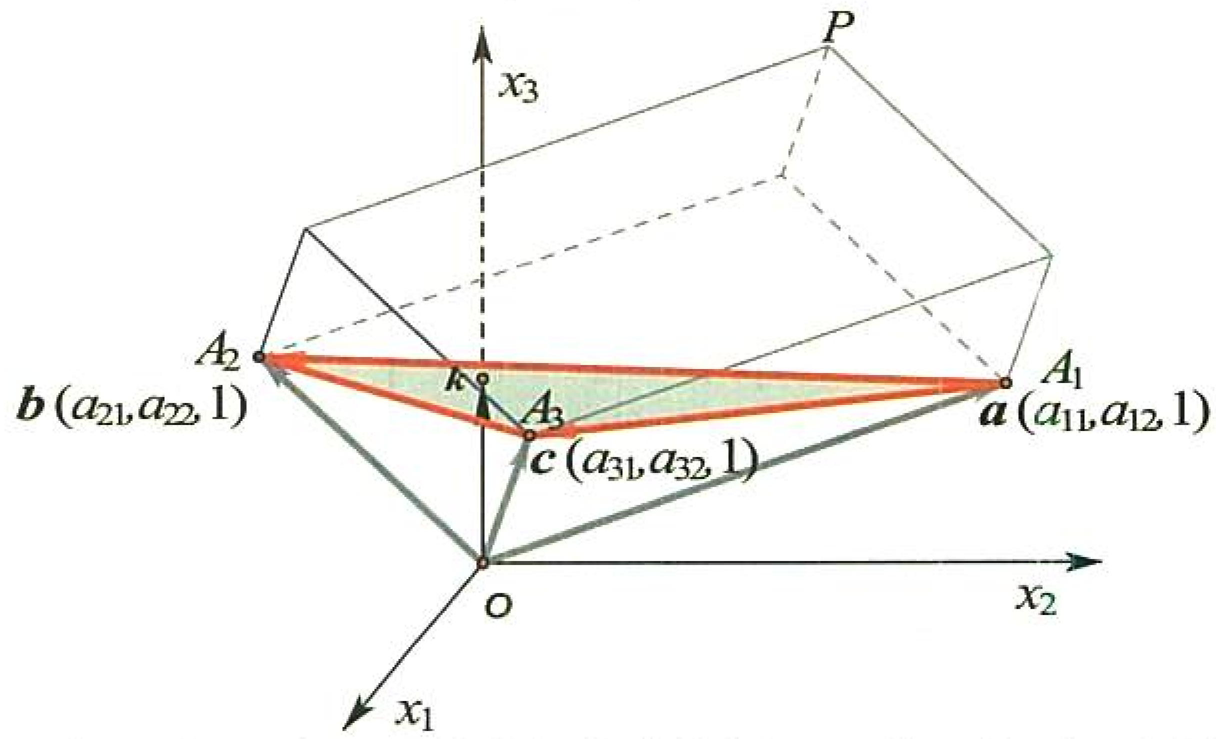


图 3-31 含“1”三阶行列式的有向体积等于顶点三角形面积的 2 倍

有个推论就是，如果行列式等于零，那么六面体的体积等于零，三角形的面积$\Delta A_1A_2A_3$也等于零，说明三个顶点A_1、A_2、A_3在一条直线上。

类似地，对于四阶行列式$\begin{vmatrix} a_{11} & a_{12} & a_{13} & 1 \\ a_{21} & a_{22} & a_{23} & 1 \\ a_{31} & a_{32} & a_{33} & 1 \\ a_{41} & a_{42} & a_{43} & 1 \end{vmatrix}$，其一般的几何意义是由四个四维行向量所张成的平行八面体的有向体积；同时也是三维空间$x_4=1$中以四个点$A_1(a_{11},a_{12},a_{13})$、$A_2(a_{21},a_{22},a_{23})$、$A_3(a_{31},a_{32},a_{33})$和$A_4(a_{41},a_{42},a_{43})$为顶点的四面体的体积的6倍；同时也是以向量$\overrightarrow{A_1A_2}$、$\overrightarrow{A_1A_3}$和$\overrightarrow{A_1A_4}$张成（为相邻棱）的平行六面体的有向体积。

3.8.2 一列为1的行列式的应用

在解析几何中我们常常应用等式$D_n=\begin{vmatrix} a_{11} & a_{12} \cdots a_{1n-1} & 1 \\ a_{21} & a_{22} \cdots a_{2n-1} & 1 \\ \vdots & \vdots \quad\quad \vdots & \vdots \\ a_{n1} & a_{n2} \cdots a_{nn-1} & 1 \end{vmatrix}=0$来判断空间中点之间的关系，比如，

$$D_n=0 \Rightarrow \begin{cases} n=2, & \text{一维数轴上两点重合} \\ n=3, & \text{二维平面上三点共线} \\ n=4, & \text{三维空间中四点共面} \end{cases}$$

这个结论是3.8.1一节的总结。下面我们看一看此类行列式在几何上的灵活应用。

例3.1 有行列式方程：

$$\begin{vmatrix} x & y & 1 \\ a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \end{vmatrix}=0$$

此行列式方程表示过平面两点$A(a_1,a_2)$和$B(b_1,b_2)$的直线，其中(x,y)为直线上的任意一点。将其推广到空间中，如例3.2所示。

例3.2

$$\begin{vmatrix} x & y & z & 1 \\ a_1 & a_2 & a_3 & 1 \\ b_1 & b_2 & b_3 & 1 \\ c_1 & c_2 & c_3 & 1 \end{vmatrix}=0$$

此行列式方程表示，不在同一条直线上的三点$A(a_1,a_2,a_3)$、$B(b_1,b_2,b_3)$、$C(c_1,c_2,c_3)$所确定的平面，其中(x,y,z)是平面上的任意一点。

这个行列式方程也可以这样得到：

从初等几何的知识知道，过不在同一条直线的三点A、B、C的平面Π是存在的。设平面

Π 的一般方程为 $Ax+By+Cz+D=0$，那么平面上的三点和任意一点 $M(x,y,z)$ 满足方程组：

$$\begin{cases} Ax+By+Cz+D=0 \\ Aa_1+Ba_2+Ca_3+D=0 \\ Ab_1+Bb_2+Cb_3+D=0 \\ Ac_1+Bc_2+Cc_3+D=0 \end{cases}$$

将这个方程看成是对于欲求平面系数 A、B、C、D 的齐次线性方程：

$$\begin{bmatrix} x & y & z & 1 \\ a_1 & a_2 & a_3 & 1 \\ b_1 & b_2 & b_3 & 1 \\ c_1 & c_2 & c_3 & 1 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \mathbf{0}$$

因为这个平面是存在的，所以方程组有非零解，根据第 6 章线性方程组的理论，四阶矩阵的行列式等于零，因此得到了题目中的结论。

第4章 向量组及向量空间的几何意义

我们知道，实数可看做一维向量（正负是实数的方向），复数可看做二维向量，那么可能还有“三维数”、“四维数”，乃至“n 维数”。不过当年哈密尔顿失望地发现，要形成有加减乘除四则运算的数系，只能是四元数，而且不得不放弃乘法的交换律。最后发现的八元数，连结合律也维持不了。除此之外，其他维数的向量根本无法定义四则运算，更谈不上构成数系。德国数学家格拉斯曼1844年引入了n维向量的概念。令人深思的是，n维向量既然不能成为有四则运算的数系，那么它的结构是什么呢？这是19世纪抽象代数思想的发展的自然思考。研究表明，n维向量全体，可以定义加法和减法，此外还有单个的“数”可以和向量相乘。这就是向量空间(线性空间)的来源。此外，两个向量可以有“内积”和“外积”，但是它们都没有逆运算，即没有除法。这是一个不同于“数系”的崭新的数学结构。果然，在向量空间的舞台上，产生了具有深远影响的数学成就。

多个向量组成向量组，向量组关键的概念有线性相关性、秩等。有了向量组，就会产生向量空间的概念，因为一个有限元素的向量组会张成一个无限元素的向量空间。

在向量组张成的向量空间里，基、维数、坐标等概念完整地勾画出了一个线性空间的全貌。如果再加上线性映射、线性变换和基变换等概念，那么向量空间就是向量活动的一个社会了。

线性变换与向量空间是相辅相成、互相依存的。向量空间是线性变换的载体，没有向量空间，线性变换无用武之地，没有归属感；反之，向量空间没有线性变换作用其上，向量空间就是死的，空洞无物。这个关系有点像运动与物质空间关系的味道。如果大家建立了一个空间的概念，作为运动概念的线性映射和线性变换就有了家园。

本章主要是从几何图形上弄清向量空间里的这些概念，让抽象的概念回归形象的几何解释。

线性空间里需要建立坐标系才能数量化，坐标系就是基，基就是由几个骨干向量构成的一个有序向量组。因此，要讨论向量空间，先讨论向量组是正点。

（注意，本章循着向量传统的写法，用黑体拉丁文$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$来表示，这与本书其他章节向量黑体小字母$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$的写法没有任何区别，只是传统或方便的选择。）

4.1 向量组的几何意义

向量组是对有限个向量集合的研究。对向量组的性质了解清楚后，就会对矩阵和线性方程组的性质认识更加深入。因为矩阵实际上也是一个有序向量组。线性方程组实际上就是向量组的线性表示。因此，在开展矩阵及方程组的研究之前，有必要先研究清楚向量组的问题。

向量组里的向量们有哪些特性呢？有什么共性？有没有不变量？

有，这个不变的特性就是一个叫“秩”的东东。下面我们来慢慢探究之。

4.1.1　向量线性表示/组合的几何意义

1. 线性表示/组合的定义

先看图 4-1 中平面向量集合，这个集合有七个二维向量，用坐标表示出来就是

$\alpha_1=(2,1)$，$\alpha_2=(3,3)$，$\alpha_3=(1,2)$，$\alpha_4=(-1,1)$，$\alpha_5=(-2,2)$，$\alpha_6=(-3,-1)$，$\alpha_7=(2,-2)$

仔细观察后发现：

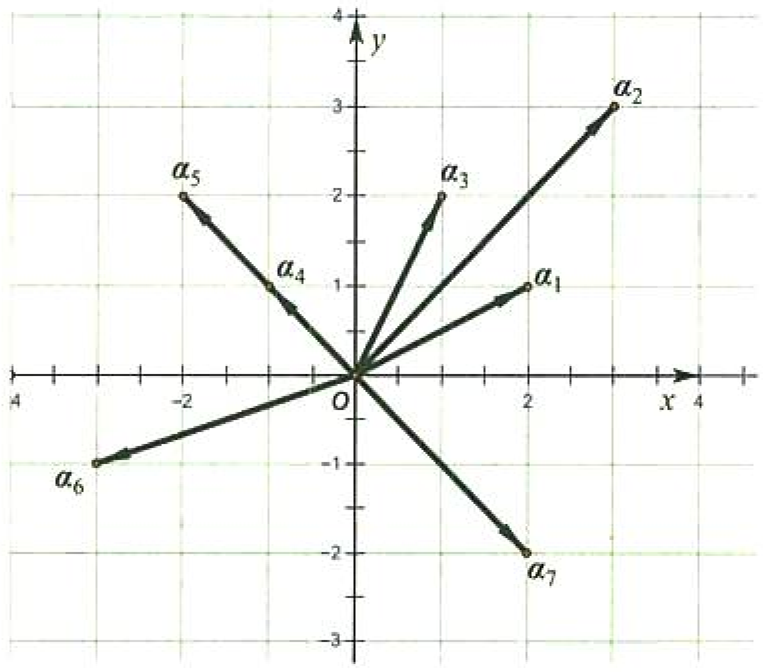


图 4-1 平面向量的例子

向量$\boldsymbol{\alpha}_2$可以由两个向量$\boldsymbol{\alpha}_1$和$\boldsymbol{\alpha}_3$相加得到，即$\boldsymbol{\alpha}_2=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_3$；

向量$\boldsymbol{\alpha}_4$可以由$\boldsymbol{\alpha}_5$乘以$\dfrac{1}{2}$得到，也可以由$\boldsymbol{\alpha}_7$乘以$-\dfrac{1}{2}$得到，即$\boldsymbol{\alpha}_4=\dfrac{1}{2}\boldsymbol{\alpha}_5$或$\boldsymbol{\alpha}_4=-\dfrac{1}{2}\boldsymbol{\alpha}_7$；

向量$\boldsymbol{\alpha}_6$可以由$\boldsymbol{\alpha}_2$的数乘和$\boldsymbol{\alpha}_7$的数乘之和得到，即$\boldsymbol{\alpha}_6=-\dfrac{2}{3}\boldsymbol{\alpha}_2-\dfrac{1}{2}\boldsymbol{\alpha}_7$；

……

上面的例子是说，一个向量可以由另外一个或几个向量（向量组）用数乘之和的形式表示出来，一般表达式就是$\boldsymbol{\beta}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_s\boldsymbol{\alpha}_s$，其中$x_1,x_2,\cdots,x_s$是常数。这里称之为向量$\boldsymbol{\beta}$可以由向量组$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s\}$线性表示。或者讲，向量$\boldsymbol{\beta}$是向量$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s$的线性组合。

我们研究一下线性表示的表达式$\boldsymbol{\beta}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_s\boldsymbol{\alpha}_s$，明显地：向量$\boldsymbol{\alpha}_1$数乘$x_1$，向量$\boldsymbol{\alpha}_2$数乘$x_2$，…，然后把数乘后的向量相加起来就得到了一个新的向量$\boldsymbol{\beta}$。反过来讲，一个向量$\boldsymbol{\beta}$被分解为几个向量的倍数。所以我们得到一个结论：

线性组合或表示式的代数意义是向量数乘和加法的综合。

在第 2 章的向量介绍中，我们知道，数乘的几何解释就是在原向量的直线上向量长度的伸长或缩短；两向量相加的几何解释就是依照平行四边形法则对向量进行合并。因此向量的线性组合的几何意义就是对向量组内的向量长度进行缩放后依照平行四边形法则进行合并相加，如图 4-2 中的$\boldsymbol{\alpha}_1$和$\boldsymbol{\alpha}_3$合并成$\boldsymbol{\alpha}_2$；线性表示的几何意义就是可以把一个向量依照平行四边形法则分解（或投影）为向量组上的和，如图 4-2 中的$\boldsymbol{\alpha}_6$分解为$\boldsymbol{\alpha}_2$与$\boldsymbol{\alpha}_7$上的分量。

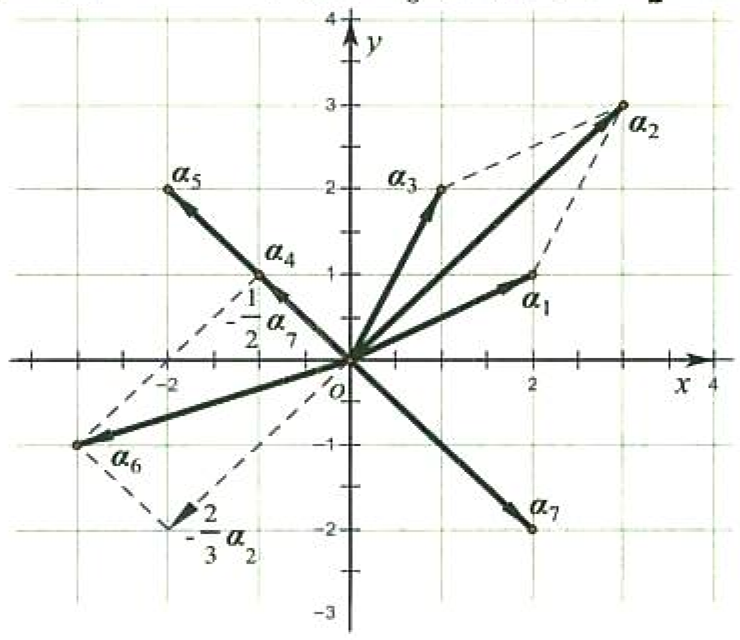


图 4-2 向量之间的线性组合及线性表示

2. 向量被向量组所线性表示的几何意义

在图 4-2 中，我们已经知道：α_4 可以被向量组 $\{\alpha_5\}$ 线性表示，也可以被 $\{\alpha_7\}$ 线性表示；α_2 可以被向量组 $\{\alpha_1,\alpha_3\}$ 线性表示，α_6 可以被向量组 $\{\alpha_2,\alpha_7\}$ 线性表示。

下面我们只研究 α_4 的线性表示方法。

α_4 可以由多种向量组线性表示，只要这个向量组所确定的直线、平面或空间包含 α_4。

比如 α_4 可以被 $\{\alpha_5\}$，$\{\alpha_5,\alpha_7\}$ 等向量组线性表示。因为 $\{\alpha_5\}$ 所能扩张的是一根直线（如图 4-3 所示的虚直线），这根直线包含了 α_4。同理，$\{\alpha_5,\alpha_7\}$ 扩张的是同一根直线。

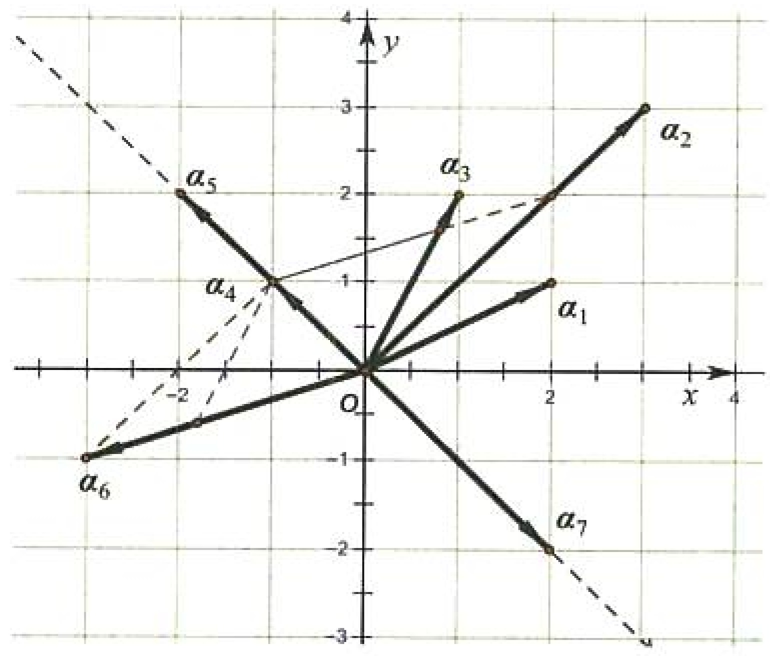


图 4-3 平面上，一个向量可以被不同的向量组线性表示

再比如 α_4 可以被向量组 $\{\alpha_3,\alpha_6\}$ 线性表示，同时也可以被向量组 $\{\alpha_2,\alpha_6\}$，$\{\alpha_1,\alpha_2,\alpha_6\}$ 等线性表示。因为，在图 4-3 的平面上，任何两个不在一条直线上的向量都可以扩张成这个坐标平面，α_4 当然包含在这个平面上，因此被这个不同方向的向量所组成的向量组线性表示。

换句话说，α_4 可以被线性表示是因为 α_4 可以分解到这两个不在一条直线上（或不同方向）的向量上。这个结论容易验证，因为我们容易把被分解向量作为平行四边形对角线，把另两个向量（或其延长线）作为邻边而构造出一个平行四边形。

我们把上述的平面二维向量扩充为空间三维向量（见图 4-4）来继续讨论。

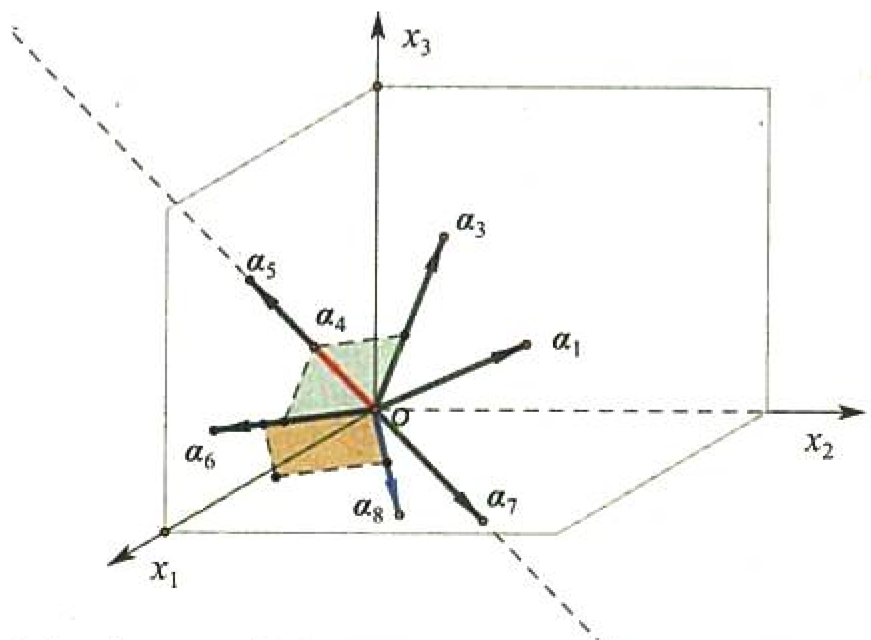


图 4-4 空间上，一个向量可以被不同的向量组线性表示

图 4-4 中 $\alpha_1 \sim \alpha_7$ 仍在同一平面，新增加的向量 α_8 与它们不在同一平面。

我们说，α_4 不能被 $\{\alpha_6,\alpha_8\}$ 线性表示，因为 α_4 不在 α_6 与 α_8 所构成的平面上。但可以说 α_4 能被 $\{\alpha_1,\alpha_6,\alpha_8\}$ 线性表示，因为 $\{\alpha_1,\alpha_6,\alpha_8\}$ 所张成的三维空间包含了 α_4。

其实我们如果引入了极大无关组（见 4.1.4 节）的概念，则可以说得更简单而清晰：向量可以被向量组线性表示的原因是，这个向量处于以向量组的极大无关组为基的线性子空间里面。向量在向量组的子空间里面就可以被表示，不在里面就不能表示——就这么简单！

4.1.2　向量组线性相关的几何意义

如果一个向量可以由一个向量组线性表示，我们就称这个向量和向量组线性相关。另外的说法就是，一个向量组里只要有一个向量可以由组内其他向量线性表示，我们就称这个向量组线性相关。反之，如果向量组里的任意一个向量都不能由其他向量线性表示，我们就称向量组线性无关。

图 4-2 中线性相关的平面向量组例举如下：

$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$ 线性相关，因为 $\boldsymbol{\alpha}_2=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_3$；

$\{\boldsymbol{\alpha}_4,\boldsymbol{\alpha}_5\}$，$\{\boldsymbol{\alpha}_4,\boldsymbol{\alpha}_7\}$，$\{\boldsymbol{\alpha}_5,\boldsymbol{\alpha}_7\}$ 都线性相关，因为 $\boldsymbol{\alpha}_5=2\boldsymbol{\alpha}_4=-2\boldsymbol{\alpha}_7$；

$\{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_6,\boldsymbol{\alpha}_7\}$ 线性相关，因为 $\boldsymbol{\alpha}_6=-\dfrac{2}{3}\boldsymbol{\alpha}_2-\dfrac{1}{2}\boldsymbol{\alpha}_7$；

……

仔细琢磨，我们就会发现一些规律，就是在二维平面上：

- 如果两个向量线性相关，那么这两个向量必然在一条直线上（估计这也是“线性相关”术语的由来），两个向量的方向或者相同或者相反；反之也成立，例如向量 $\boldsymbol{\alpha}_4$、$\boldsymbol{\alpha}_5$、$\boldsymbol{\alpha}_7$ 在一条直线上，因此两两相关。

解释：因为在一条直线上的所有向量中的两个向量都具有 x 倍数的关系，即 $\boldsymbol{\beta}=x\boldsymbol{\alpha}$，所有向量都线性相关。

- 不在一条直线上的任意两个向量一定线性无关，如向量 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、$\boldsymbol{\alpha}_3$、$\boldsymbol{\alpha}_4$（或 $\boldsymbol{\alpha}_5$ 或 $\boldsymbol{\alpha}_7$）、$\boldsymbol{\alpha}_6$ 两两线性无关。

解释：因为一个向量必然不能被另外一个向量线性表示。如果一个向量能够被另外一个向量线性表示，即 $\boldsymbol{\beta}=x\boldsymbol{\alpha}$，那么这两个向量就是 x 倍数的关系，就会在一条直线上。这与命题的前提矛盾。

- 在二维平面空间上，任意三个向量必然相关，如 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$，$\{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_6,\boldsymbol{\alpha}_7\}$ 等。当然，三个以上的向量也必然线性相关。

解释：我们看看任意一个三向量 $\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$，如果其中任意两个如 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 线性相关，不用说这个三向量的向量组也线性相关；下一步我们设 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 线性无关，看看为什么这个三向量的向量组就线性相关了呢？如图 4-5 所示，我们随意画出不在一条直线上的两个向量 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$，分别作出过此两个向量的直线（图中虚线所示）；然后过第三个向量 $\boldsymbol{\gamma}$ 的头部点分别作 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 的平行线（作投影），且交 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 直线的交点为（x_1,x_2）。至此，我们构造出了向量 $\boldsymbol{\gamma}$ 的分解式 $\boldsymbol{\gamma}=x_1\boldsymbol{\alpha}+x_2\boldsymbol{\beta}$。

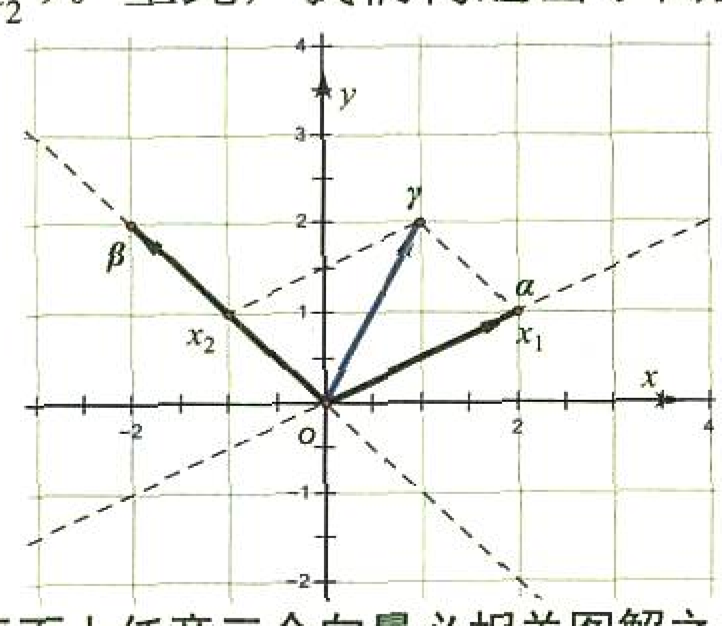
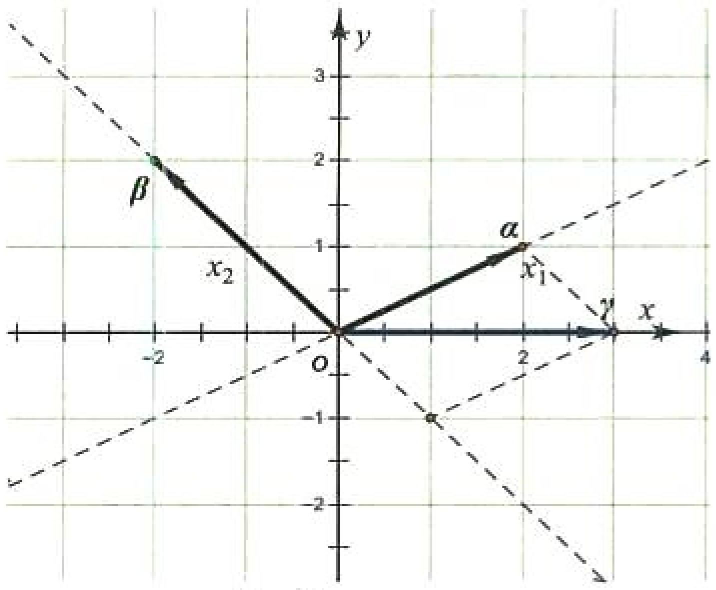


图 4-5 平面上任意三个向量必相关图解之一

改变向量 $\boldsymbol{\gamma}$ 的位置，如图 4-6 所示，我们可得到同样的表达式，只是 x_1、x_2 的值不同罢了。注意这里 x_1、x_2 的值分别是投影分量和 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 的比值，或称为坐标值，这个 4.2.3 节有详细讲解。



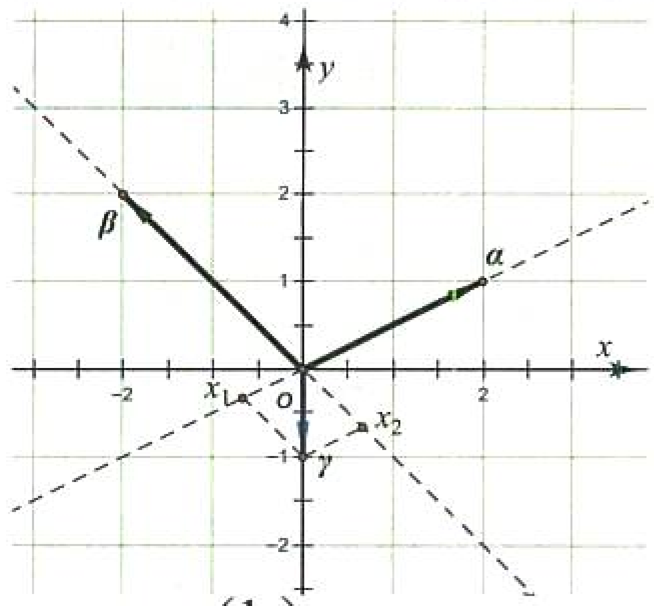


图 4-6 平面上任意三个向量必相关图解之二

这个分解式就是线性表示式。因此$\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$线性相关（再强调一下，此命题在二维向量空间中成立）。

对于通常的三维空间的向量，线性相关和线性无关也有直观的几何意义。先考察三维空间中两个向量的线性关系。由定义，它们线性相关是说其中一个是另一个的线性组合，即只与另一个向量相差一个常数因子。这就是说，这两个向量落在同一条直线上，方向相同或者方向相反。反之，如果两个向量落在空间里同一条直线上，那么它们就线性相关。两个向量线性无关则说这两个向量不平行，不能平移地搬到同一条直线上去。

现在来考察三维空间中三个向量$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$的线性关系（如图 4-7 所示）。

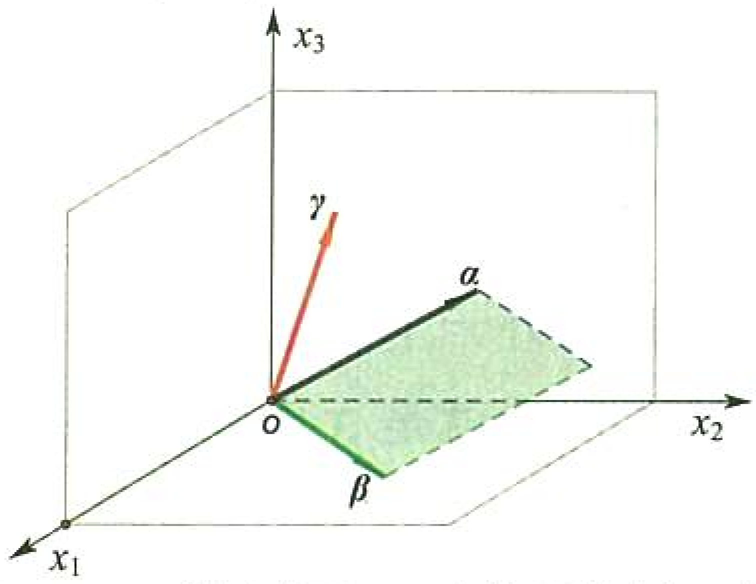


图 4-7 三维空间里三个向量线性无关

设向量$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$线性无关。那么就有$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$两两线性无关，任意假设$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$线性无关，则说明$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$不在同一条直线上，这两个向量所在的两根相交直线构成了一个平面（或者讲$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$张成的平面），这个平面上的所有向量是$x_1\boldsymbol{\alpha}+x_2\boldsymbol{\beta}$（$x_1$、$x_2$为任意实数）。

既然$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$线性无关，那么就没有$x_1\boldsymbol{\alpha}+x_2\boldsymbol{\beta}=\boldsymbol{\gamma}$的可能。也就是说向量$\boldsymbol{\gamma}$必然不在这个平面上，而是在这个平面之外。

实际上，这三个向量中任意一个向量都不在其余两个向量所张成的平面内（见图 4-8），如果用这三个三维向量构成一个三阶行列式，那么必然张成一个平行六面体，同时行列式的值不等于零。

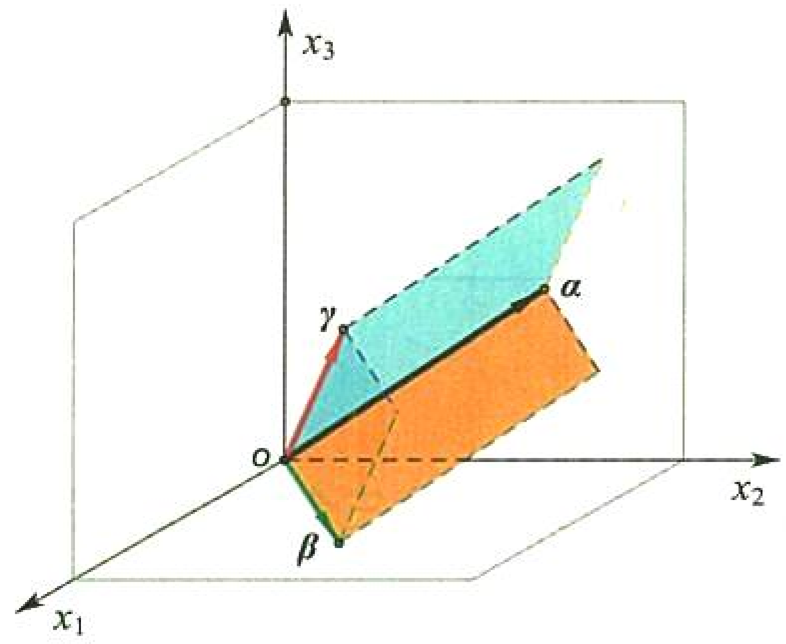


图 4-8 三个线性无关向量两两张成一个平面

如果设向量 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 线性相关，这里有两种情况：

一种情况是三个向量在一个平面上（三向量共面）。正如前面讨论三个向量线性无关的情况相反。任意地，如果 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 不相关，那么 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 会张成一个平面 $\{x_1\boldsymbol{\alpha}+x_2\boldsymbol{\beta}\}$；又因为 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 线性相关，则必有 $\boldsymbol{\gamma}$ 落在这个平面上满足 $x_1\boldsymbol{\alpha}+x_2\boldsymbol{\beta}=\boldsymbol{\gamma}$（$x_1$、$x_2$ 因 $\boldsymbol{\gamma}$ 的不同而不同），见图 4-9（a）。如果 $\boldsymbol{\gamma}$ 不落在这个平面上，会得到 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 线性无关的矛盾。

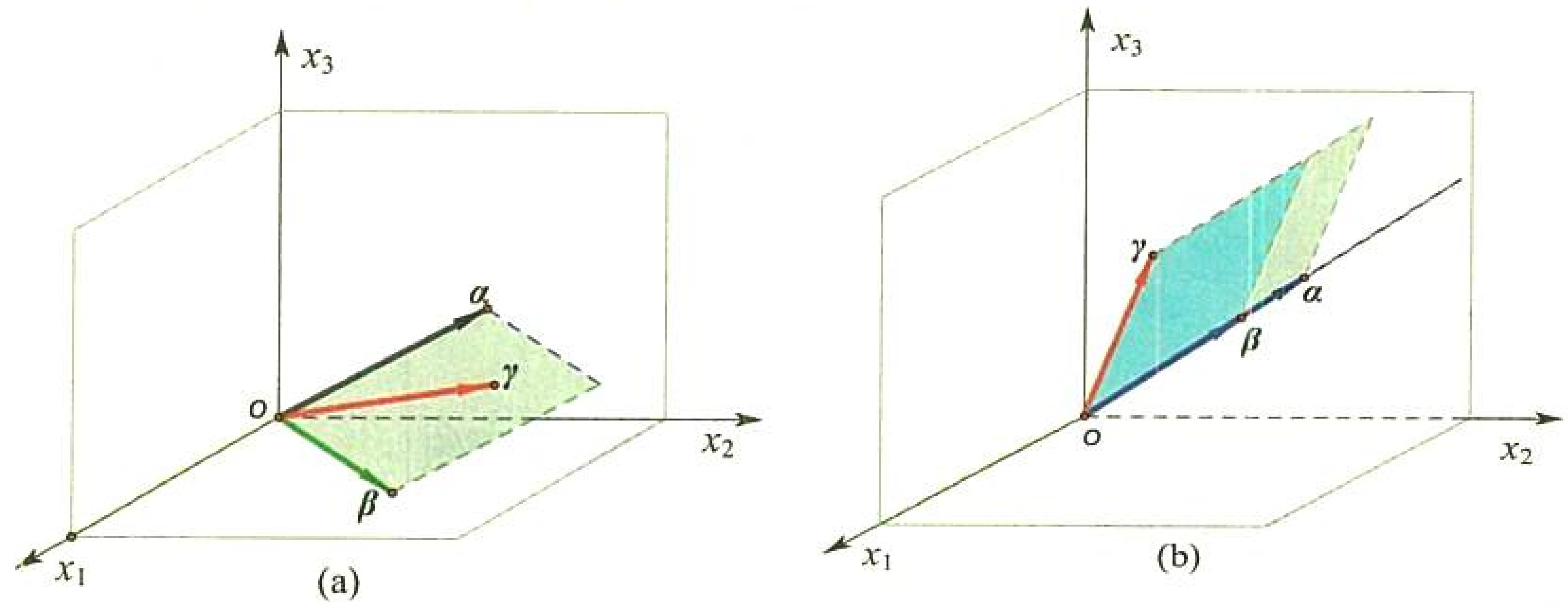


图 4-9 三维空间里三个向量线性相关图解之一

如果 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 线性相关，但第三个向量 $\boldsymbol{\gamma}$ 与 $\boldsymbol{\alpha}$（或 $\boldsymbol{\beta}$）线性无关，则也会有一个平面张出来，仍然会得到三个向量共面的情况，见图 4-9（b）。

另一种情况是三个向量在一条直线（共线）上。继续前面的讨论，如果 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 线性相关，第三个向量 $\boldsymbol{\gamma}$ 也与 $\boldsymbol{\alpha}$（或 $\boldsymbol{\beta}$）线性相关。那么三个向量会无意外地落在同一条直线上，如图 4-10 所示。

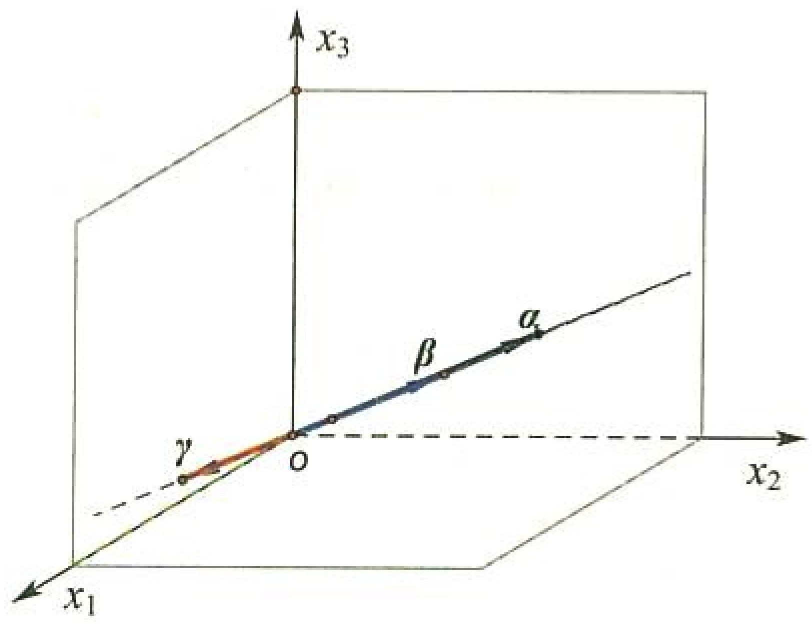


图 4-10 三维空间里三个向量线性相关图解之二

4.1.3　向量组等价的几何解释

两个向量组 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的等价就是这两个向量组能够互相被线性表示。详细地说，向量组 $\boldsymbol{A}$ 中的每一个向量都可以被向量组 $\boldsymbol{B}$ 线性表示；同样，向量组 $\boldsymbol{B}$ 中的每一个向量也可以被向量组 $\boldsymbol{A}$ 线性表示。或者说，如果把一个向量组中的任意一个向量拿出来放到另外一个向量组中，那么另外这个扩大的向量组就会线性相关，而且不论原向量组是否线性相关。

根据前面分析的向量组线性表示的几何意义，我们很容易理解向量组等价的几何意义：

两个向量组等价就是两个向量组所扩张成的直线、平面或空间相互重合。

下面讨论的是三维空间中向量组的等价关系。

1. 直线上的等价向量组

如图 4-11 所示的三维空间中，共有三条分离的不共面直线，每条直线上分别有两个、三

个和四个向量。两向量$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$在一条直线上；三向量$\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$、$\boldsymbol{\beta}_3$在另外一条直线上；四向量$\boldsymbol{\gamma}_1$、$\boldsymbol{\gamma}_2$、$\boldsymbol{\gamma}_3$、$\boldsymbol{\gamma}_4$在第三条直线上。

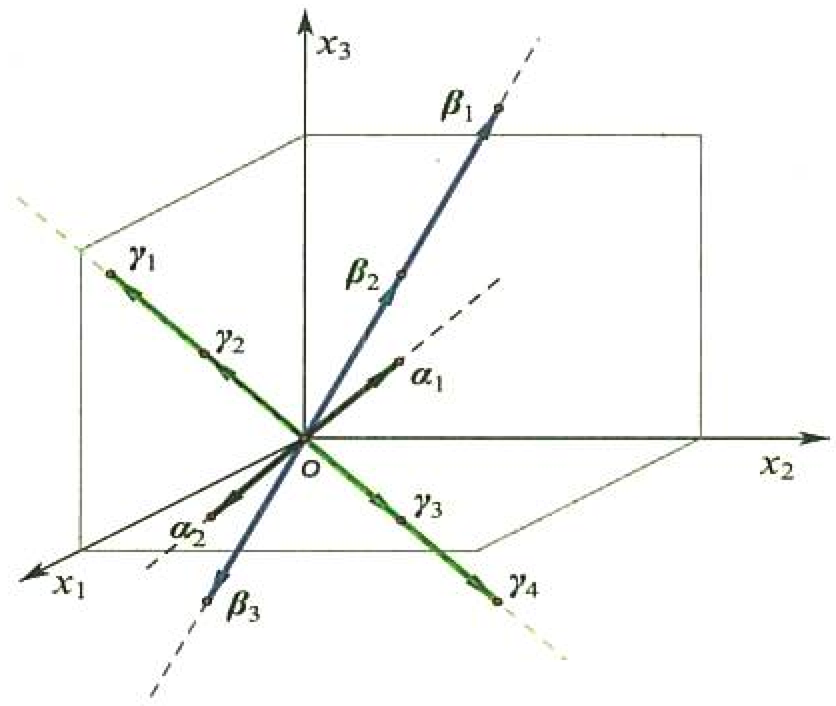


图 4-11 直线上的等价向量组

由此，我们可以验证以下命题：

- $\{\boldsymbol{\alpha}_1\}$，$\{\boldsymbol{\alpha}_2\}$，$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$是等价向量组；
- $\{\boldsymbol{\beta}_1\}$，$\{\boldsymbol{\beta}_2\}$，$\{\boldsymbol{\beta}_3\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_3\}$，$\{\boldsymbol{\beta}_2,\boldsymbol{\beta}_3\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3\}$是等价向量组；
- $\{\boldsymbol{\gamma}_1\}$，$\{\boldsymbol{\gamma}_2\}$，$\{\boldsymbol{\gamma}_3\}$，$\{\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_3\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3\}$，$\{\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_3,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_3,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3,\boldsymbol{\gamma}_4\}$，$\{\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3,\boldsymbol{\gamma}_4\}$是等价向量组。

其实上述命题不用验证也可以知道，因为我们罗列的等价向量组里的向量都包含在一条直线上，每个向量组都扩张成同一根直线。从代数意义上来说，在一条直线上的向量也是可以互相线性表示的。因此，可以有这样一个结论:

对于同一条直线上的向量，含有一个向量的向量组和含有多个向量的向量组等价。

2. 平面及三维空间上的等价向量组：

类似地，三维空间中，我们看看平面上的向量组之间的等价关系。如图 4-12 所示，向量$\boldsymbol{\alpha}_i$、$\boldsymbol{\beta}_i$、$\boldsymbol{\eta}_i$分别所在的三条直线共一平面（阴影平行四边形），因此向量$\boldsymbol{\alpha}_i$、$\boldsymbol{\beta}_i$、$\boldsymbol{\eta}_i$中的任何一类可以被其他两类线性表示，例如有关系$\boldsymbol{\alpha}_i = x_1\boldsymbol{\beta}_i + x_2\boldsymbol{\eta}_i$。

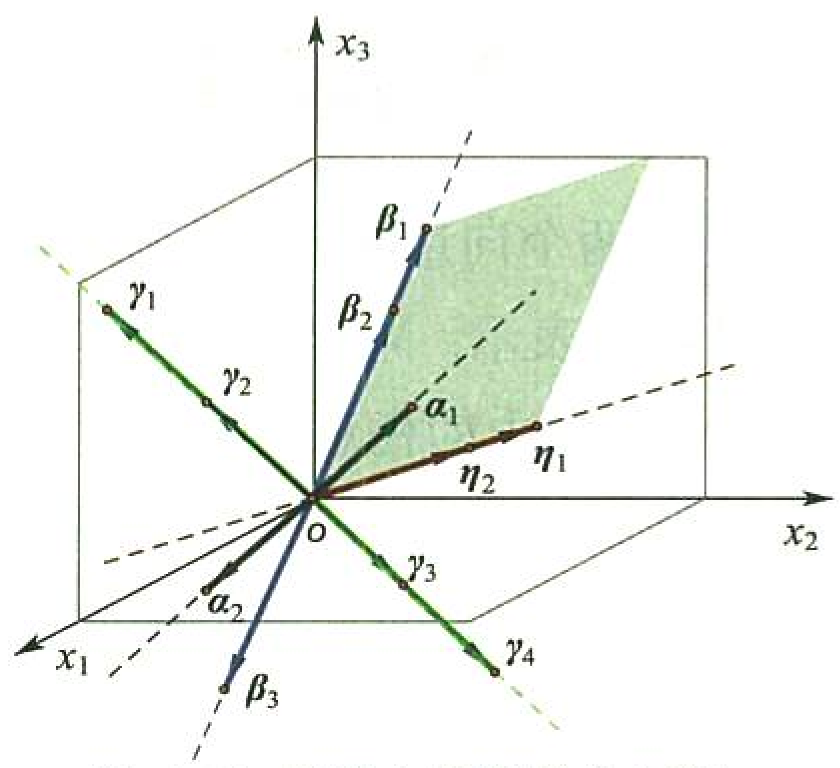


图 4-12 平面上的等价向量组

换句话说，向量$\boldsymbol{\alpha}_i$、$\boldsymbol{\beta}_i$、$\boldsymbol{\eta}_i$的任意C_3^2和C_3^3组合的向量组所张成的向量空间都是同一个平面。因此，所有组合都是等价向量组。具体的等价关系如下：

- $\{\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i\}$，$\{\boldsymbol{\alpha}_i,\boldsymbol{\eta}_i\}$，$\{\boldsymbol{\beta}_i,\boldsymbol{\eta}_i\}$，$\{\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\boldsymbol{\eta}_i\}$是等价向量组，比如$\{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2\}$，$\{\boldsymbol{\alpha}_2,\boldsymbol{\eta}_1\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2\}$，

$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2\}$ 是等价向量组。

如果一个平面再加一个平面外的一条直线 $\boldsymbol{\gamma}$ 就是向量组所张成的三维空间，则不难理解以下等价关系：

- $\{\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\boldsymbol{\gamma}_i\}$，$\{\boldsymbol{\alpha}_i,\boldsymbol{\eta}_i,\boldsymbol{\gamma}_i\}$，$\{\boldsymbol{\beta}_i,\boldsymbol{\eta}_i,\boldsymbol{\gamma}_i\}$，$\{\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\boldsymbol{\eta}_i,\boldsymbol{\gamma}_i\}$ 是等价向量组，比如 $\{\boldsymbol{\alpha}_1,\boldsymbol{\beta}_2,\boldsymbol{\gamma}_1\}$，$\{\boldsymbol{\alpha}_2,\boldsymbol{\eta}_1,\boldsymbol{\gamma}_2\}$，$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_2,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_3\}$，$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3,\boldsymbol{\gamma}_4\}$ 是等价向量组。

其实采用向量子空间（4.2 节会探讨向量空间）基的概念，就会清晰地明了向量组等价的几何意义：向量组 $\boldsymbol{A}$ 中的每一个向量都在向量组 $\boldsymbol{B}$ 张成的向量子空间中；同样，向量组 $\boldsymbol{B}$ 中的每一个向量也在向量组 $\boldsymbol{A}$ 张成的向量子空间中。反之也成立，两个向量组如果张成的向量子空间相同或重合，就说这两个向量组等价。

4.1.4　向量组的秩和极大无关组的几何意义

向量空间中的向量无穷多，因此，可以有无数个向量组等价，而且等价的向量组中的向量个数也不尽相同。

从前面的罗列中，还可以看出最短的等价向量组是只有一个向量元素的向量组；长的等价向量组的元素可以无穷多。这里，最短的向量组实际上就是极大无关组，极大无关组的元素的个数就是等价向量组的秩。如果等价向量组最小只有一个向量，则等价向量组的秩等于 1。

我们可以这样理解极大无关组：从原来的较长的向量组中挑出一部分向量组成了一个新的向量组，这个新的向量组在某种意义下可以代表原来的向量组（因为两者等价，可以互相表出）；同时这个新的向量组很纯净，没有躲在别人后面滥竽充数的向量，多余的向量被剔出了，向量之间互相独立，个顶个，既不代表谁也不被代表（任一个向量都不能被其他向量线性表示）。这些个顶个的向量个数就是这些互相等价的向量组的秩。

从几何意义上讲，在一个向量组里，如果有多个向量在一条直线上，那么直线上这些向量只要一个向量就可以了，其他的同直线的向量可以被代表了。这个向量代表可以是直线上任意一个非零向量；进而，如果向量组里还有多个向量构成或存在于一个平面上，那么只要有两个非零非共线的向量就可以代表其他的共面向量了；继续，如果向量组里还有多个向量构成或存在于一个立体空间里，那么只要有三个非零非共线非共面的向量就可以代表其他的同立体向量了……

所以，一个 n 维向量组可以通过几何意义上筛选得到一个极大无关向量组，筛选的过程可以这样：

(1) 把共线的向量全部找出来；然后对每一个直线，各留一个向量代表，线上其余向量删除。

(2) 把上述精简后的向量组里所有共面的向量全部找出来，然后对一个平面，各留两个向量，面上其余的向量删除；

(3) 把上述精简后的向量组里所有共立方体的向量全部找出来，然后对每一个立方体，各留三个向量，立方体内其余的向量删除；

……

(4) 把上述精简后的向量组里所有共超立方体（n-1 维）的向量全部找出来，然后对每一

个超立方体，各留 n-1 个向量，超立方体内其余的向量删除；

(5) 留下的向量数小于等于 n，筛选结束，剩下的向量则为极大无关向量组。

注意： 最后精简后的向量的个数就是原 n 维向量组的秩。在整个精简过程中，每一步留下的向量组虽然个数逐步减少，但每一步向量组的秩却一直没有变。秩是一个不变量。

作为例子，我们把图 4-12 中的向量组进行筛选操作，首先一根直线留一个向量(可任意)，见图 4-13（a），得到了一个向量组 $\{\alpha_1,\beta_1,\eta_1,\gamma_1\}$；然后把共面的向量留两个向量（可任意），得到了包含三个向量的向量组 $\{\beta_1,\eta_1,\gamma_1\}$，如图 4-13（b），筛选过程结束。

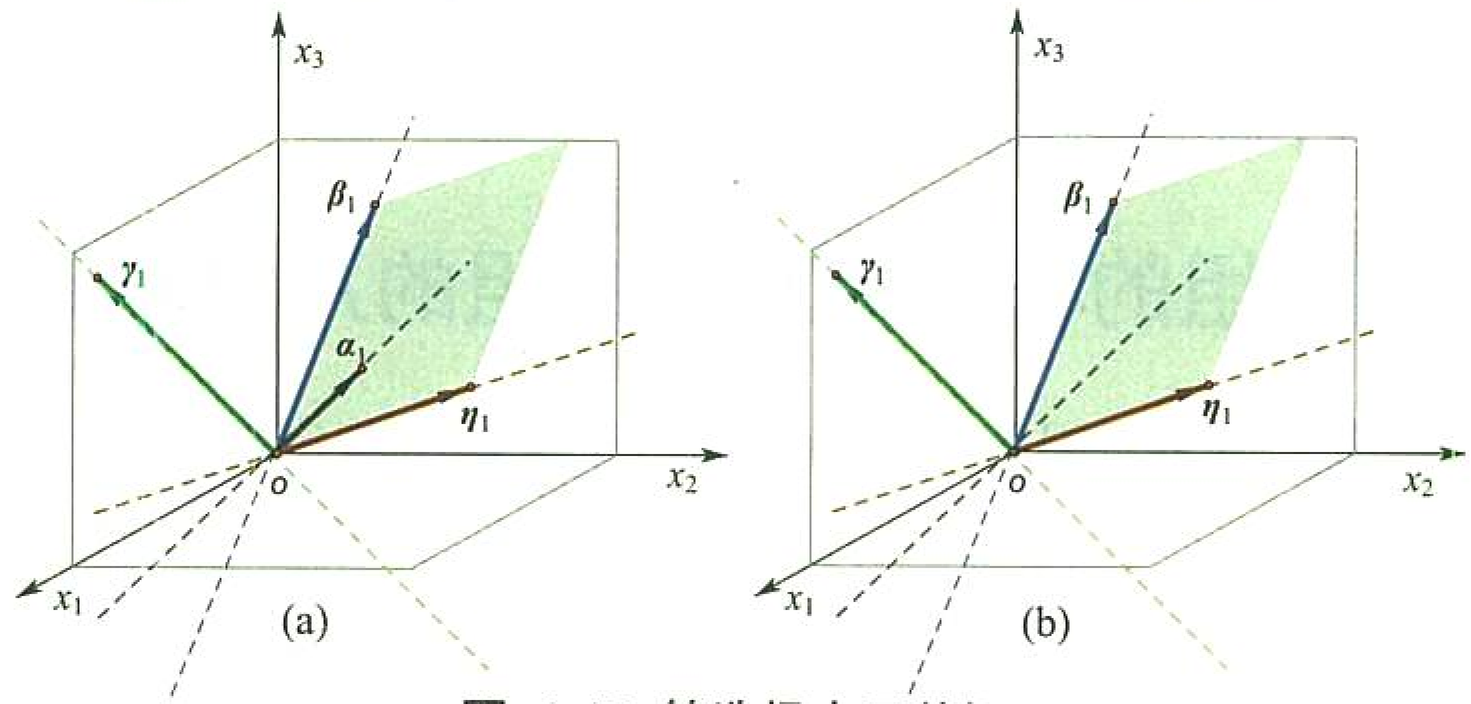


图 4-13 筛选极大无关组

总结一下，采用向量空间的概念，我们可有一个更全面的关于线性相关的几何意义的结论：

- **一个向量空间中，一个向量组线性相关的话，那么这个向量组的全部向量会张成向量空间上的一个子向量空间，且子空间的维数小于向量组元素的个数；**
- **一个向量空间中，一个向量组线性无关的话，那么这个向量组的全部向量会张成向量空间上的一个子向量空间，且子空间的维数等于向量组元素的个数；**

（在理想情况下，如果向量组里的元素没有冗余，一个向量组可以张成一个和向量组元素个数相同的子空间，一个向量张成一维的子空间，两个向量张成二维的子空间……但残酷的现实是向量世界也有多余的有裙带关系的混子，有关系的冗余存在，向量组的作用效率就低。所以，如果一个向量组 n 个元素张成一个小于 n 的子空间，那么这个向量组就线性相关；如果总是张成一个 n 维的子空间，那么这个向量组就线性无关。）

- **线性相关或无关的向量组的秩就是该向量组可以张成的最大子空间的维数；**
- **两个向量组等价，就是两个向量组张成的向量子空间相同或重合。**

4.1.5 向量组例题的图解

下面介绍两个容易搞错的命题，以加深印象。

例 4.1 错误命题 1：若向量组 $\{\alpha_1,\alpha_2,\alpha_3\}$ 中的向量两两线性无关，则 $\{\alpha_1,\alpha_2,\alpha_3\}$ 线性无关。

用二维空间的向量即可证明。

如图 4-14 所示，向量组 $\{\alpha_1,\alpha_2,\alpha_3\}$ 的向量定义如下：$\alpha_1=(2,1)$，$\alpha_2=(3,3)$，$\alpha_3=(1,2)$，显然 α_1 和 α_2，α_1 和 α_3，α_2 和 α_3 线性无关（不在一条直线上），但 $\alpha_2=\alpha_1+\alpha_3$，所以向量组 $\{\alpha_1,\alpha_2,\alpha_3\}$ 线性相关。

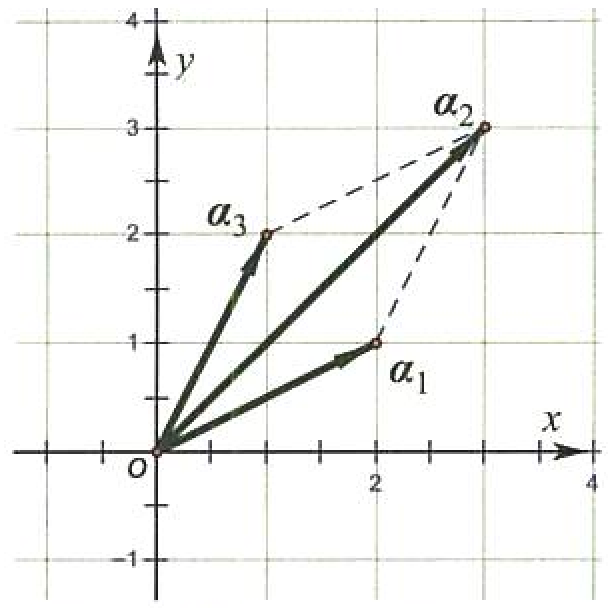


图 4-14 三个二维向量线性相关，但两两线性无关

这个例题说明，向量α_1、α_2、α_3属于二维的平面向量空间，而向量组的元素个数是 3，超过了向量空间的维数 2，因而线性相关。一般的结论是，n 维向量空间里 n+1 个以上的向量必线性相关。

例 4.2 错误命题 2：若α_1和α_2线性相关，β_1和β_2也线性相关，那么，$\alpha_1+\beta_1$和 $\alpha_2+\beta_2$也线性相关。

也用二维空间的向量证明。

如图 4-15 所示，向量定义如下：$\alpha_1=(2,1)$，$\alpha_2=(-2,-1)$，$\beta_1=(-1,2)$，$\beta_2=(-2,4)$，则有$\alpha_1+\beta_1=(1,3)$，$\alpha_2+\beta_2=(-4,3)$；$\alpha_1+\beta_1$和$\alpha_2+\beta_2$不在一条直线上，因而$\alpha_1+\beta_1$和$\alpha_2+\beta_2$线性无关。

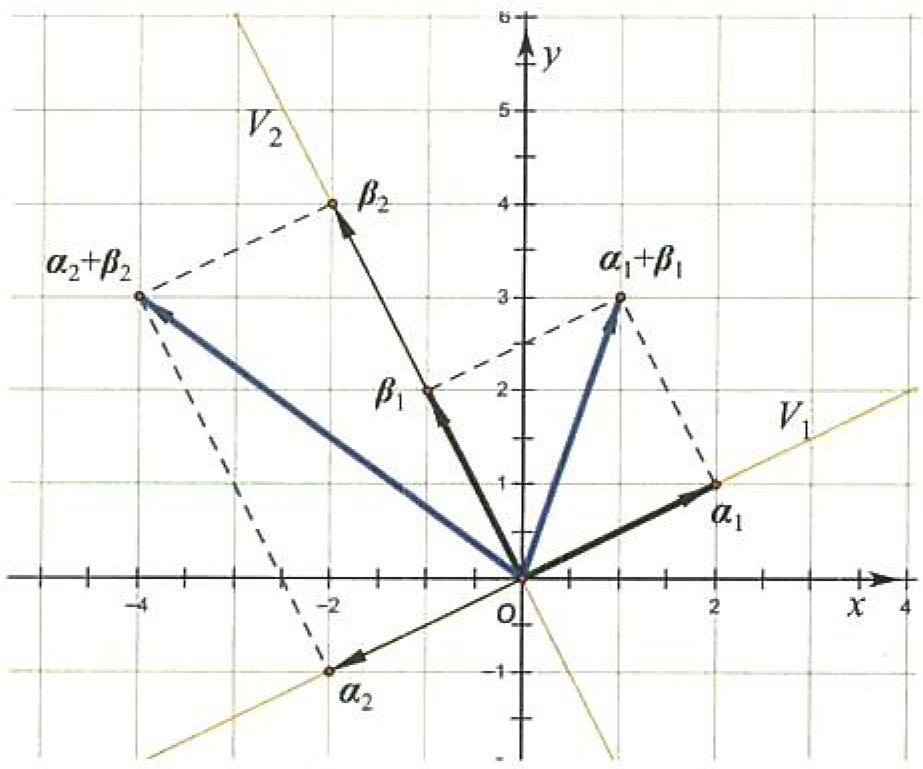


图 4-15　线性相关组的向量和不一定相关

这个例题说明，α_1和α_2在空间 V_1（直线）上，β_1和β_2在空间V_2（直线）上，两个不同空间上的向量（零向量除外）相加，必然会进入第三个空间（直线）、第四个空间（直线）……以致布满整个二维空间（平面），显然$\alpha_1+\beta_1$和$\alpha_2+\beta_2$绝大部分线性无关。

4.2 向量空间的几何意义

向量空间也就是线性空间，顾名思义，当然具有直观的几何意义。线性来源于向量直线段的几何概念，空间来源于二维平面或三维立体的几何概念。让我们真正幸运的是，所有的五花八门的线性空间（这些线性空间大多隐藏在我们的物理世界中而难以发现，隐藏在电子电路世界里面的由电阻、电感或电容组成的电路网络，比如隐藏在代数里面的多项式和高等数学里面的满足微积分运算的数的集合等）都可以和实数域上的线性空间 $\mathbf{R}^n$ 同构——空间的结构相同。什么意思？就是所有类型的线性空间都和直线、平面、三维立体以及高维空间里的变换性质一

样，所有类型的线性空间里的元素都可以和 $\mathbf{R}^n$ 空间的点（向量）相互对应。一句话，向量空间 $\mathbf{R}^n$ 的几何图形化就是所有的线性空间的几何意义或几何解释。

实际上，本书对线性代数进行几何意义上的解释正是从实数域上的向量和低维的线性空间 $\mathbf{R}^2$ 或 $\mathbf{R}^3$ 的角度全面解读的。

形形色色的向量方向、长短各异，应该给它们分类，划分成集合便于分析研究。由于向量的概念具有几何的特质，因此向量的集合通常叫做向量空间。这个向量空间里的规矩很多，有人给出八条铁律，还有的是十条。其实只有两项基本原则：一是任意两个向量叠罗汉（相加）不能超出空间；另一个是任意一个向量伸头缩脑（数乘）也不能超出空间。

我们常常发现，在向量空间这所大房子里又可以划分出好多居室，每个居室里的向量们也严格坚守着自己居室的同样的两项基本原则：叠罗汉和伸缩头脑不能出室（呵，有些像小学生，端坐笔直像一个个向量，下课了活动活动还不能出教室），这些大大小小的居室就是子空间。

需要注意的是，这些居室有个特点，就是共有一个原点，或者说都要包括零向量。空间和子空间的图形大致有如图 4-16 所示的几种类型。

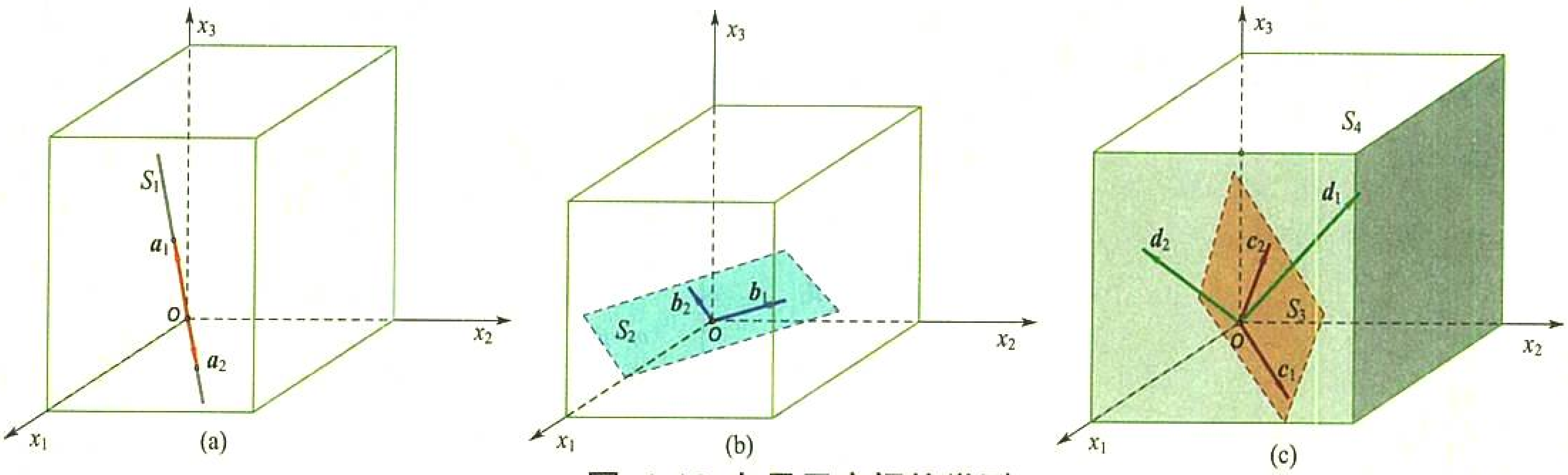


图 4-16 向量子空间的类型

向量子空间 S_1 是一根直线（见图 4-16（a）），包含向量 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$，因为两向量在直线上，$\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 相加和数乘运算不会超出直线的范围。向量子空间 S_2 是一个平面（见图 4-16（b）），包含向量 $\boldsymbol{b}_1$、$\boldsymbol{b}_2$；显然向量 $\boldsymbol{b}_1$、$\boldsymbol{b}_2$ 相加和数乘运算也不会超出平面 S_2 的范围。同样，向量子空间 S_3 的 $\boldsymbol{c}_1$、$\boldsymbol{c}_2$ 也遵守相同的规则（见图 4-16（c））；S_1、S_2 和 S_3 都是 S_4 的子空间，它们包含的向量 $\boldsymbol{a}_i$、$\boldsymbol{b}_i$、$\boldsymbol{c}_i$、$\boldsymbol{d}_i$ 都属于 S_4 空间，所有向量的相加和数乘都不会超出三维空间的范围。

在数学教科书中，向量空间的标准定义一般是这样的：

设 V 是非空的 n 维向量的集合（n=1，2，3,…），如果 V 中的向量对加法和数乘两种运算封闭，也即

- 若 $\boldsymbol{a}, \boldsymbol{b} \in V$，则 $\boldsymbol{a}+\boldsymbol{b} \in V$;
- $\boldsymbol{a} \in V$，则 $k\boldsymbol{a} \in V$，k 为任意实数，

则称 V 为向量空间。

空间和子空间的说法可以不加区别，一个空间也可以是自己的子空间。

向量空间主要有两种：一种是由 V 中的一个向量组张成的空间（比如由特征向量张成的特征子空间等）；另一种是由齐次线性方程组的解集组成的解空间。实际上，线性方程组的解空间也是由解向量所张成的。下面先看看由向量组所张成的空间的意义。

4.2.1　向量张成的空间

实际上，向量空间的概念就是对向量集合的一个划分。那么一个向量空间如何用数学式子表达呢？换句话说，一个空间里面的所有向量(无穷多)如何用有限而简洁的数学算式表达呢？

前面已说过，一个向量空间满足两个基本原则：对加法和数乘的运算封闭。把加法和数乘综合到一块，就是线性组合式 $x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$。所以，我们可以使用一个向量组的线性组合式来表达一个空间里的全部的无穷向量。这个向量组常常是极大无关向量组，也可以是向量的相关组。

一个向量组可以线性表示出一个空间里的所有向量，反过来讲，空间里的所有向量都可以分解为这个向量组的线性表示。那么这个空间我们就叫向量组张成的空间。

下面我们看看它的数学定义式。

设一个向量组 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$，这个向量组的所有的线性组合生成一个向量集合：

$$\{x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n \mid x_1,x_2,\cdots,x_n\in\mathbf{R}\}$$

此集合常称为 $\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$，即为由向量组 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$ 张成的向量空间。

请注意：这个向量空间的代数定义和前面的加法与数乘的定义是等价的。

图 4-17 中给出了由向量组 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$ 张成的向量空间平面 S 的例子：

$$S=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}=\{x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2 \mid x_1,x_2\in\mathbf{R}\}$$

图 4-17 中显示，由两个不相关的向量使用平行四边形法则可以生成平面上所有的向量。

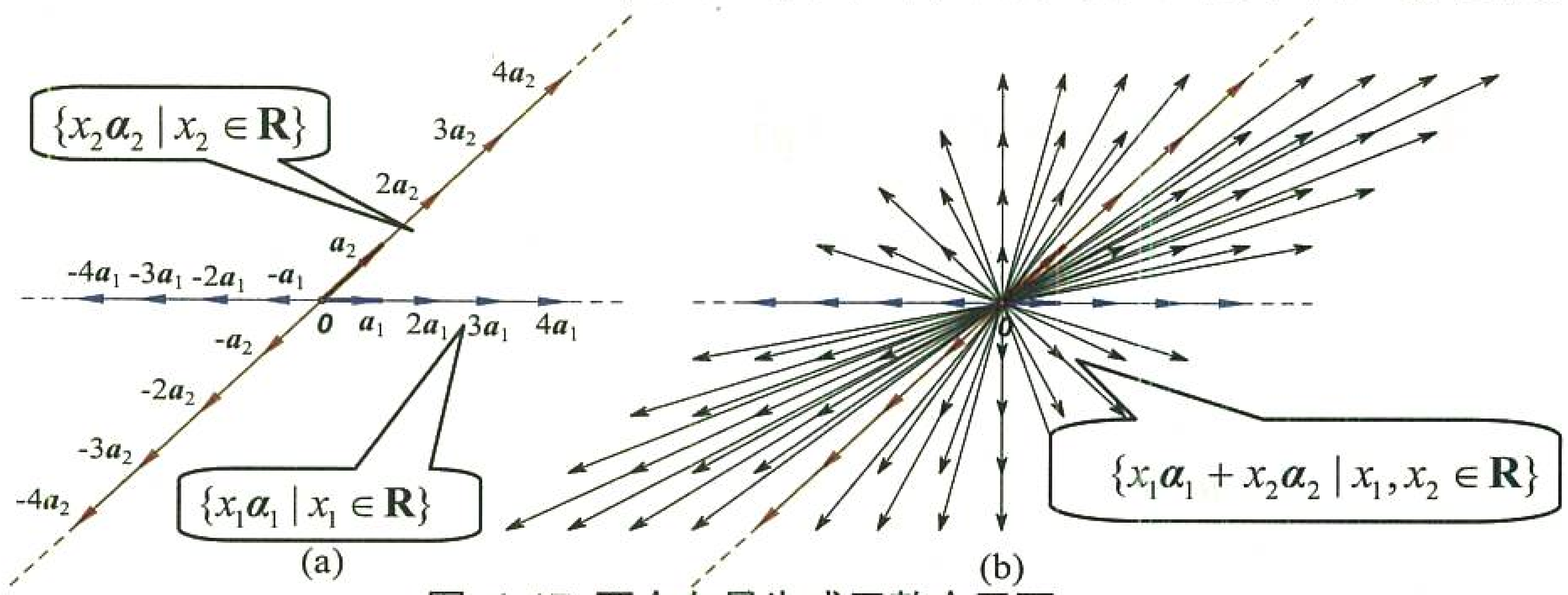


图 4-17　两个向量生成了整个平面

首先图 4-17（a）生成两个主轴：$\boldsymbol{\alpha}_1$ 的任意数乘 $x_1\boldsymbol{\alpha}_1$ 生成横轴，$\boldsymbol{\alpha}_2$ 的任意数乘 $x_2\boldsymbol{\alpha}_2$ 生成纵轴；然后图 4-17（b）显示了生成的整个平面：横轴和纵轴上的无穷多向量配对相加（几何上是平行四边形法则）生成覆盖整个平面的无穷多向量。

我们在讲行列式的几何意义时说，行列式的 m 维超平行多面体像是一个枝繁叶茂的大树所构成的一个物理空间，主枝干就是行列式的 m 个行向量或列向量。虽然向量数量可以是无穷多，但这个物理空间是有限的，空间的体积就是行列式的值。

由向量所张成的线性空间是无穷大的，空间里的向量也是无穷多的。因为在向量空间的数学定义式 $\{x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n \mid x_1,x_2,\cdots,x_n\in\mathbf{R}\}$ 中，系数可以无穷大，所以可以张成无穷大的空间。

4.2.2　子空间的几何意义

子空间的一般定义是这样的：

如果V和H都是向量空间，而且$H \subset V$，则称H是V的子空间。

具体说来，由向量空间中的一些向量张成的子空间，其定义如下：

设$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_m\}$是$n$维向量空间$V$的一个向量组，$m \leqslant n$，这个向量组的所有的线性组合生成一个向量空间：

$$\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m\}=\{x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_m\boldsymbol{\alpha}_m \mid x_1,x_2,\cdots,x_m \in \mathbb{R}\}$$

向量空间$\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m\}$称为由向量$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$张成的子空间。

这里要提醒一下，$\mathbf{0}$向量是唯一的，既属于V空间也属于H空间。任意一个子空间H都要包含$\mathbf{0}$向量，否则就不能满足加法和数乘的封闭运算。

下面来一个三维空间中由两个三维向量$\boldsymbol{a}=(a_1,a_2,a_3)$和$\boldsymbol{b}=(b_1,b_2,b_3)$张成的一个平面二维空间的图例（见图 4-18）。

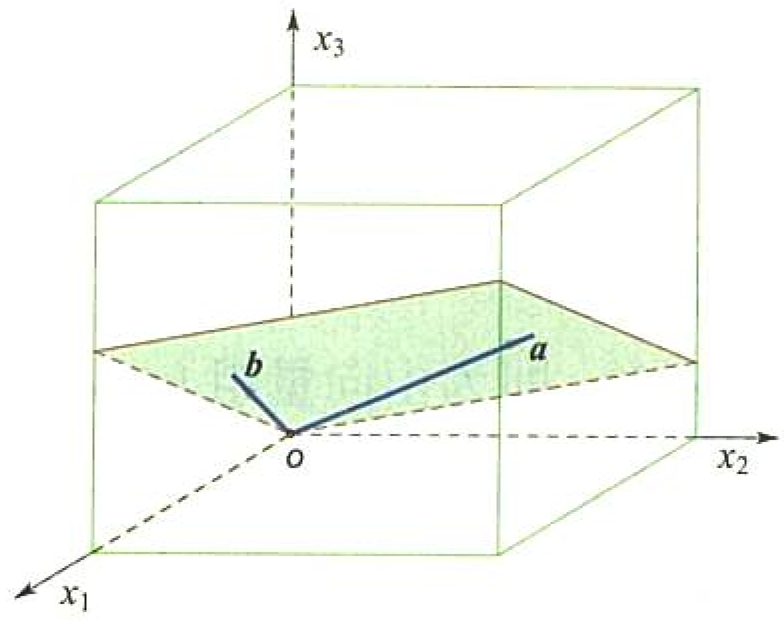


图 4-18 三维空间中的二维子空间

这里注意一个小细节：两个**三维**的向量张成了一个**二维**的平面，这个过原点的二维平面是三维空间的二维子空间。二维子空间里的向量和三维空间里的向量一样都是三维向量。

1. *n*维实线性空间$\mathbf{R}^n$的子空间

$\mathbf{R}^n$表示所有n维实向量所构成的集合。每个向量中的元素是实数，元素个数是n个。如$\mathbf{R}^2$表示平面实向量集合，$\mathbf{R}^3$表示三维空间实向量集合。

三维向量空间$\mathbf{R}^3$的所有子空间包括：

三维子空间：本身$\mathbf{R}^3=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$（$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、$\boldsymbol{\alpha}_3$线性无关），作为自身的子空间表现为一个立体空间，同自身一样，也包含原点；

二维子空间：如$\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$（$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$线性无关），表现为通过原点的任意一个平面（**注意：二维空间$\mathbf{R}^2$不是$\mathbf{R}^3$的子空间**）；

一维子空间：如$\mathrm{Span}\{\boldsymbol{\alpha}_1\}$ $(\boldsymbol{\alpha}_1 \neq \mathbf{0})$，表现为通过原点的任意一条直线；

零维子空间：只包含原点$\mathbf{0}$向量，只有零空间。

图 4-19 给出了$\mathbf{R}^3$的所有子空间的图形。

图 4-19 中，V为三维向量空间即$\mathbf{R}^3$，它可以由$\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$（$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、$\boldsymbol{\alpha}_3$线性无关）表示；$H=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$（$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$线性无关），表示一个二维子空间；$K=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_3\}$（$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_3$线性无关），表示另一个二维子空间；$H$和$K$的公共集合交集$H \cap K=\mathrm{Span}\{\boldsymbol{\alpha}_1\}$ $(\boldsymbol{\alpha}_1 \neq \mathbf{0})$，是一维向量子空间；上述所有的子空间皆包含零向量$\mathbf{0}$=（0，0，0）；当然零向量自身可以组成一个零空间。

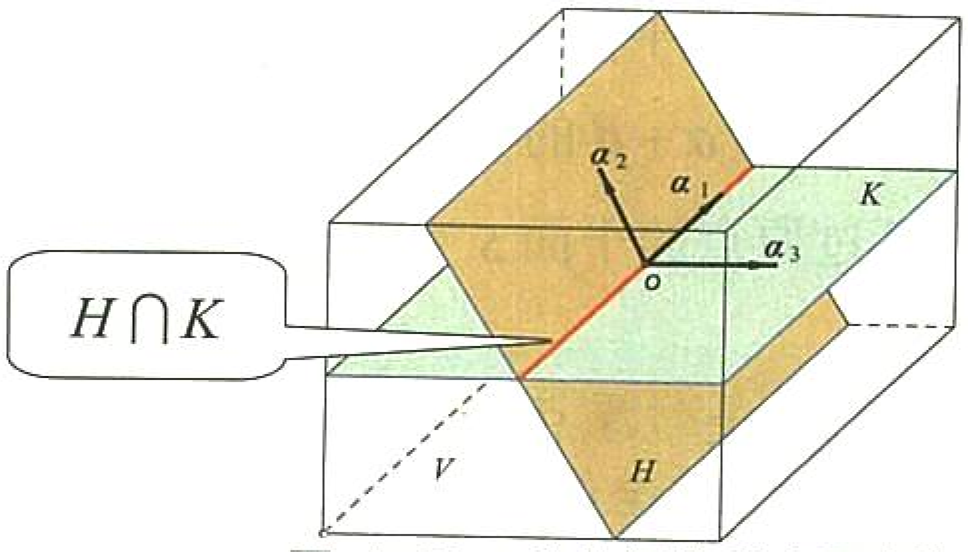


图 4-19 $\mathbf{R}^3$ 空间的所有子空间

2. 子空间要过原点的几何意义

前面在介绍子空间的概念时，总是在强调过原点，或者所有的子空间一定要包含零空间在内。为什么？这是硬性规定吗？

实际上，在坐标系下讨论的向量，不能称之为自由向量，因为所有的向量的尾部都被拉到了原点上，或者说，空间里的所有向量都是从原点出发的（见图 4-20），大家都有一个共同的零空间，这就是为什么所有的子空间一定要包含零空间的原因了。

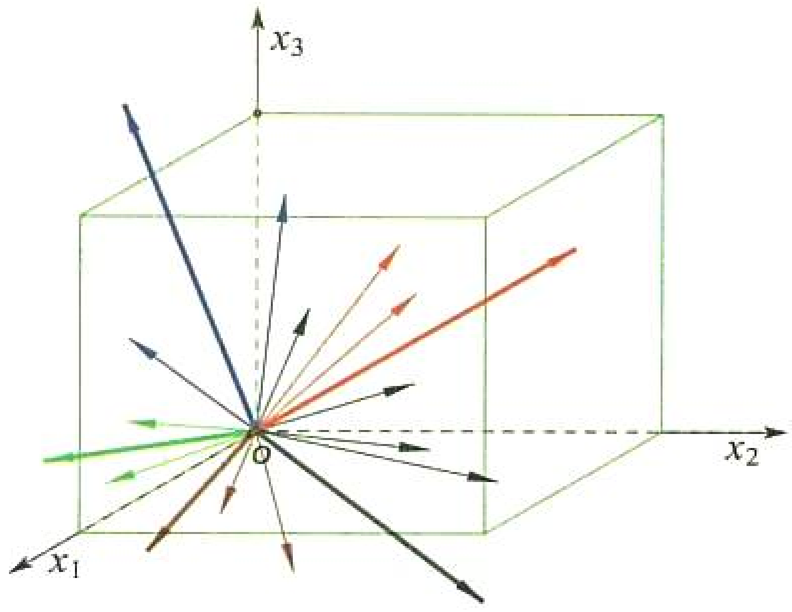


图 4-20 空间里的所有向量都是从原点出发的

为什么要把向量的尾部都拉到原点呢？在前面向量的基础几何意义一章讲过，那就是为了用解析方法研究向量的方便，因为这样就可以把向量和空间中的点一一对应起来。空间中一旦建立起了坐标系，点有坐标值，那么我们就用点的坐标表示与点对应的向量，这样向量就有了解析式，就有了向量的坐标表达式，我们就可以方面地使用代数中的矩阵技术进行分析计算了。

假设一个子空间没有通过原点，那么从原点出发的向量必然“头尾不顾”，造成了向量头部在子空间上而尾部在空间外（因为原点在空间外）。当然，向量的加法和数乘也都跑出这个子空间外面去了，如图 4-21 所示。这说明对线性变换运算不封闭，因此这个“子空间”不是真正的向量空间。

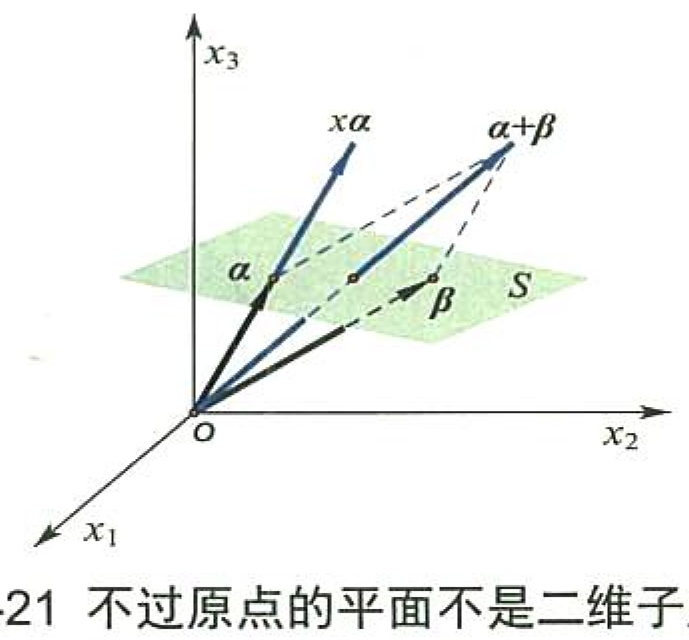


图 4-21 不过原点的平面不是二维子空间

图 4-21 中，严格来讲向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 并不在平面空间 S 中，因为只有向量的身体的全部在上面才算数，而向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 只有头部在平面上。但如果 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 向量是采用点坐标解析式表达的话，

我们常会误认为这两个向量（实际是点）在平面上。即便这样，向量$\boldsymbol{\alpha}$的数乘$x\boldsymbol{\alpha}$的头部还是超出了平面S之上，$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$相加的和向量$\boldsymbol{\alpha}+\boldsymbol{\beta}$的头部也超出了平面之上。因此，它们对向量的加法和数乘运算不封闭。所以，不过原点的平面S绝对不是二维向量子空间。

实际上，在三维几何向量空间中，凡是过原点的平面或直线上的全体向量组成的集合都是$\mathbf{R}^3$的子空间，而不过原点的平面或直线上的全体向量组成的集合都不是$\mathbf{R}^3$的子空间。

4.2.3 基、维数及其坐标的几何意义

在解析几何中，为了研究几何图形的变换，我们总是在一个固定的坐标系中讨论，进而把几何问题转化为代数问题。同样，在n维空间几何中，我们也要选定一组基底来应用向量和矩阵分析的工具。

对于向量空间V中的一个有序向量组$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$，若满足：

- $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$线性无关；
- V中任意一个向量$\boldsymbol{\alpha}$都可以由$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$线性表示，即$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\ldots+x_n\boldsymbol{\alpha}_n$，

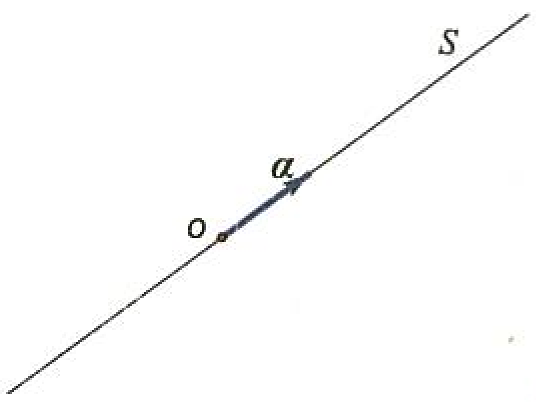
那么称向量组$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$为向量空间$V$的一个**基**；称向量组$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$的元素个数$n$为向量空间$V$的**维数**；称有序数组$(x_1,x_2,\cdots,x_n)$为向量$\boldsymbol{\alpha}$在基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$上的**坐标**。

1. 基的几何意义

基是向量空间的一组很“结实”的向量集合，每一个基向量可以像房屋地基的每一块石块一样支撑衍生出空间中的全部向量。因此，首先一个基能代表或衍生出空间里的所有的向量，缺一不可；其次，作为基的每一个向量都是个顶个，谁也不能代表谁，它们必须线性无关，它是一个极大无关向量组。

我们给一个向量空间找一个基，目的是为了给这个空间定一个坐标系，以方便我们定位和计算向量。一个基实际上就是选取的一个坐标系，另外一个基就是选取的一个新的坐标系。基是坐标系在线性空间中的推广。基向量对应坐标系的坐标轴，有几个基向量就有几个坐标轴，n维空间的一个基就需要有n个基向量。下面我们看看$\mathbf{R}^n$空间中的几个基的例子。

图 4-22（a）中，一维向量空间S是一条过 0 点的直线，向量$\boldsymbol{\alpha}_1\neq\mathbf{0}$并属于直线$S$，因而可以讲$S$是向量$\boldsymbol{\alpha}$张成的向量空间$S=\mathrm{Span}(\boldsymbol{\alpha})$，所以向量组$\{\boldsymbol{\alpha}\}$是向量空间$S$的一个基。



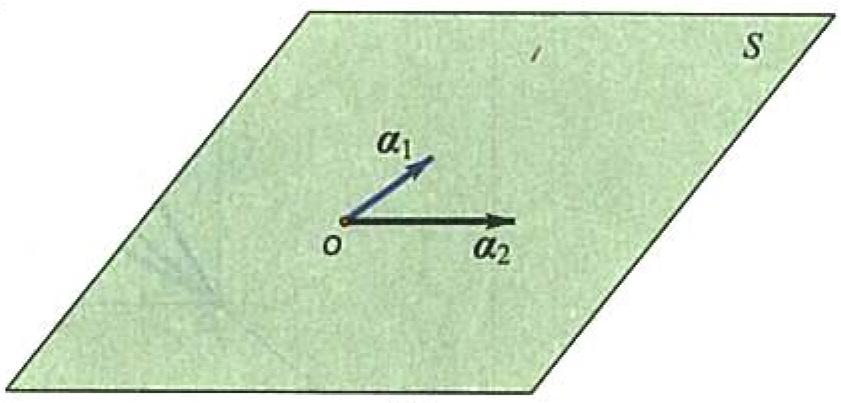


图 4-22 一维和二维空间的基

图 4-22（b）中，如果二维向量空间S是$\mathbf{R}^3$中的一个平面，且$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$是平面S上的任意的两个向量，其中任意一个都不是另外一个向量的倍数，因此向量组$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$线性无关 。平面$S$

可以看做向量α_1、α_2张成的向量空间$S=\text{Span}\{\alpha_1,\alpha_2\}$，所以向量组$\{\alpha_1,\alpha_2\}$是向量平面的一个基。

在图 4-23 中，三维向量空间S是$\mathbf{R}^3$，三个标准单位向量$\{e_1,e_2,e_3\}=\{(1,0,0),(0,1,0),(0,0,1)\}$，因为$e_1$、$e_2$、$e_3$彼此线性无关，可以生成$\mathbf{R}^3$，因此向量组$\{e_1,e_2,e_3\}$是$\mathbf{R}^3$的一个基。这个基的基向量是由标准单位向量组成的，因此又称为标准基。

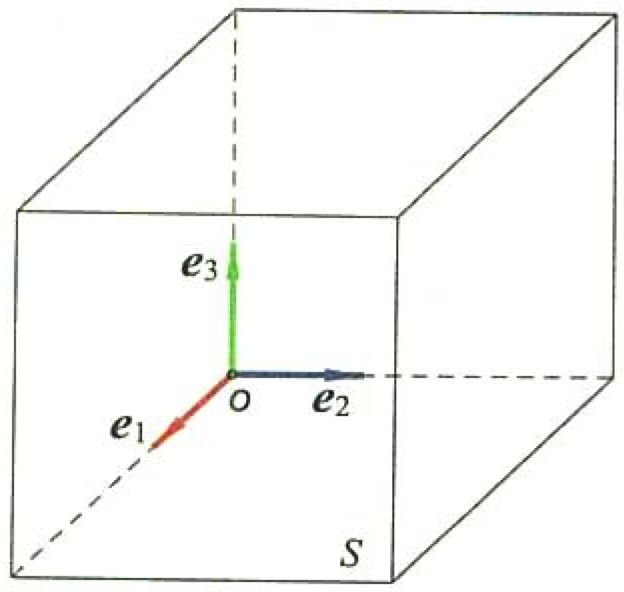


图 4-23 三维空间的标准基

2. 坐标与维数的几何意义

一个基包含的向量个数就是坐标轴的个数，也就是向量空间的维数。维数是空间的一个本质特征，它不依赖于基的选取。选取不同的基，基向量的个数不会改变，维持支撑空间的维数不会改变。这就是为何称之为“维数”的原因。

一个向量空间的基选定后，其坐标是什么？如何求取？下面我们接着看看几个图示的例子。

1) 一维基及其坐标刻度

一维空间S，维数为 1，只需一个基向量。当选取的基为$\{\alpha\}$时，坐标选取见图 4-24（a)；当选取的新基向量β为0.5α时，坐标刻度的密度加大一倍，见图 4-24（b)；当选取的新基向量γ为$-\alpha$时，坐标轴方向也随之反转，见图 4-24（c）。

显然，坐标轴的刻度是以所选基向量的长度为基本单位的。

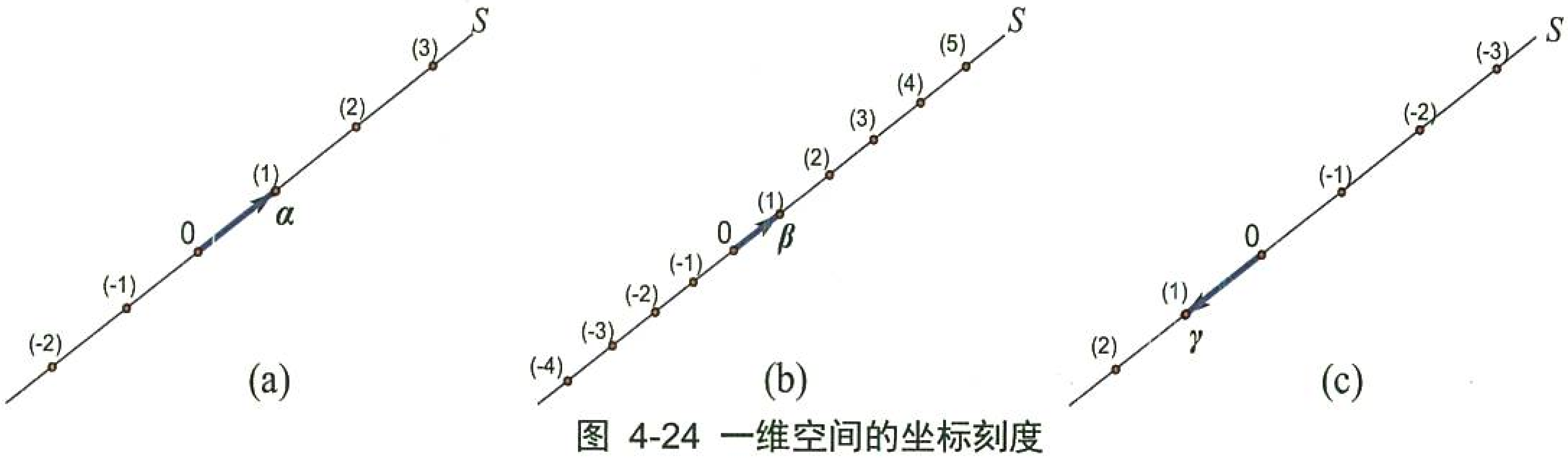


图 4-24 一维空间的坐标刻度

2) 二维基及其坐标刻度

二维空间S，维数为 2，需要两个基向量。当选取的基为$\{\alpha_1,\alpha_2\}$时，坐标选取见图 4-25（a）。两个坐标轴分别与向量α_1、α_2共线，刻度的划分是遵循向量加法的平行四边形规则；或者说，坐标网络就是由坐标轴上的基向量为基本单位作平行线所构成（注意，在这里我们不要沿袭笛卡尔坐标系的习惯，试图把空间中的一点（一个向量）向坐标轴作垂直投影）。

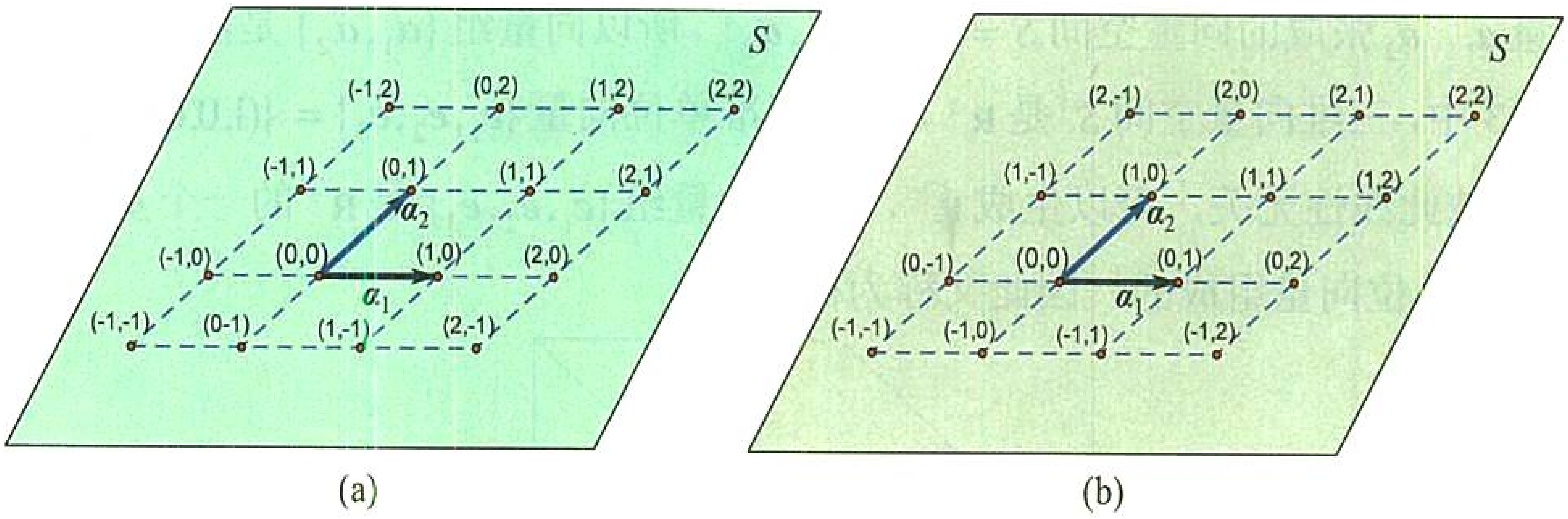


图 4-25 二维空间的坐标刻度

另外，在二维空间中，确定基向量的顺序是必要的。在向量组的讨论中我们不强调向量组中向量们的顺序，但作为一个基的向量组就要有顺序了。显然，如果基向量的顺序进行了调整，坐标值也相应进行调整。在图 4-25（b）中，我们把图 4-25（a）的空间 S 的基 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$ 调换了顺序成为一个新的基 $\{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_1\}$，当然空间中的坐标也变了。

另外，我们在上述的例子中也看到了基与直角坐标系的不同，两个基向量不一定垂直；在刻画坐标网络时不是直角坐标系的垂直投影，而是平行四边形坐标网络，分割一个坐标轴的坐标线是与另外一个坐标轴平行的关系。一个基向量的方向是对应坐标轴的正方向，坐标单位是基向量的长度。

3) 三维基及其坐标刻度

三维空间的一个基包含了三个线性无关的向量 $\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$，空间以 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 为基的坐标刻画满足平行六面体法则，见图 4-26，向量 $(1,1,1)$ 是与原点相对应的平行六面体的对角点。

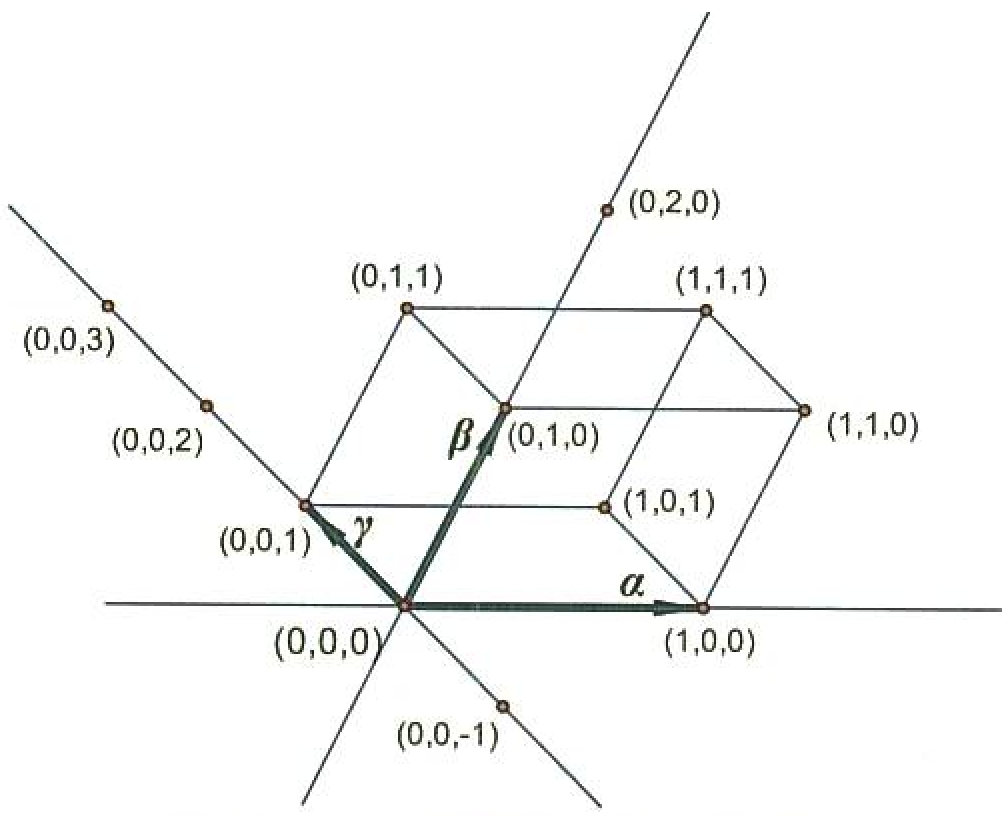


图 4-26 三维空间的坐标刻度

下面对于一个三维空间中的二维子空间，我们看看它的基及其坐标是如何刻画的。

4) 三维空间的子空间的基及其向量坐标

在图 4-26 的三维空间的例子中，设向量 $\boldsymbol{a}$ 的坐标是 $(1,1,0)$。如果我们要研究由向量组 $\{\boldsymbol{\alpha},\boldsymbol{\beta}\}$ 张成的子空间 $S=\mathrm{Span}\{\boldsymbol{\alpha},\boldsymbol{\beta}\}$，这个二维的子空间现以向量组 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 为基，那么向量 $\boldsymbol{a}$ 在 S 中的坐标是什么？从图 4-27 中可以看出，向量 $\boldsymbol{a}$ 的新坐标是 $(1,1)$。

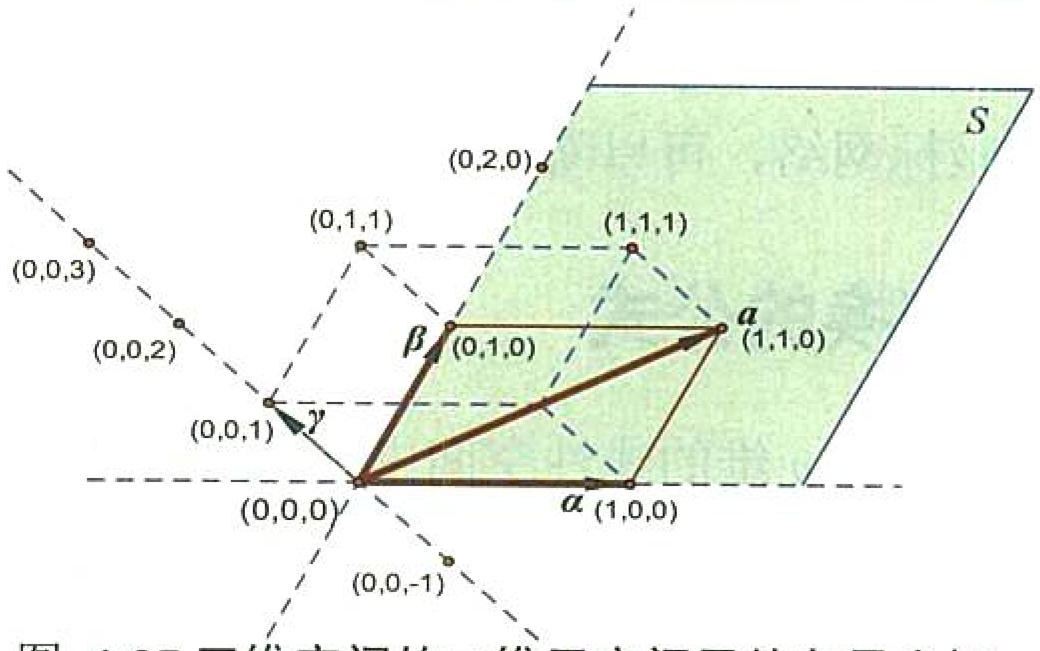


图 4-27 三维空间的二维子空间里的向量坐标

总结一下：向量 $\boldsymbol{a}$ 在三维空间 $\text{Span}\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$ 中的 $\boldsymbol{B}$ 坐标是 $(1,1,0)$，其中 $\boldsymbol{B}=\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$；而向量 $\boldsymbol{a}$ 在二维空间 $\text{Span}\{\boldsymbol{\alpha},\boldsymbol{\beta}\}$ 中的 $\boldsymbol{B}$ 坐标变成了 $(1,1)$，其中 $\boldsymbol{B}=\{\boldsymbol{\alpha},\boldsymbol{\beta}\}$。

空间坐标系

前面说过，建立坐标系的目的就是把空间向量的线性变换转化为坐标的运算。在我们讨论向量和矩阵以及向量方程中，仿射坐标系和直角坐标系最为有用。

在空间中任取一点 0，以点 0 为起点任意作三个不共面的向量 ε_1、ε_2、ε_3，这就建立了一个仿射坐标系，记为 $\{0;\varepsilon_1,\varepsilon_2,\varepsilon_3\}$，有了这个仿射坐标系，我们就可以建立空间向量和三元有序数组（即坐标值）之间的一一对应关系了。就是说，在坐标系中的任意向量 $\boldsymbol{a}$ 可用唯一的有序数组 (a_1,a_2,a_3) 来表示关系：$\boldsymbol{a}=a_1\varepsilon_1+a_2\varepsilon_2+a_3\varepsilon_3$。

在这里，点 0 称为坐标原点，ε_1、ε_2、ε_3 称为坐标向量或基，过原点且与坐标向量同向的直线称为坐标轴，有序数组 (a_1,a_2,a_3) 称为向量 $\boldsymbol{a}$ 的坐标。

仿射坐标系按手征性分为左手坐标系和右手坐标系。

特别地，如果仿射坐标系 $\{0;\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\}$ 的基 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 是两两垂直的单位向量，则称之为直角坐标系。$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 所在的坐标系分别称为 x 轴、y 轴和 z 轴。我们常用的直角坐标系为右手直角坐标系。

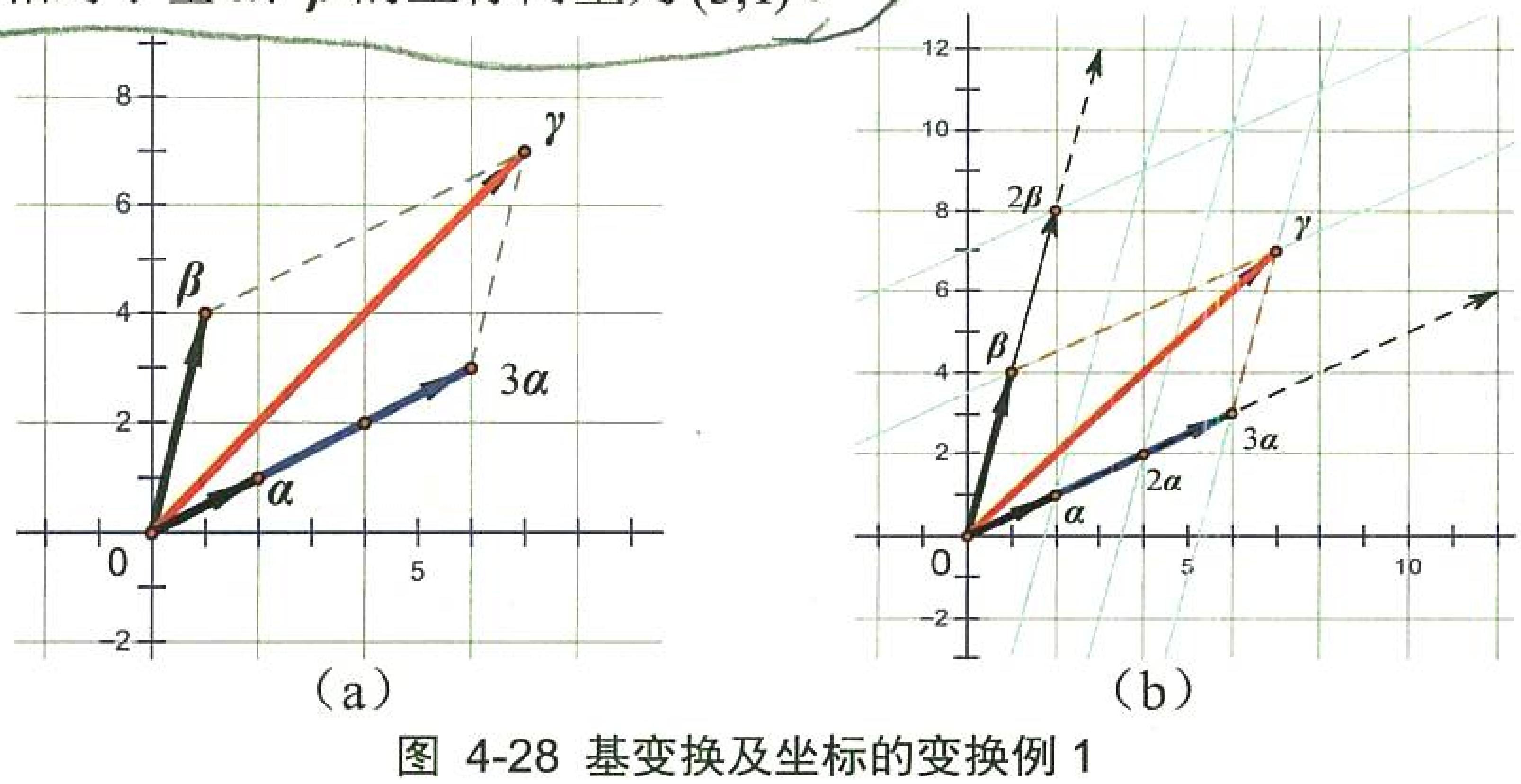
4.2.4 基变换的几何意义

1. 基变换及坐标变换的图解

在直角平面坐标系（实际上为单位正交基 $\boldsymbol{i}$、$\boldsymbol{j}$）的空间中，有两个向量 $\boldsymbol{\alpha}=(2,1)$ 和 $\boldsymbol{\beta}=(1,4)$，如图 4-28（a）所示。这两个向量线性无关（不成倍数关系），让它们构成平面空间一对新基。有另外一个向量 $\boldsymbol{\gamma}=(7,7)$，则可以把它拼凑为这两个新基的线性组合（而且组合是唯一的）：

$$\boldsymbol{\gamma}=3\boldsymbol{\alpha}+\boldsymbol{\beta}$$

那么，向量 $\boldsymbol{\gamma}$ 相对于基 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 的坐标向量为 $(3,1)$。



（a）　（b）

图 4-28 基变换及坐标的变换例 1

图 4-28（a）为原坐标系下的三个向量，向量 γ 可以通过扩张 $\boldsymbol{\alpha}$ 到三倍，并与 $\boldsymbol{\beta}$ 合并而成；图 4-28（b）为重新划分的坐标网络，可以看出，向量 γ 在新网络下的坐标值确实是 $(3,1)$。

2. 基变换及坐标变换的公式

前面的讨论告诉我们，在一个 n 维的线性空间中，可以取不同的 n 元素无关向量组作为基。那么这个线性空间的任意两个基之间必有关联，这个关联是什么？还有，一个向量 $\boldsymbol{a}$ 在一个确定的基下有一个确定的坐标，这个向量在不同的基下有不同的坐标。第二个问题是，任意两个基上的坐标之间有什么关联？

一个 n 维线性空间 V ，$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 是 V 的两个基。先将 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 作为空间的基，那么向量组 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 可以由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 线性表示为

$$\boldsymbol{\beta}_1 = a_{11}\boldsymbol{\alpha}_1 + a_{12}\boldsymbol{\alpha}_2 + \cdots + a_{1n}\boldsymbol{\alpha}_n = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \end{pmatrix} = (a_{11},\ a_{12}, \cdots, a_{1n})\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix}$$

$$\boldsymbol{\beta}_2 = a_{21}\boldsymbol{\alpha}_1 + a_{22}\boldsymbol{\alpha}_2 + \cdots + a_{2n}\boldsymbol{\alpha}_n = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2n} \end{pmatrix} = (a_{21},\ a_{22}, \cdots, a_{2n})\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix}$$

$$\vdots$$

$$\boldsymbol{\beta}_n = a_{n1}\boldsymbol{\alpha}_1 + a_{n2}\boldsymbol{\alpha}_2 + \cdots + a_{nn}\boldsymbol{\alpha}_n = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nn} \end{pmatrix} = (a_{n1},\ a_{n2}, \cdots, a_{nn})\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix}$$

将以上 n 个表示式合并为

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{bmatrix} a_{11} & a_{21} & \dots a_{n1} \\ a_{12} & a_{22} & \dots a_{n2} \\ \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \dots a_{nn} \end{bmatrix} \tag{4-1}$$

或者

$$\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots a_{nn} \end{bmatrix}\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix} \tag{4-2}$$

式（4-1）或式（4-2）称为由基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的基变换公式，其中，矩阵 $\boldsymbol{P} = \begin{bmatrix} a_{11} & a_{21} & \dots a_{n1} \\ a_{12} & a_{22} & \dots a_{n2} \\ \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \dots a_{nn} \end{bmatrix}$ 或 $\boldsymbol{P}' = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots a_{nn} \end{bmatrix}$ 称为由基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵。

注意两点：

（1） $\boldsymbol{P}$ 和 $\boldsymbol{P}'$ 互为转置矩阵，这取决于你将基向量 $\boldsymbol{\alpha}_i$ 和 $\boldsymbol{\beta}_i$ 看做是行向量还是列向量；

（2） 过渡矩阵 $\boldsymbol{P}$ 的列向量和 $\boldsymbol{P}'$ 的行向量分别是 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 在基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 上的坐标（按序排列成矩阵），换句话说，把 $\boldsymbol{\beta}_i$ 在基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 上的坐标一列列地排列而成过渡矩阵 $\boldsymbol{P}$ 。

以上得到的矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{P}'$ 给出了第一个问题的答案，就是线性空间的两个基之间是可以互

相转换或变换的，变换的矩阵称为过渡矩阵。下面我们看看一个向量的坐标的转换是什么？

设向量 $\boldsymbol{a}$ 在两个基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 下的坐标分别是 $(x_1,x_2,\cdots,x_n)$ 和 $(x'_1,x'_2,\cdots,x'_n)$，则向量 $\boldsymbol{a}$ 在基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 上表示为

$$\boldsymbol{a}=(x_1,x_2,\cdots,x_n)\begin{pmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\\ \vdots\\ \boldsymbol{\alpha}_n\end{pmatrix}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix} \tag{4-3}$$

同时，向量 $\boldsymbol{a}$ 在基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 上表示为

$$\boldsymbol{a}=(x'_1,x'_2,\cdots,x'_n)\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\\ \vdots\\ \boldsymbol{\beta}_n\end{pmatrix}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\begin{pmatrix}x'_1\\ x'_2\\ \vdots\\ x'_n\end{pmatrix} \tag{4-4}$$

把前面的基变换公式（4-2）和公式（4-1）代入式（4-4）得到

$$\boldsymbol{a}=(x'_1,x'_2,\cdots,x'_n)\boldsymbol{P}'\begin{pmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\\ \vdots\\ \boldsymbol{\alpha}_n\end{pmatrix}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{P}\begin{pmatrix}x'_1\\ x'_2\\ \vdots\\ x'_n\end{pmatrix} \tag{4-5}$$

因为一个向量在同一个基下的坐标是唯一的，所以，对比式（4-5）和式（4-3）中的坐标部分，得到

$$(x_1,x_2,\cdots,x_n)=(x'_1,x'_2,\cdots,x'_n)\boldsymbol{P}' \tag{4-6}$$

或

$$\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=\boldsymbol{P}\begin{pmatrix}x'_1\\ x'_2\\ \vdots\\ x'_n\end{pmatrix} \tag{4-7}$$

至此，我们得到了坐标转换的公式。

假设已知坐标 $(x'_1,x'_2,\cdots,x'_n)$，直接使用式（4-7）即得到新基下的坐标 $(x_1,x_2,\cdots,x_n)$。反之，如果已知 $(x_1,x_2,\cdots,x_n)$ 求 $(x'_1,x'_2,\cdots,x'_n)$，即公式

$$\begin{pmatrix}x'_1\\ x'_2\\ \vdots\\ x'_n\end{pmatrix}=\boldsymbol{P}^{-1}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix} \tag{4-8}$$

在上述推导中，一并给出了行向量和列向量的情形。不过大多教材只处理列向量的情形，我们以后也只处理列向量，或把向量统统默认为列向量。

基过渡矩阵和向量转换矩阵

请注意一个容易搞错的地方：在列向量的习惯下，由旧基矩阵计算得到新基矩阵的变换矩阵是基过渡矩阵或转换矩阵 $\boldsymbol{P}$，公式为式（4-1），右乘旧基矩阵；而任一个向量由其旧坐标计算新坐标的公式是式（4-8），其向量坐标的转换矩阵是 $\boldsymbol{P}^{-1}$，左乘向量，用的是基过渡矩阵的逆阵。

下面我们举例说明具体的基过渡矩阵和坐标转换计算。

例 4.3 如图 4-29 所示，二维平面空间的一组基为 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$（没有具体值），平面有一向量 $\boldsymbol{a}$，

由图知 $\boldsymbol{a}=(-2,2)^{\mathrm{T}}$。我们再取一组基 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$，新基在原来基 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 上的坐标分别是 $(-\frac{1}{2},\frac{3}{2})^{\mathrm{T}}$，$(-1,-1)^{\mathrm{T}}$。把它们按顺序排列起来，得到了由基 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$ 过渡到基 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 的过渡矩阵为

$$\boldsymbol{P}=\left[\begin{array}{c:c}-\frac{1}{2} & -1 \\ \frac{3}{2} & -1\end{array}\right]$$

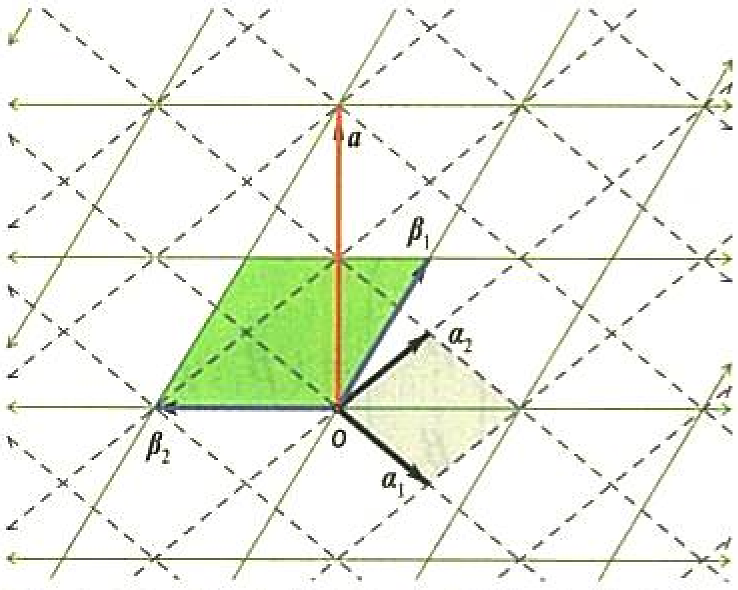


图 4-29 基变换及坐标的变换例 2

过渡矩阵也可以由列出的基的表示式方程组得到。把 $\boldsymbol{\beta}_1=-\frac{1}{2}\boldsymbol{\alpha}_1+\frac{3}{2}\boldsymbol{\alpha}_2$，$\boldsymbol{\beta}_2=-\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2$ 按列向量排列即可析出过渡矩阵

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\begin{bmatrix}-\frac{1}{2} & -1 \\ \frac{3}{2} & -1\end{bmatrix}$$

过渡矩阵对不对，我们验证一下。用过渡矩阵求向量 $\boldsymbol{a}=(-2,2)^{\mathrm{T}}$ 在新基 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 上的坐标，用坐标转换公式：

$$\begin{pmatrix}x'_1 \\ x'_2\end{pmatrix}=\boldsymbol{P}^{-1}\begin{pmatrix}x_1 \\ x_2\end{pmatrix}=\begin{bmatrix}-\frac{1}{2} & -1 \\ \frac{3}{2} & -1\end{bmatrix}^{-1}\begin{pmatrix}-2 \\ 2\end{pmatrix}=\frac{1}{2}\begin{bmatrix}-1 & 1 \\ -\frac{3}{2} & -\frac{1}{2}\end{bmatrix}\begin{pmatrix}-2 \\ 2\end{pmatrix}=\begin{pmatrix}2 \\ 1\end{pmatrix}$$

看看图 4-29，向量 $\boldsymbol{a}$ 在基 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 的坐标网络中的坐标正是 $\begin{pmatrix}2 \\ 1\end{pmatrix}$，过渡矩阵和坐标转换公式得到验证。

4.2.5 欧式空间及内积推广

诸位可能注意到了，前面讨论的向量空间没有谈到向量长度之类的问题。是的，不涉及长度和角度的向量空间一样可以自成体系。如果在向量空间里再定义向量的长度和角度等概念必须定义内积，定义了内积的向量空间称为欧氏空间。

在第第 2 章章中，我们把向量内积称为点积，简单的内积运算写做 $\boldsymbol{\alpha}\cdot\boldsymbol{\beta}$ 很方便，如果多向量作内积运算估计容易弄丢小点点，常把内积运算表示式的向量用逗点分开再用小括号括起来，如 $(\boldsymbol{\alpha},\boldsymbol{\beta})$。

1. 欧式空间是加了内积运算的线性空间

在现实世界的二、三维空间里，我们是需要对向量进行度量的，比如要知道一个向量的长度是多少啊、两个向量的夹角是多少啊、两个向量是否正交啊等问题（见图 4-30）；同时，由

向量构成的几何图形也有距离和角度这两个重要的性质等。咋度量呢？或者说在一个线性空间中如何定义长度和夹角呢？

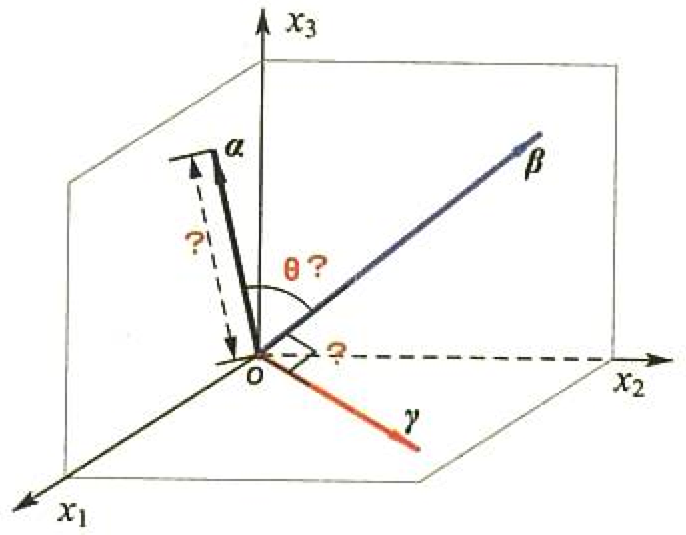


图 4-30 欧式空间的度量

实际上，如果我们定义了两个向量的内积，那么空间里向量的长度、夹角、正交等问题皆可以迎刃而解。

比如以 $\mathbf{R}^2$ 上的向量为例，一个向量 $\boldsymbol{\alpha}=(x_1,x_2)^{\mathrm{T}}$，那么它的长度是以直角边边长为 $|x_1|$、$|x_2|$ 的直角三角形斜边的长度，据勾股定理得到该向量的长度为 $\sqrt{x_1^2+x_2^2}$，用内积表述向量的长度就是

$$|\boldsymbol{\alpha}|=\sqrt{x_1^2+x_2^2}=\sqrt{(x_1,x_2)(x_1,x_2)^{\mathrm{T}}}=\sqrt{(\boldsymbol{\alpha}^{\mathrm{T}},\boldsymbol{\alpha})} \qquad (4\text{-}9)$$

两个向量的夹角可以间接地用其夹角的三角函数表示。假设另一个向量 $\boldsymbol{\beta}=(y_1,y_2)^{\mathrm{T}}$，$\boldsymbol{\alpha}$ 与 x 轴的夹角为 θ_1，$\boldsymbol{\beta}$ 与 y 轴的夹角为 θ_2，$\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的夹角为 θ，则有

$$\cos\theta_1=\frac{x_1}{\sqrt{x_1^2+x_2^2}},\quad \sin\theta_1=\frac{x_2}{\sqrt{x_1^2+x_2^2}},\quad \cos\theta_2=\frac{y_1}{\sqrt{y_1^2+y_2^2}},\quad \sin\theta_2=\frac{y_2}{\sqrt{y_1^2+y_2^2}}$$

因此 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的夹角余弦表示为

$$\cos\theta=\cos(\theta_1-\theta_2)=\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2=\frac{x_1y_1+x_2y_2}{\sqrt{x_1^2+x_2^2}\sqrt{y_1^2+y_2^2}}$$

由此得到内积的夹角（或夹角余弦）表达式为

$$\cos\theta=\frac{x_1y_1+x_2y_2}{\sqrt{x_1^2+x_2^2}\sqrt{y_1^2+y_2^2}}=\frac{(\boldsymbol{\alpha},\boldsymbol{\beta})}{|\boldsymbol{\alpha}||\boldsymbol{\beta}|} \qquad (4\text{-}10)$$

向量的长度和夹角这两个主要的概念可以用内积表示，说明内积是一个更一般的概念（双线性是比内积更加一般的概念，咱以后再提），一个线性空间如加了内积的定义，就不是一般的线性空间了。

欧式空间其实是特殊的线性空间，是定义了内积的线性空间，完整地讲是实数域 $\mathbf{R}$ 上具有内积的线性空间。从同构意义上讲，n 维欧式空间是以 $\mathbf{R}^n$ 为代表的。在一般的 n 维欧式空间 V 中，只要取一组标准正交基后，以 V 中向量在这组基下的坐标（即 $\mathbf{R}^n$ 中的向量）代替此向量来研究，就可以将 V 中的问题化为 $\mathbf{R}^n$ 中的问题来解决了。

2. 内积定义的推广

我们要好好地端详端详一下内积这个东东喽。

喂，内积有什么好讲的，它的定义不是 $\boldsymbol{a}\cdot\boldsymbol{b}=ab\cos\theta$ 吗？如果用坐标计算的话就是公式 $\boldsymbol{a}^{\mathrm{T}}\cdot\boldsymbol{b}=a_xb_x+a_yb_y+a_zb_z$ 嘛！前面向量一章中（2.3 节）不是说得很明白吗？！

喏，喏（作卑猥状）。看官，俺只是想在这里把内积的定义推广一下。

实际上，上述的内积定义是在一个比较标准的正交坐标系下定义的，我们下面试图在不同的基坐标系下定义同样一个内积的定义式。为啥这样做呢？吃饱了撑的？不全是。

我们常需要把向量看作像刚体一样在不同坐标系下具有不变性，即不变的长度和方向（夹角）；向量的长度和方向是用内积定义的，那么就要求向量也有不变的内积。另外，内积的物理意义是和做功联系在一起的。功是力在位移上的积累，是能量；能量守恒是基本的物理定律。伟大的导师爱因斯坦用相对论再一次教导我们说，物理定律是不随参考系（或坐标系）的不同而改变的。因此，我们就认为，欧式空间里的内积也是不应该随着不同基的不同而不同的。

但实际情况是，空间里的向量的坐标值都是定义在某一个基的上面的，基不同，向量的坐标值不同，显然两个向量的内积也是不同的。我们拿上面的例子来验算一下。

如图 4-31 所示，不用计算，我们就可以通过读图得到两个向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 分别在两个基下的坐标数组，并计算内积为

基 $\{\boldsymbol{\alpha}_1 0 \boldsymbol{\alpha}_2\}$ 下：$\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b} = (-2, 2)\begin{pmatrix} 1 \\ 3 \end{pmatrix} = 4$

基 $\{\boldsymbol{\beta}_1 0 \boldsymbol{\beta}_2\}$ 下：$\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b} = (2,1)\begin{pmatrix} 1 \\ -\dfrac{3}{2} \end{pmatrix} = \dfrac{1}{2}$

显然内积不同。内积之所以不同，直接与基的不同有关。如何就不同了呢？分析下。

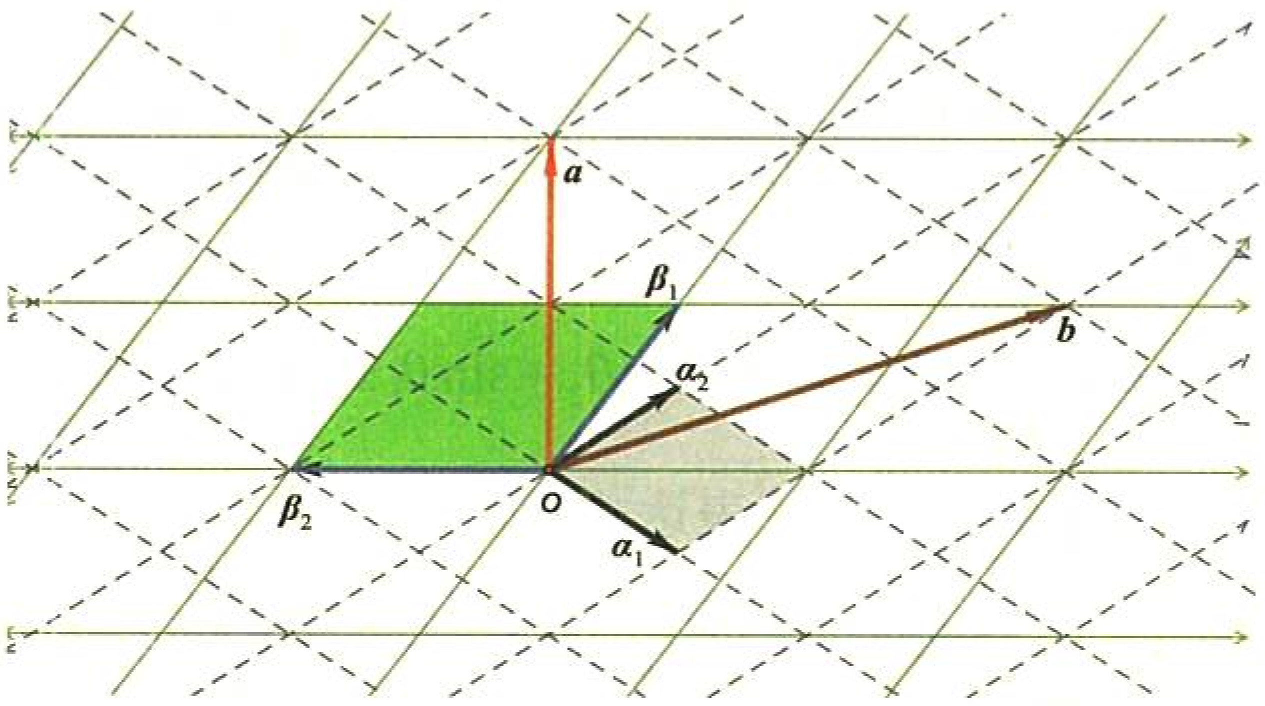


图 4-31 不同基下的向量内积

为了推导，我们把向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 在基 $\{\boldsymbol{\beta}_1 0 \boldsymbol{\beta}_2\}$ 下的对应的向量表示为 $\boldsymbol{c}$、$\boldsymbol{d}$，基 $\{\boldsymbol{\alpha}_1 0 \boldsymbol{\alpha}_2\}$ 到基 $\{\boldsymbol{\beta}_1 0 \boldsymbol{\beta}_2\}$ 的过渡矩阵为 $\boldsymbol{P}$。那么有

$$\boldsymbol{a} = \boldsymbol{P}\boldsymbol{c},\ \boldsymbol{b} = \boldsymbol{P}\boldsymbol{d}$$

则内积为

$$\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b} = (\boldsymbol{P}\boldsymbol{c})^{\mathrm{T}}(\boldsymbol{P}\boldsymbol{d}) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{d} = \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})\boldsymbol{d} \tag{4-11}$$

式（4-11）看明白没？提示一下，看等式的两头：$\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b}$ 是基 $\{\boldsymbol{\alpha}_1 0 \boldsymbol{\alpha}_2\}$ 下的内积，$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}$ 是第二个基 $\{\boldsymbol{\beta}_1 0 \boldsymbol{\beta}_2\}$ 下的内积，它们之间差一个乘积矩阵 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$。有了这个矩阵，两个内积就相等了，验证一下：

因为 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \begin{pmatrix} -\dfrac{1}{2} & \dfrac{3}{2} \\ -1 & -1 \end{pmatrix}\begin{pmatrix} -\dfrac{1}{2} & -1 \\ \dfrac{3}{2} & -1 \end{pmatrix} = \begin{pmatrix} \dfrac{5}{2} & -1 \\ -1 & 2 \end{pmatrix}$，所以 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{d} = (2,1)\begin{pmatrix} \dfrac{5}{2} & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -\dfrac{3}{2} \end{pmatrix} = 4$。

果然两个基下的内积都等于 4，没有变啊。

这个神奇的矩阵 $P^{\mathrm{T}}P$ 就是传说中的**内积度量矩阵**，它是基向量的逐个内积值所构成的矩阵，也称为 Gram 方阵，常用 S 表示，即 $S = P^{\mathrm{T}}P$。

由度量矩阵的定义式，你可以弄个具体的可逆方阵验算一下：**度量矩阵是一个正定的对称方阵**。

由此，如果知道了基向量，我们就可以这样定义内积：

任取向量空间 V 的一组基 M，将 V 中的每个向量用它们在此基下的坐标数组来表示；再任取一个 n 阶正定实对称方阵 S，对任意的向量 $\alpha, \beta \in V$，它们在基下 M 的坐标为 x、y，则有内积

$$(x, y) = x^{\mathrm{T}}Sy$$

虽然我们通过对任意的一组基 M 和任意的正定实对称方阵 S 来定义内积，但一旦内积定义好，有了长度和角度，就可以重新选择由两两垂直且长度为 1 的向量组成新基来替代基 M。这样的新基就是所谓的标准正交基，标准正交基的逐个内积刚好构成单位矩阵 E，也就是 $S = E$。那么在标准正交基下，内积的定义式就重新具有了早先的简单形式：

$$(x, y) = x^{\mathrm{T}}Sy = x^{\mathrm{T}}Ey = x^{\mathrm{T}}y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

这实际上就是大家习惯的直角坐标系及其内积定义。

为了方便解决实际问题，贴近直角坐标系下的物理空间，欧式线性空间里的基常常采用标准正交基。下面我们来讨论标准正交基概念的几何含义。

内积度量矩阵

真是神奇的矩阵，它解决了坐标系变换而保持内积计算值不变的问题。无论你变多少次基，每变一次基同时就改变一次度量矩阵，因此就保证了内积值的不变性，这样也就保证了向量空间度量的一致性，达到了度量不随坐标改变而改变的目的。度量矩阵就像爱因斯坦的尺子和钟表，时空坐标系（速度）在变，尺子变钟表也在变，就是光速不变。

4.2.6 标准正交基的几何解释

标准正交基也叫规范正交基。实际上，如果这些基向量互相垂直，就叫正交基，而且每个基向量的长度等于单位 1 的话，那么这个基就叫标准正交基。

1. 标准正交基的好处

为什么要得到标准正交基呢？主要原因是如果基是正交且标准的，就很容易计算任一向量到子空间的投影（坐标）。

对一空间 V，H 是它的 r 维子空间，即 $H \subset V$。若 $e_1, e_2, \cdots, e_r$ 是子空间 H 的一个标准正交基，则 V 中任一向量 α 到 H 的投影向量 α' 都能用标准正交基 $e_1, e_2, \cdots, e_r$ 线性表示，即

$$\alpha' = x_1e_1 + x_2e_2 + \cdots + x_ie_i + \cdots + x_re_r$$

其中 $(x_1, x_2, \cdots, x_r)$ 就是向量 α' 的坐标。

如果想求坐标 $(x_1, x_2, \cdots, x_r)$ 中一个 x_i，只要把向量 α 向对应的基向量 e_i 投影求其投影长度即可，求投影长度就是求内积，即公式 $\left(e_i^{\mathrm{T}}, \alpha\right) = x_i$。

图 4-32 给出了三维空间的一个二维子空间H（一个过原点平面），为H取一个不同于父空间基的标准正交基$\boldsymbol{e}_1$、$\boldsymbol{e}_2$。父空间任一向量$\boldsymbol{\alpha}$到子空间H的正交投影为$\boldsymbol{\alpha}'$，坐标为$(x_1,x_2)^{\mathrm{T}}$；其分别对基$\boldsymbol{e}_1$、$\boldsymbol{e}_2$坐标轴上的正交投影是$x_1\boldsymbol{e}_1$、$x_2\boldsymbol{e}_2$，那么坐标的计算方法就是$x_1=\boldsymbol{e}_1^{\mathrm{T}}\cdot\boldsymbol{\alpha}$，$x_2=\boldsymbol{e}_2^{\mathrm{T}}\cdot\boldsymbol{\alpha}$，或者写为矩阵方程的形式：

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{bmatrix}\boldsymbol{e}_1^{\mathrm{T}}\\\boldsymbol{e}_2^{\mathrm{T}}\end{bmatrix}\boldsymbol{\alpha}$$

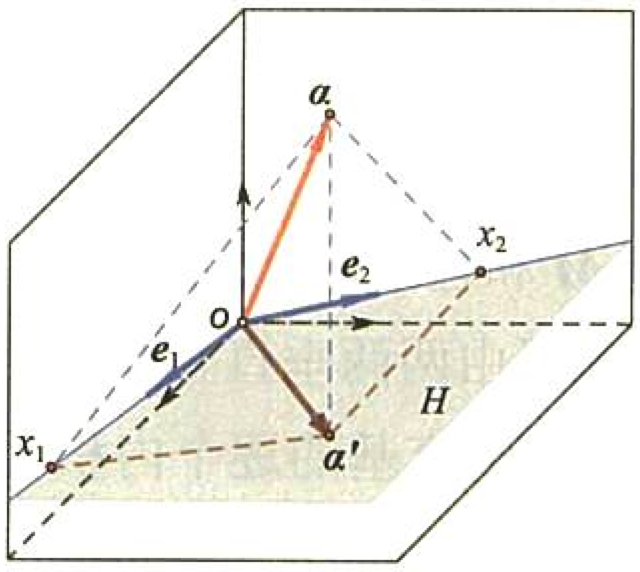


图 4-32 向量到子空间上的正交投影

坐标计算公式方便、简捷，是因为向量空间的基是标准正交基。所以，我们在给出向量空间的基时总是求取规范正交基。实际上，我们最常用的笛卡尔坐标系就是标准正交基的坐标系，只是在使用中忽略掉了x、y和z轴上的单位向量罢了。

2. 一维空间$\mathbf{R}^1$的标准正交基

一维$\mathbf{R}^1$的标准正交基只有两个：$\boldsymbol{i}=\{(1)\}$和$-\boldsymbol{i}=\{(-1)\}$。在x直线上的任意一个向量(x)都可以表示成为$\boldsymbol{i}$的倍数，其坐标是x；任意向量(x)也能表示成$-\boldsymbol{i}$的倍数，其坐标是$-x$。

图 4-33 一维空间的标准正交基

3. 二维空间$\mathbf{R}^2$的标准正交基

二维欧几里德空间$\mathbf{R}^2$的单位坐标向量组$\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\}$就是一个标准正交基。除了这个标准正交基外，还有其他的标准正交基吗？

有！比如把向量组内的向量位置交换一下：$\left\{\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right\}$，就是一个新的标准正交基。

还有！比如下面的向量组也是一个标准正交基：

$$\left\{\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)^{\mathrm{T}},\left(-\frac{1}{2},\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}\right\}$$

验证一下：

两个向量的内积：

$$\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)\begin{pmatrix}-\frac{1}{2}\\\frac{\sqrt{3}}{2}\end{pmatrix}=0$$

内积为 0，因此两个向量正交。

向量长度：

$$\left\|\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)\right\|=\sqrt{\left(\frac{\sqrt{3}}{2}\right)^2+\left(\frac{1}{2}\right)^2}=1，\quad \left\|\left(-\frac{1}{2},\frac{\sqrt{3}}{2}\right)\right\|=\sqrt{\left(-\frac{1}{2}\right)^2+\left(\frac{\sqrt{3}}{2}\right)^2}=1$$

长度为 1，单位长度，因此两个向量都是单位向量。

实际上，我们可以对二维笛卡尔单位坐标向量进行镜像或旋转来得到无数个包括上述的二维空间中的标准正交基。

二维平面的笛卡儿坐标空间中与坐标轴重合的单位向量有四个（参见图 4-34（a））：

$$\boldsymbol{i}=\begin{pmatrix}1\\0\end{pmatrix},\ \boldsymbol{j}=\begin{pmatrix}0\\1\end{pmatrix},\ -\boldsymbol{i}=\begin{pmatrix}-1\\0\end{pmatrix},\ -\boldsymbol{j}=\begin{pmatrix}0\\-1\end{pmatrix}$$

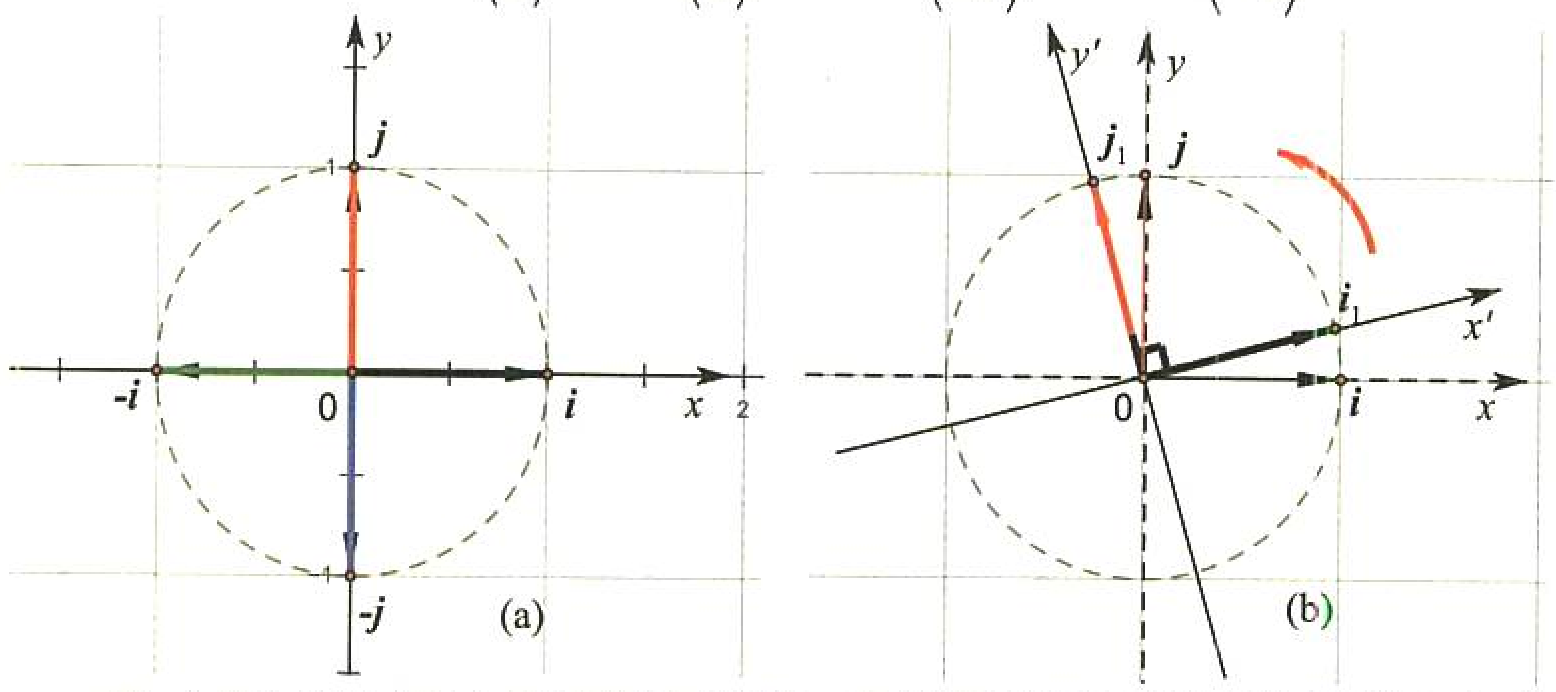


图 4-34 坐标轴上的标准正交基，旋转可得到任意多的标准正交基

其中，$\boldsymbol{i}$ 和 $-\boldsymbol{i}$ 互为镜像，$\boldsymbol{j}$ 和 $-\boldsymbol{j}$ 也互为镜像。这四组单位向量中的任意两个向量如果满足正交要求，那么这两个向量组成的有序向量组就构成一个标准正交基。所有的八个与坐标轴重合的标准正交基列举如下：

$$\{\boldsymbol{i},\boldsymbol{j}\},\ \{-\boldsymbol{i},\boldsymbol{j}\},\ \{\boldsymbol{i},-\boldsymbol{j}\},\ \{-\boldsymbol{i},-\boldsymbol{j}\},\ \{\boldsymbol{j},\boldsymbol{i}\},\ \{-\boldsymbol{j},\boldsymbol{i}\},\ \{\boldsymbol{j},-\boldsymbol{i}\},\ \{-\boldsymbol{j},-\boldsymbol{i}\}$$

另外，要得到不与坐标轴重合的标准正交基，我们可以把上述八个标准正交基的任意一组进行旋转就可以得到新的标准正交基。比如，把标准正交基 $\{\boldsymbol{i},\ \boldsymbol{j}\}$ 中的向量 $\boldsymbol{i}$ 和 $\boldsymbol{j}$ 同时进行逆时针旋转到 $\{\boldsymbol{i}_1,\boldsymbol{j}_1\}$（见图 4-34（b），保持垂直角度不变），那么 $\{\boldsymbol{i}_1,\boldsymbol{j}_1\}$ 也是一个标准正交基。如果旋转角等于 $\pi/6$，则这个标准正交基就是前边说的 $\left\{\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)^{\mathrm{T}},\left(-\frac{1}{2},\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}\right\}$。

实际上，如果用一个通用的向量表达式表示出二维空间 $\mathbf{R}^2$ 上的标准正交基就是 $\left\{\begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix},\begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix}\right\}$，其中 θ 为基向量逆向旋转的角度。这正是单位向量圆的参数表示式。

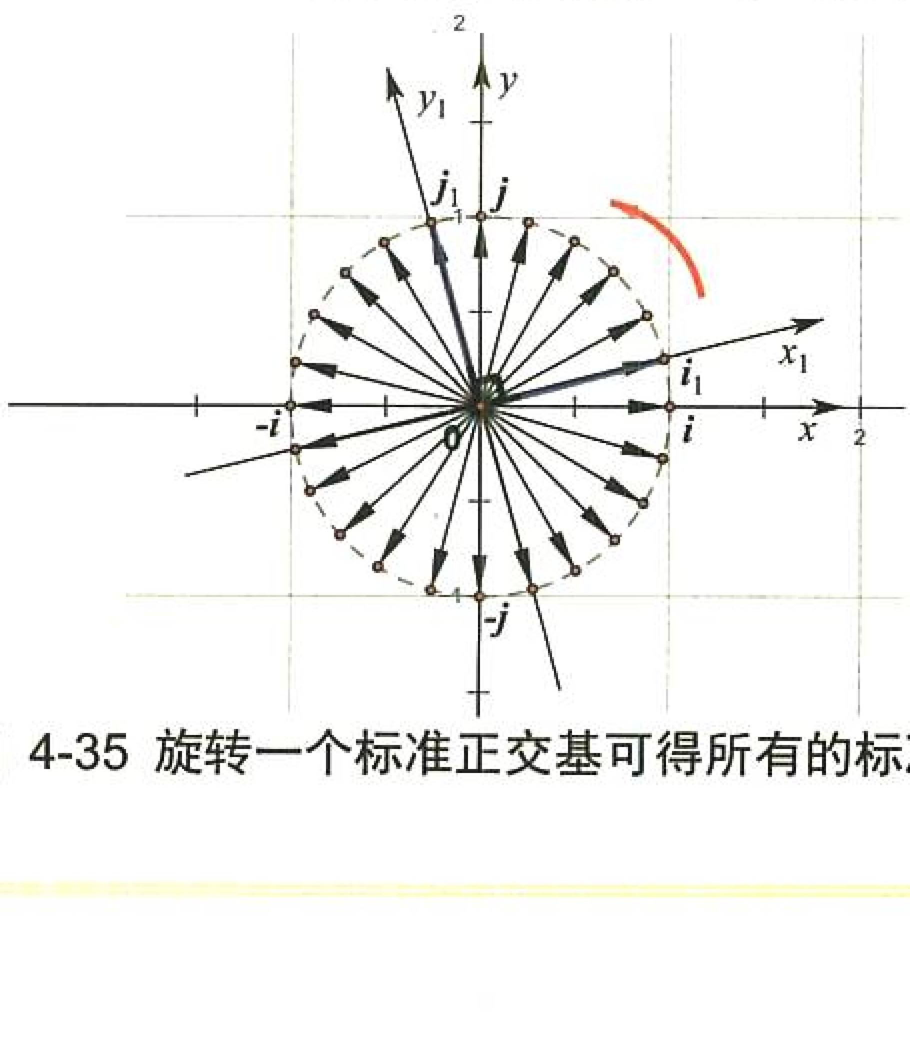


图 4-35 旋转一个标准正交基可得所有的标准正交基

对于上述的二维标准正交基的取法，我们有一个统一的几何图形（如图 4-35 所示），就是在把标准正交基 $\{\boldsymbol{i},\boldsymbol{j}\}$ 保持其角度垂直的同时进行逆时针或者顺时针旋转，即可得到全部的二维标准正交基。

4. 三维空间 $\mathbf{R}^3$ 标准正交基

类似的，三维空间 $\mathbf{R}^3$ 的标准正交基是由单位坐标向量组 $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ 在保持彼此的相对正交位置的同时进行全方位任意旋转，即可得到全部的标准正交基。显然，全部的标准正交基向量（无数的）的末端组成一个单位球面（见图 4-36）。

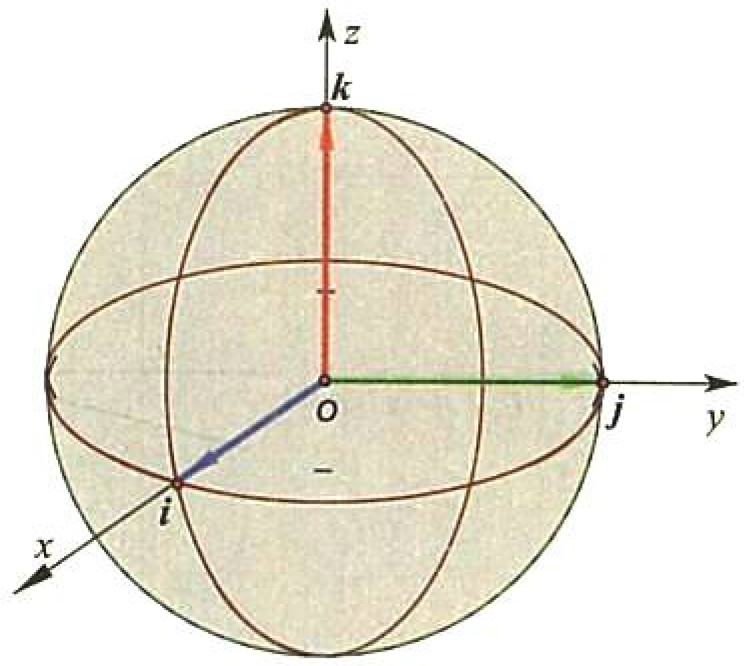


图 4-36 全部标准正交基构成一个单位球

使用球坐标系的定义式，我们可以得到所有的三维空间 $\mathbf{R}^3$ 的右手系标准正交基的数学表达式为

$$\left\{\begin{pmatrix}\sin\theta\cos\varphi\\ \sin\theta\sin\varphi\\ \cos\theta\end{pmatrix},\begin{pmatrix}\cos\theta\cos\varphi\\ \cos\theta\sin\varphi\\ -\sin\theta\end{pmatrix},\begin{pmatrix}-\sin\varphi\\ \cos\varphi\\ 0\end{pmatrix}\right\}$$

其中 $0\leqslant\theta\leqslant\pi$，$0\leqslant\varphi\leqslant 2\pi$。

下面我们给出以上结论导出过程的几何解释。首先看看球的半径为 r 的球坐标系的定义（见图 4-37）：

$$\begin{cases}x=r\sin\theta\cos\varphi\\ y=r\sin\theta\sin\varphi\\ z=r\cos\theta\end{cases}\qquad(0\leqslant\theta\leqslant\pi,\ 0\leqslant\varphi\leqslant 2\pi)$$

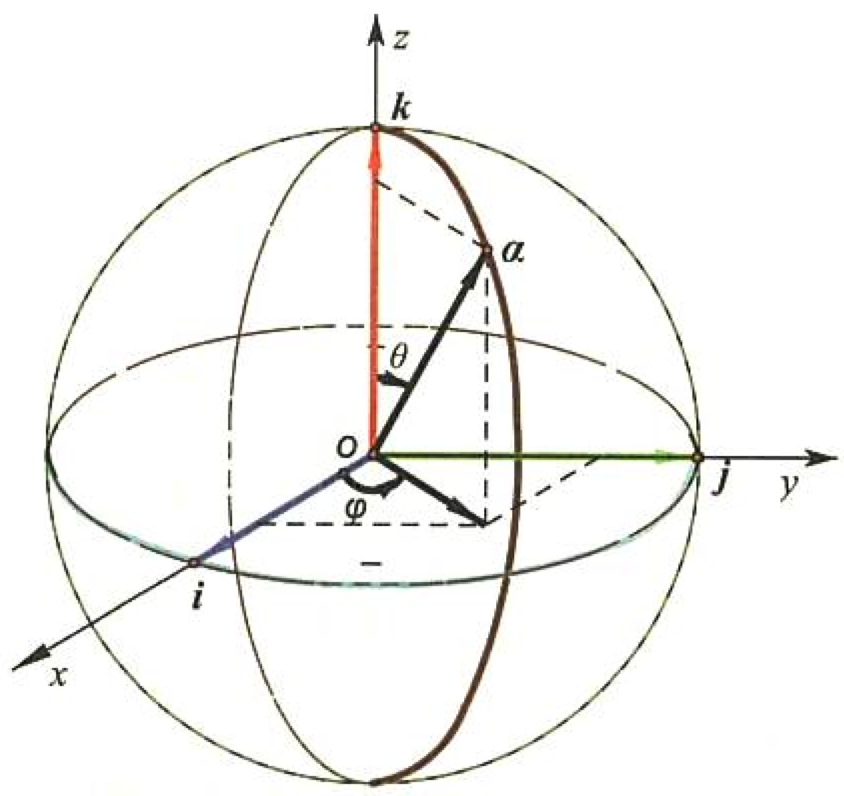


图 4-37 球坐标系的定义

实际上，球坐标系的定义式就是球面上任意一点的表达式。对于单位球来说，$r=1$。那么末端在球面上的任意一个向量$\boldsymbol{\alpha}$为

$$\boldsymbol{\alpha}=\begin{pmatrix}\sin\theta\cos\varphi\\ \sin\theta\sin\varphi\\ \cos\theta\end{pmatrix}$$

为了组合成一个正交基，我们还要找到另外两个向量。

第二个向量的找法是：从上面得到的一个任意向量$\boldsymbol{\alpha}$出发，在向量$\boldsymbol{\alpha}$和 z 轴所确定的平面上，沿着 θ 角的方向再继续旋转一个$\pi/2$角度，得到一个与$\boldsymbol{\alpha}$正交的单位向量，我们称之为向量$\boldsymbol{\beta}$（见图 4-38（a））。把 $\theta=\theta+\pi/2$ 代入向量$\boldsymbol{\alpha}$的表达式，即可得到向量$\boldsymbol{\beta}$的表达式，即

$$\boldsymbol{\beta}=\begin{pmatrix}\cos\theta\cos\varphi\\ \cos\theta\sin\varphi\\ -\sin\theta\end{pmatrix}$$

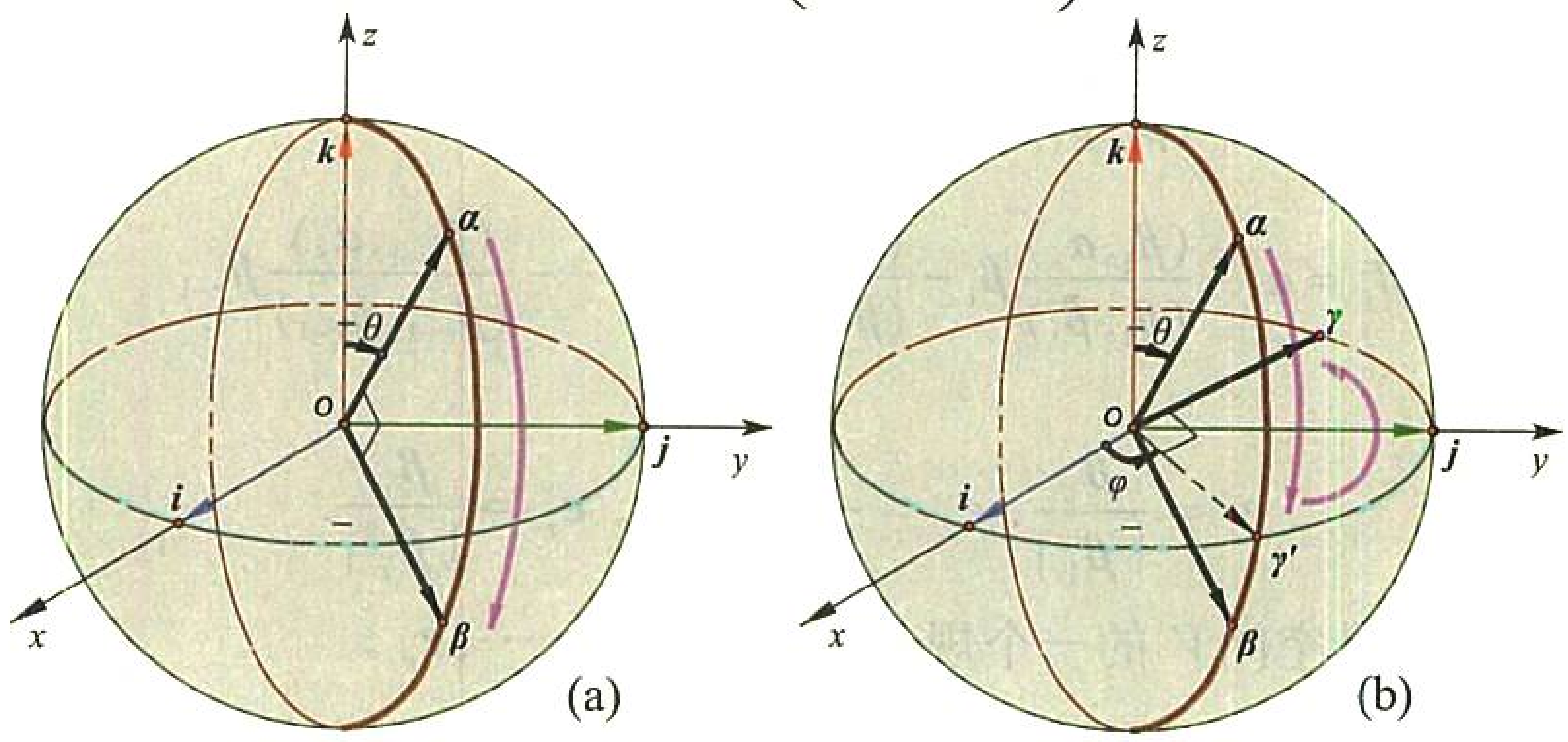


图 4-38 标准正交基表达式的图解

第三个基向量必须要与$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$正交才行。因为$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$在一个平面上，这个平面就是向量$\boldsymbol{\alpha}$和 z 坐标轴的单位向量$\boldsymbol{k}$所确定的平面。那么第三个向量只要与这个平面垂直就可以了。由此我们得到了第三个向量的求法：从向量$\boldsymbol{\alpha}$出发，在向量$\boldsymbol{\alpha}$和$\boldsymbol{k}$所确定的平面上，沿着θ角的方向再继续旋转一个$\pi/2-\theta$角度，得到一个与$\boldsymbol{k}$正交的单位向量，我们称之为向量$\boldsymbol{\gamma}'$（见图 4-38（b））。继续地，把向量$\boldsymbol{\gamma}'$在 xoy 平面沿着φ角方向旋转一个$\pi/2$角度，得到向量$\boldsymbol{\gamma}$。

向量$\boldsymbol{\gamma}$就是第三个单位正交基。因为$\boldsymbol{\gamma}$垂直于$\boldsymbol{k}$和$\boldsymbol{\gamma}'$，又因为$\boldsymbol{k}$、$\boldsymbol{\gamma}'$、$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$在同一个平面上，因此$\boldsymbol{\gamma}$正交于$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$。把 $\theta=\pi/2$ 和 $\varphi=\varphi+\pi/2$ 代入向量$\boldsymbol{\alpha}$的表达式，即可得到向量$\boldsymbol{\gamma}$的表达式，即

$$\boldsymbol{\gamma}=\begin{pmatrix}-\sin\varphi\\ \cos\varphi\\ 0\end{pmatrix}$$

至此，我们搞定了三维空间$\mathbf{R}^3$的右手系标准正交基为$\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$。

当然你也可以不用这么麻烦，直接叉积亦可得到$\boldsymbol{\gamma}$：$\boldsymbol{\gamma}=\boldsymbol{\alpha}\times\boldsymbol{\beta}$。另外，在得到$\boldsymbol{\gamma}$的过程中，如果把向量$\boldsymbol{\gamma}'$在 xoy 平面沿着φ角方向旋转一个$-\pi/2$角度，将得到向量$-\boldsymbol{\gamma}$，那么就可以得到左手系的标准正交基$\{\boldsymbol{\alpha},\boldsymbol{\beta},-\boldsymbol{\gamma}\}$。

对于高维空间的标准正交基，使用上述几何的方法就比较麻烦了。如果已知 n 维空间的一

个基，如何得到此空间的一个标准正交基呢？解决这个问题的一个著名方法是施密特正交化方法，下面我们将进行讨论。

4.2.7 施密特正交化的几何解释

设$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$是向量空间$V$的任意一个基，我们要在这组基上重新构造出一个新基，新的基要求是一个标准正交基。换句话说，要找一组两两正交的单位向量组$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_r$，这个向量组与原基$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$等价。把向量组$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$化为$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_r$的过程就是规范正交化。过程及公式如下：

第一步正交化（施密特正交化）：令

$$
\begin{aligned}
&\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1\\
&\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-\frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_2)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1\\
&\boldsymbol{\beta}_3=\boldsymbol{\alpha}_3-\frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1-\frac{(\boldsymbol{\beta}_2,\boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2\\
&\quad\vdots\\
&\boldsymbol{\beta}_r=\boldsymbol{\alpha}_r-\frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1-\frac{(\boldsymbol{\beta}_2,\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)}\boldsymbol{\beta}_2-\cdots-\frac{(\boldsymbol{\beta}_{r-1},\boldsymbol{\alpha}_r)}{(\boldsymbol{\beta}_{r-1},\boldsymbol{\beta}_{r-1})}\boldsymbol{\beta}_{r-1}
\end{aligned}
$$

第二步单位化：

$$\boldsymbol{e}_1=\frac{\boldsymbol{\beta}_1}{|\boldsymbol{\beta}_1|},\ \boldsymbol{e}_2=\frac{\boldsymbol{\beta}_2}{|\boldsymbol{\beta}_2|},\ \cdots,\ \boldsymbol{e}_r=\frac{\boldsymbol{\beta}_r}{|\boldsymbol{\beta}_r|}.$$

至此，我们得到了向量空间V的一个规范正交基$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_r$。

下面我们看看求取一至三维实向量空间基的规范正交化几何意义是什么？

1. 一维向量空间$\boldsymbol{R}^1$的规范正交基求法的几何意义

一维向量中只有一个元素a，设$\{(a)\}$是一维向量空间的一个基，显然，对这个向量只进行规范化即得到规范正交基：$\{(\frac{a}{|a|})\}$。图 4-39 中画出了负方向的(a)向量，向量(a)的长度被压缩到单位长度 1，方向保持不变，即规范正交基为$-\boldsymbol{i}$。

图 4-39 一维空间基的规范正交化

2. 二维向量空间$\mathbf{R}^2$的规范正交基求法的几何意义

从前面的分析知道，二维向量空间的规范基集中在平面坐标系的单位圆上。设$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\}$是二维向量空间上任意的一个基（见图 4-40（a）），设想由这个任意基通过施密特方法得到一个规范正交基，就是想办法用向量的伸缩和加减计算把向量$\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$变化到单位圆上得到新向量$\boldsymbol{e}_1$、$\boldsymbol{e}_2$，而且新向量$\boldsymbol{e}_1$、$\boldsymbol{e}_2$要互相垂直。

第一步$\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1$，这个就是向量改个名，没什么意义；

施密特正交化公式的第二步及其变形为

$$\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-\frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_2)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1=\boldsymbol{\alpha}_2-\frac{(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_2)}{|\boldsymbol{\beta}_1|^2}\boldsymbol{\beta}_1=\boldsymbol{\alpha}_2-\left(\frac{\boldsymbol{\beta}_1}{|\boldsymbol{\beta}_1|},\boldsymbol{\alpha}_2\right)\frac{\boldsymbol{\beta}_1}{|\boldsymbol{\beta}_1|}=\boldsymbol{\alpha}_2-(\boldsymbol{e}_1,\boldsymbol{\alpha}_2)\boldsymbol{e}_1$$

在上式中，$(\boldsymbol{e}_1,\boldsymbol{\alpha}_2)$ 是单位向量 $\boldsymbol{e}_1$ 和 $\boldsymbol{\alpha}_2$ 的内积，即求向量 $\boldsymbol{\alpha}_2$ 在 $\boldsymbol{e}_1$ 上的坐标值，因而 $(\boldsymbol{e}_1,\boldsymbol{\alpha}_2)\boldsymbol{e}_1$ 是向量 $\boldsymbol{\alpha}_2$ 在 $\boldsymbol{e}_1$ 上投影的分向量。向量 $\boldsymbol{\alpha}_2$ 减去自己在 $\boldsymbol{e}_1$ 上这个投影，则得到一个正交于向量 $\boldsymbol{e}_1$ 的分向量 $\boldsymbol{\beta}_2$（见图 4-40（b））。

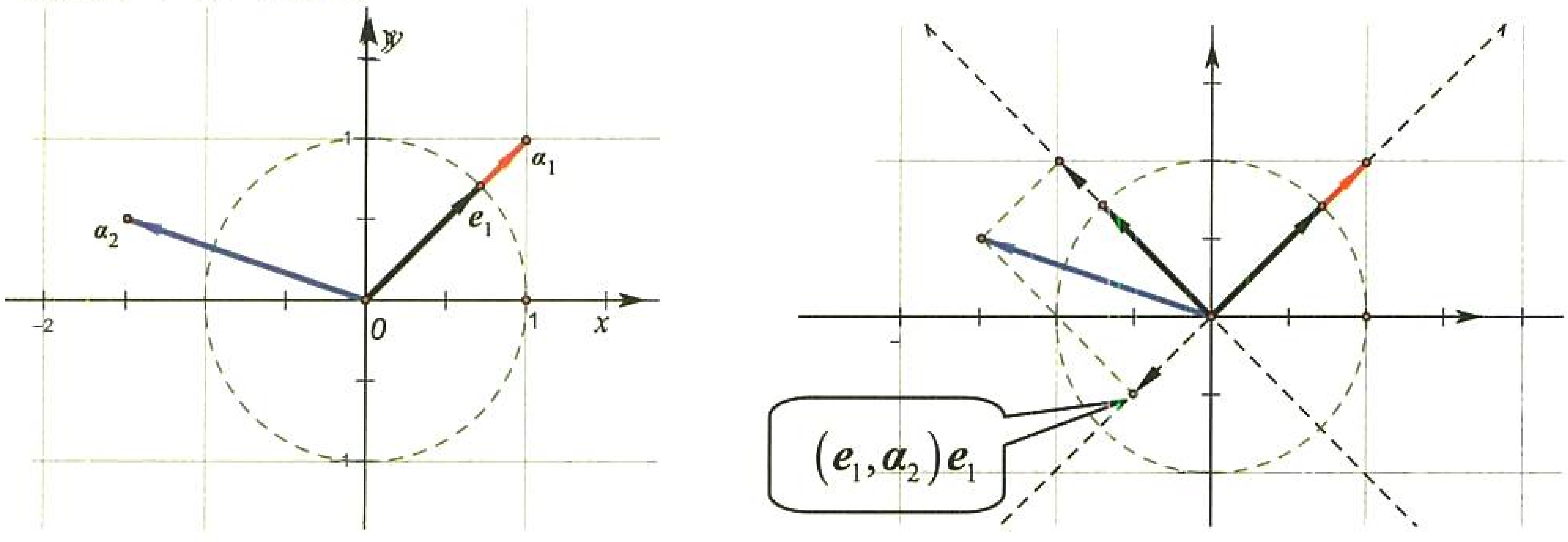


图 4-40 二维空间基的规范正交化

第三步对正交向量 $\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\beta}_2$ 单位化，得到位于单位圆上的标准正交基 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$。

3. 三维向量空间 $\mathbf{R}^3$ 的规范正交基求法的几何意义

类似地，三维向量空间基的规范正交化过程是二维上的推广。对于一个基 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$（见图 4-41（a）），我们先给出施密特正交化的变形公式：

$$\begin{aligned}\boldsymbol{\beta}_1&=\boldsymbol{\alpha}_1\\ \boldsymbol{\beta}_2&=\boldsymbol{\alpha}_2-(\boldsymbol{e}_1,\boldsymbol{\alpha}_2)\boldsymbol{e}_1\\ \boldsymbol{\beta}_3&=\boldsymbol{\alpha}_3-(\boldsymbol{e}_1,\boldsymbol{\alpha}_3)\boldsymbol{e}_1-(\boldsymbol{e}_2,\boldsymbol{\alpha}_3)\boldsymbol{e}_2\end{aligned}$$

好，前面两个公式是二维空间的公式，我们已经讨论过。根据上节的方法我们假设得到了 $\{\boldsymbol{e}_1,\boldsymbol{e}_2\}$，如图 4-41（b）所示。

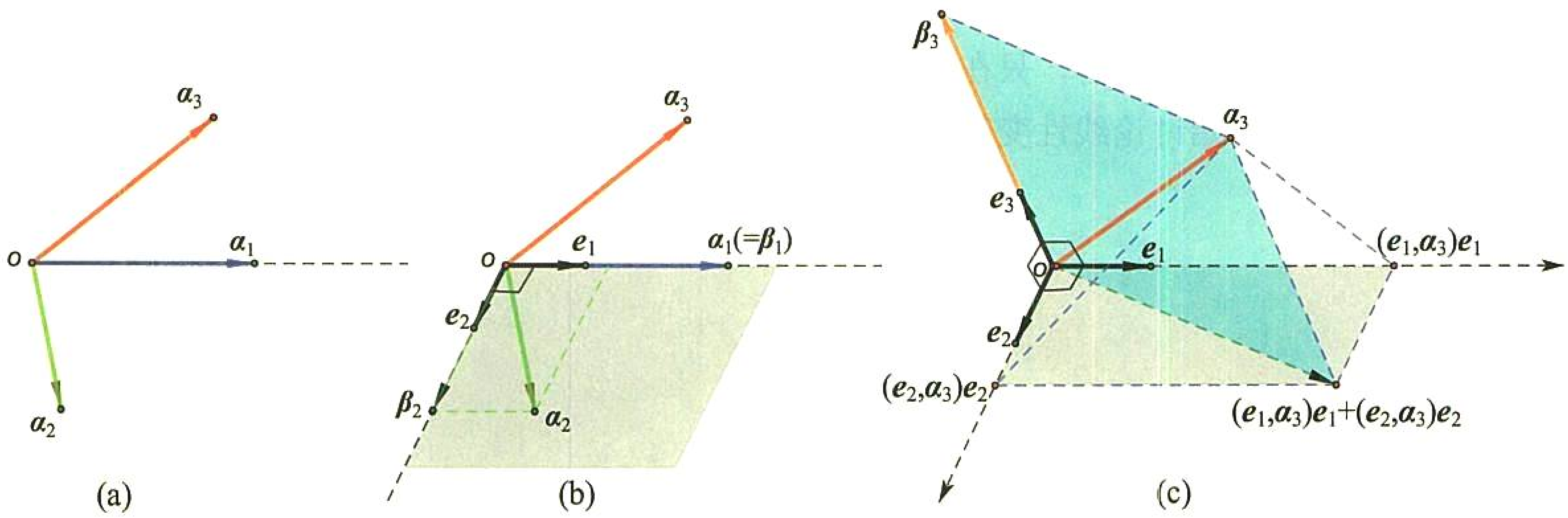


图 4-41 三维空间基的规范正交化图

新增的第三个公式是求取 $\boldsymbol{\beta}_3$，由式子的几何含义，我们不难看出：向量 $\boldsymbol{\alpha}_3$ 减去自己在二维正交规范基上的投影得到 $\boldsymbol{\beta}_3$。其中 $(\boldsymbol{e}_1,\boldsymbol{\alpha}_3)\boldsymbol{e}_1$ 是 $\boldsymbol{\alpha}_3$ 在 $\boldsymbol{e}_1$ 上的投影，$(\boldsymbol{e}_2,\boldsymbol{\alpha}_3)\boldsymbol{e}_2$ 是 $\boldsymbol{\alpha}_3$ 在 $\boldsymbol{e}_2$ 上的投影。两者的和 $(\boldsymbol{e}_1,\boldsymbol{\alpha}_3)\boldsymbol{e}_1+(\boldsymbol{e}_2,\boldsymbol{\alpha}_3)\boldsymbol{e}_2$ 就是 $\boldsymbol{\alpha}_3$ 在以 $\{\boldsymbol{e}_1,\boldsymbol{e}_2\}$ 为基的子空间（平面）上的投影。

这里，向量α_3减去自己在e_1上的投影的差向量必垂直于e_1，同时也减去自己在e_2上的投影的差向量亦必垂直于e_2，因而总的差向量$\alpha_3-[(e_1,\alpha_3)e_1+(e_2,\alpha_3)e_2]$即$\beta_3$垂直于$e_1$和$e_2$张成的平面空间（见图 4-41（c））。

对β_3再进行单位化得到了第三个基向量e_3。为清楚起见，图中没有画出α_1、α_2等向量。稍细心些，我们将发现，原基为左手系，规范化后的新基仍为左手系。

4. r维向量空间的规范正交基求法的几何意义

至此，多维向量空间的规范正交几何意义显而易见。求取第 n 个正交基的方法就是把第 n 个基向量分别减去在先前 n-1 个正交基向量上的投影，从而得到正交于全部 n-1 个正交基的新基向量。从几何作图方便性的角度出发，我们建议把施密特正交规范化公式修改一下，不再需要过渡向量$\beta_1,\beta_2,\cdots,\beta_r$，公式如下：

$$
\begin{aligned}
e_1 &= \frac{\alpha_1}{|\alpha_1|}\\
e_2 &= \frac{\alpha_2-(e_1,\alpha_2)e_1}{|\alpha_2-(e_1,\alpha_2)e_1|}\\
e_3 &= \frac{\alpha_3-(e_1,\alpha_3)e_1-(e_2,\alpha_3)e_2}{|\alpha_3-(e_1,\alpha_3)e_1-(e_2,\alpha_3)e_2|}\\
&\vdots\\
e_r &= \frac{\alpha_r-(e_1,\alpha_r)e_1-(e_2,\alpha_r)e_2-\cdots-(e_{r-1},\alpha_r)e_{r-1}}{|\alpha_r-(e_1,\alpha_r)e_1-(e_2,\alpha_r)e_2-\cdots-(e_{r-1},\alpha_r)e_{r-1}|}
\end{aligned}
$$

上式看起来比较复杂，其实概念清晰自然，计算方便。

到此为止，众多的向量们终于有了家——向量空间。在这个空间里，我们已经给子民们规定了宪法——基，树立了典范——标准正交基，给每个子民定义了身份证号——坐标，并规定社会团体——子空间也同样遵守国家宪法。好了国家有了大法，社会有了秩序，下面就看子民们如何玩了，玩就是映射或变换，因此，大法下面就是刑法——保持线性；各个向量在遵守刑法下全国各地的玩就是线性映射，只在省内玩的就是线性变换。

下面的章节我们开始讨论线性变换的具体表现形式——矩阵是如何玩的。

第5章 矩阵的几何意义

前面章节除了向量空间的基变换用到了矩阵概念外，其余的章节基本没提矩阵，憋到现在咱终于进入了线性代数的重要内容：矩阵。矩阵是线性代数中的主要研究对象，它作为一种新型的数学表示工具，正在或者已经成为现代公民的常识——掌握是必须的，因此本章内容的篇幅较大。

前面说过，矩阵是“比例函数”概念的推广，是描述向量之间变换关系的（见1.2.3节 ）。如果说，比例函数的系数是“数”与“数”之间的线性对应关系，是把一个数变成另一个数，那么矩阵则是向量与向量之间的线性对应关系，是把一个向量变成另一个向量。

当然，矩阵把一个向量变成另一个向量是发生在向量空间里的变换运动，该变换有个专业名词叫线性变换或线性映射。这可以称为矩阵的几何意义。与行列式类似，我们仍然主要使用向量来解释矩阵，矩阵是由向量组成的，矩阵是对向量进行作用的等。因为没有向量我们就无从谈起几何意义，矩阵的几何意义就是矩阵的向量意义。

矩阵独立的几何意义表现为对向量的作用结果。矩阵对一个向量是如何作用的？矩阵对多个向量是如何作用的？矩阵对一个几何图形（由无数向量组成的几何图形）是如何作用的？矩阵对空间上的坐标基向量又是如何作用的？在矩阵对一个几何图形的变换研究中，我们会发现一些有规律的东西，比如特征向量、秩等。在矩阵对一个基的变换研究中，我们也会发现更高等的东西，比如和微分几何有深刻联系的雅可比矩阵的几何意义。

一个矩阵就描述了向量空间中的一个运动——变换，这个矩阵规定了所有向量的变换规则。这么重要的矩阵概念是如何产生的呢？

通过前面的章节我们初步了解到，解线性方程组的克莱姆法则使用了行列式理论，但克莱姆法则只能用于解方程个数等于未知数个数的方程组，而且系数行列式不能等于0。即使以上条件都满足，也要计算$n+1$个n阶行列式。实际工程中的n一般很大（比如$n=1000$），即使在现代计算机技术面前，计算效率也不能使人满意。

那就用消元法解。在用消元法解各种类型的线性方程组时，一系列问题出现了：当系数行列式等于0时 ，方程组是否有解？若有解又如何求出？当未知量个数与方程的个数不等时，线性方程组的解又如何？

要深入探讨这些问题，除了向量概念外还必须引入矩阵的理论。到了1858年，哈密尔顿和凯莱的著作中出现了矩阵的运算，从行列式到矩阵的出现，大约经过了100年的时间。

5.1 矩阵的概念及物理意义

矩阵的外观就是一个长方形的数表，生活中的一些长方形的数表也可以看成是矩阵。矩阵和行列式相似，也是以行和列组织的矩形数字阵列，因此称为矩阵。与行列式的绝对值表示法

不同，矩阵是用方括号把数字块括起来，表示一个有顺序有组织的数据块；而行列式是对这些数据块进行的一个运算，是一个算式，故称为行列“式”。矩阵的一个三阶例子如下：

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$$

如果用数组来统一定义标量、向量和矩阵的话，就是：标量是一维向量，向量是标量的数组，矩阵则是向量的数组。例如，上面介绍的矩阵可以使用列向量 $\boldsymbol{a}=(a_1,a_2,a_3)^{\mathrm{T}}$，$\boldsymbol{b}=(b_1,b_2,b_3)^{\mathrm{T}}$，$\boldsymbol{c}=(c_1,c_2,c_3)^{\mathrm{T}}$ 来表示，写做数组的形式 $[\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}]$。

其实，对矩阵概念的理解可以从几个方面或层次入手。初步的层次就是认为矩阵是一数表，无它；深入点的层次是矩阵也可以通过分块看作是一向量，在矩阵的乘法运算中符合向量点积的运算规律；更深入的层次是先前提过的：把矩阵看做线性变换或者说向量的比例函数的映射系数，这个有清晰的几何意义，也是本章论述的中心；离线性代数课程远点的层次就是把矩阵看做由多个向量构成的一个新的数学量，又称为二阶张量。矩阵是张量的概念这个咱不作讨论，有兴趣的读者可以看张量分析方面的资料。

为感性理解，下面举两个例子说说矩阵是数表和映射系数两个概念的理解。

5.1.1 矩阵是统计数表的例子

比如某家电器公司的制造厂有几个生产线，生产线在2009年和2010年的上半年的产出量的统计表如表 5-1和表 5-2所示。

表 5-1　2009 年上半年的产出量的统计表

顺序	生产线名	2009 年上半年的每月产出量					
		1 月	2 月	3 月	4 月	5 月	6 月
1	冰箱线	22	35	30	23	25	12
2	吸尘器线	25	43	32	34	35	30
3	电视线	23	23	34	44	40	45

表 5-2　2010 年上半年的产出量的统计表

顺序	生产线名	2010 年上半年的每月产出量					
		1 月	2 月	3 月	4 月	5 月	6 月
1	冰箱线	22	34	30	23	25	12
2	吸尘器线	24	43	32	34	35	34
3	电视线	23	23	34	45	41	45
4	手机线	34	34	35	45	23	43
5	平板电脑线	45	24	31	34	45	12

我们将表 5-1对应的矩阵记为 $\boldsymbol{A}$，表 5-2对应的矩阵记为 $\boldsymbol{B}$，则有

$$\boldsymbol{A}=\begin{bmatrix} 22 & 35 & 30 & 23 & 25 & 12 \\ 25 & 43 & 32 & 34 & 35 & 30 \\ 23 & 23 & 34 & 44 & 40 & 45 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{和} \quad \boldsymbol{B}=\begin{bmatrix} 22 & 34 & 30 & 23 & 25 & 12 \\ 24 & 43 & 32 & 34 & 35 & 34 \\ 23 & 23 & 34 & 45 & 41 & 45 \\ 34 & 34 & 35 & 45 & 23 & 43 \\ 45 & 24 & 31 & 34 & 45 & 12 \end{bmatrix}$$

于是

$\boldsymbol{A}+\boldsymbol{B}$ 的实际意义为2009、2010年1~6月各生产线每月产量的和(2009年手机，平板电脑的产量为0)；

$\boldsymbol{B}-\boldsymbol{A}$ 的实际意义为2010年1~6月各生产线每月产量比上年同期的增产情况。

设冰箱、吸尘器、电视机、手机、平板电脑的价格分别为1500元/台、900元/台、3000元/台、2500元/台、5000元/台，则有单价向量 $\boldsymbol{c}=(1500,900,3000,2500,5000)$，那么矩阵乘法，即 $\boldsymbol{cB}$ 的实际意义为2010年1~6月该制造厂每种产品的月产值。

如果每生产一件产品职工得到的奖励积分为3分，则数乘运算 $3\boldsymbol{A}$ 的实际意义为2009年1~6月该厂各车间的职工月积分。

5.1.2 矩阵是线性函数系数的例子　例子

矩阵是个数表，这认识是静态的概念，无法深入展开，必须进入函数系数以至线性映射的概念。

矩阵是“比例函数”概念的推广，是向量与向量之间的线性对应关系，是向量的比例函数的系数。看起来和听起来似乎是那么回事，但如果没个例子的话，总有一点存疑在那里。买房差钱不？差钱咱来个化石成金的例子哈。例子哈

例5.1 话说甲类金矿石一吨可以炼出a克黄金和b克白银，乙类金矿石一吨可以炼出c克黄金和d克白银。您就是矿主，今开采了甲类矿石 x_1 吨，乙类矿石 x_2 吨。请问矿主能炼出多少克真金和白银？

小儿科，好解。设能炼出 y_1 克黄金和 y_2 克白银，有以下的线性方程组：

$$\begin{cases} ax_1+cx_2=y_1 \\ bx_1+dx_2=y_2 \end{cases}$$

方程

将其改写成向量和矩阵的形式为

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{5-1}$$

咱们来理解下式（5-1）里面的向量和矩阵实际代表的物资的意义，也可算做物理意义吧。

甲类、乙类矿石都是矿石，两者的质量合成为一个二维列向量$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$，可称之为土石向量 $\boldsymbol{x}$；黄金、白银都是钱啊，两者的质量也合成为一个二维列向量$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$，可称之为金银向量 $\boldsymbol{y}$。土石的质量和金银的质量之间有个函数关系，这个关系也可以被认为是一个“正比例”关系（即式（5-1））。$\boldsymbol{x}$ 吨的土石可以炼出 $\boldsymbol{y}$ 克的金银；$2\boldsymbol{x}$ 吨的土石可以炼出$2\boldsymbol{y}$ 克的金银。矩阵 $\boldsymbol{A}=\begin{bmatrix} a & c \\ b & d \end{bmatrix}$ 刚好是这个“复合比例”函数的系数。“复合比例”的意思是，矩阵是两种矿石和两种贵重金属的合一起的比例函数。一点也不奇怪，像一次比例线性函数一样，式（5-1）也被称为线性函数。

既然矩阵可称为线性函数的比例系数，那么它是否和一次比例函数系数的物理意义一样呢？一样的。我们来看：

一次比例函数（亦可称为一维向量的比例函数）$y=kx$ 的系数 k 的物理意义比较容易理解。

例如物体的匀速直线运动函数$s=vt$，速度v就是比例系数。这个比例系数是什么呢？它是当自变量时间t等于单位时间1秒时的函数值，即$s=v$。换句话说，系数就是物体走单位时间1秒所通过的距离值。

那么这个二维向量的比例函数$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$的系数$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$的物理意义是否作同样的理解？是的，它也是当自变向量$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$取单位向量时得到的函数值，这个单位向量有两个：$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$,$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$。具体说来，1吨甲类矿石和0吨乙类矿石可得到$a$克黄金和$b$克白银，即有$\begin{pmatrix} 1 \\ 0 \end{pmatrix}\to\begin{pmatrix} a \\ b \end{pmatrix}$；0吨甲类矿石和1吨乙类矿石可得到$c$克黄金和$d$克白银，即有$\begin{pmatrix} 0 \\ 1 \end{pmatrix}\to\begin{pmatrix} c \\ d \end{pmatrix}$。我们把得到的两个金银向量组合成比例系数即矩阵$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$。

换句话说，矩阵就是自变向量取单位向量组所得到因变向量组值的有序组合。

如果从映射的运动角度理解，$\boldsymbol{x}$吨的土石变换出来$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$克的金银；矩阵$\boldsymbol{A}$正是变换的决定因素，它是点石成金的仙人指哦（由映射或变换的角度解释矩阵的几何及物理意义详见5.5.2节）！

好，矩阵虚的概念介绍过了。下面的章节咱开始来点实在的，具体考察矩阵的各类运算及其意义。

5.2 矩阵加法的几何意义

矩阵的加法和乘法等简单运算可看做源自于线性方程组的简单运算，读者可以参看6.8.1节的详细介绍。在下面介绍矩阵加法和乘法的几何意义时，我们仍然不能离开向量的有力帮助。矩阵中的行向量或列向量的意义可以有效地帮助我们看清矩阵所蕴含的几何变换的意义。

多个矩阵的加法比较简单，即使不用给出几何意义，我们也能轻松掌握它。不过画出矩阵加法的几何图形，可以帮助我们对多个向量所组成的几何图形的叠加有个形象的认知。

图 5-1显示了三个三阶矩阵连加的过程。三阶矩阵有三个行向量，三个矩阵连加相当于三组向量同时连加，结果得到的三个向量组成了矩阵的和。

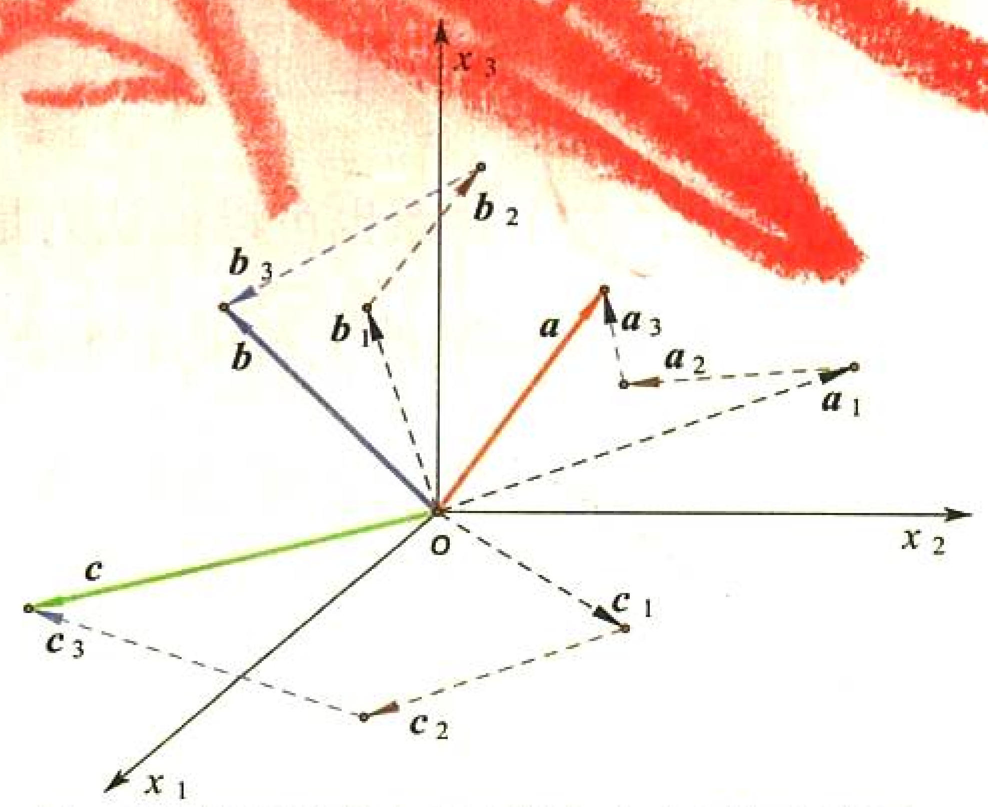


图 5-1 矩阵相加可看做多个向量同时相加

三阶方矩阵的连加表述出来就是：

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ b_{11} & b_{12} & b_{13} \\ c_{11} & c_{12} & c_{13} \end{bmatrix} + \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ b_{21} & b_{22} & b_{23} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} + \begin{bmatrix} a_{31} & a_{32} & a_{33} \\ b_{31} & b_{32} & b_{33} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

简写为行向量为

$$\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{b}_1 \\ \boldsymbol{c}_1 \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_2 \\ \boldsymbol{b}_2 \\ \boldsymbol{c}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_3 \\ \boldsymbol{b}_3 \\ \boldsymbol{c}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix}$$

上面是把每个矩阵都分解为三个行向量来给出图形的，其实因为矩阵的加法是对每个元素分别对应相加，因此对于列向量同样等效。

5.3　矩阵与向量乘法的几何意义

矩阵与向量乘积比如 $\boldsymbol{Ax}$ 表现为矩阵 $\boldsymbol{A}$ 对一个向量 $\boldsymbol{x}$ 作用的结果。其作用的主要过程是对一个向量进行旋转和缩放的综合过程（即线性变换的过程），一个向量就变换为另一个向量。一个 m 行 n 列的实矩阵 $\boldsymbol{A}_{m\times n}$ 就是一个 $\mathbf{R}^n \to \mathbf{R}^m$ 上的线性变换。或者说，矩阵 $\boldsymbol{A}_{m\times n}$ 把一个 n 维空间的 n 维向量变换为一个 m 维空间的 m 维向量。

在矩阵与向量、矩阵与矩阵的乘法过程中，常常通过分块技术可以把矩阵看作一个向量，而向量则有时蜕化为数，其各层次的乘法都符合向量点积的运算规律。

5.3.1　矩阵与向量的乘积的概念

矩阵 $\boldsymbol{A}$ 与向量 $\boldsymbol{c}$ 的乘积（矩阵左乘向量，记为 $\boldsymbol{Ac}$ ）是一个向量，这个向量的每个分量是以矩阵 $\boldsymbol{A}$ 的每个行向量分别与列向量 $\boldsymbol{c}$ 的数量积作为元素的。乘法公式如下：

$$\boldsymbol{A}\cdot\boldsymbol{c} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} (a_1\ \ a_2\ \ a_3)\cdot\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \\ (b_1\ \ b_2\ \ b_3)\cdot\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} a_1c_1 + a_2c_2 + a_3c_3 \\ b_1c_1 + b_2c_2 + b_3c_3 \end{pmatrix} \quad (5\text{-}2)$$

上式 $\boldsymbol{Ac}$ 的乘积是把矩阵 $\boldsymbol{A}$ 看作两个行向量，实质上还是向量与向量的点乘积。

类似地，向量 $\boldsymbol{c}$ 与矩阵 $\boldsymbol{A}$ 的乘积（矩阵右乘向量，记为 $\boldsymbol{cA}$ ）也是一个向量，这个 $\boldsymbol{cA}$ 向量的每个分量是行向量 $\boldsymbol{c}$ 与矩阵 $\boldsymbol{A}$ 的每个列向量的点乘积。乘法公式如下：

$$\boldsymbol{cA} = (c_1\ \ c_2\ \ c_3)\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix} = \left((c_1\ \ c_2\ \ c_3)\cdot\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad (c_1\ \ c_2\ \ c_3)\cdot\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right) = (c_1a_1 + c_2a_2 + c_3a_3 \quad c_1b_1 + c_2b_2 + c_3b_3)$$

以上的乘积运算都是“行向量·列向量”的形式。下面我们换一下思维方式，把乘积运算看成“列向量·行向量”的形式是否说得通？先从 $\boldsymbol{Ac}$ 的乘积开始。

如果把矩阵 $\boldsymbol{A}$ 分解为3个列向量，我们可以这样展开式（5-2）：

$$\boldsymbol{A}\cdot\boldsymbol{c} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}c_1 + \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}c_2 + \begin{pmatrix} a_3 \\ b_3 \end{pmatrix}c_3 = \begin{pmatrix} a_1c_1 + a_2c_2 + a_3c_3 \\ b_1c_1 + b_2c_2 + b_3c_3 \end{pmatrix} \quad (5\text{-}3)$$

式）的含义是 $\boldsymbol{Ac}$ 的乘积可以理解为矩阵 $\boldsymbol{A}$ 的列向量的线性组合，组合系数是向量 $\boldsymbol{c}$ 的三个分量。这个展开的实质仍然是向量点乘积的乘法。

再看 $\boldsymbol{c}\cdot\boldsymbol{A}$ 的展开：

$$\boldsymbol{c}\cdot\boldsymbol{A}=(c_1\ c_2\ c_3)\begin{bmatrix}a_1 & b_1\\ a_2 & b_2\\ a_3 & b_3\end{bmatrix}=c_1(a_1\ \ b_1)+c_2(a_2\ \ b_2)+c_3(a_3\ \ b_3)=(c_1a_1+c_2a_2+c_3a_3\quad c_1b_1+c_2b_2+c_3b_3)$$

这个式子可以理解为矩阵 $\boldsymbol{A}$ 的行向量的线性组合，组合系数是向量 $\boldsymbol{c}$ 的三个分量。

实际上，在以上的各种乘法中，我们使用了矩阵和向量的分块技术（全部是“行向量 · 列向量”的形式），每一个分块都要看成是矩阵最基本的元素“数”进行运算。至此，我们较全面地理解了向量与矩阵乘积的展开实质。显然，左乘与右乘的结果是不同的。为什么不同，答案就在随后的章节里。

5.3.2 矩阵与向量乘积的几何意义

为了更具体地观察矩阵和向量乘积的几何意义，我们下面先考察欧式空间中一个矩阵与单位坐标向量的乘积的过程，然后再看一个矩阵与任意向量的乘积的几何意义。

1. 矩阵与单位坐标向量的乘积的几何解释

三维的单位坐标向量就是 $\boldsymbol{i}=(1,0,0)$，$\boldsymbol{j}=(0,1,0)$，$\boldsymbol{k}=(0,0,1)$。我们取 x 坐标轴上的单位向量 $\boldsymbol{i}$ 与 $\boldsymbol{A}=\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}$ 相乘，得乘式如下：

$$\boldsymbol{iA}=(1,0,0)\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}=(a_1\ \ a_2\ \ a_3),\qquad \boldsymbol{Ai}^{\mathrm{T}}=\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}a_1\\b_1\\c_1\end{pmatrix}$$

$\boldsymbol{iA}$ 的结果是 $(a_1\ \ a_2\ \ a_3)$，这恰是矩阵 $\boldsymbol{A}$ 的第一行。$\boldsymbol{Ai}^{\mathrm{T}}$ 的结果是 $\begin{pmatrix}a_1\\b_1\\c_1\end{pmatrix}$，这恰是矩阵 $\boldsymbol{A}$ 的第一列。

为了更明了，下面把 $\boldsymbol{j}$ 和 $\boldsymbol{k}$ 与矩阵 $\boldsymbol{A}$ 的乘积也一并列出来：

$$\boldsymbol{jA}=(0,1,0)\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}=(b_1\ \ b_2\ \ b_3),\qquad \boldsymbol{kA}=(0,0,1)\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}=(c_1\ \ c_2\ \ c_3)$$

$$\boldsymbol{Aj}^{\mathrm{T}}=\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}a_2\\b_2\\c_2\end{pmatrix},\qquad \boldsymbol{Ak}^{\mathrm{T}}=\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}a_3\\b_3\\c_3\end{pmatrix}$$

从**向量对矩阵**的作用方面上，我们可以这样理解上述乘式给出的操作意义上的几何内涵：

（1）x 轴上的单位向量 $\boldsymbol{i}$ 左乘一个矩阵就是把矩阵的 x 轴上的行向量（第一行）给取出来；类似地，y 轴上的单位向量 $\boldsymbol{j}$ 左乘一个矩阵就是把矩阵的 y 轴上的行向量（第二行）给取出来；z 轴上的单位向量 $\boldsymbol{k}$ 左乘一个矩阵就是把矩阵的 z 轴上的行向量（第三行）给取出来。

(2) x轴上的单位向量$\boldsymbol{i}$右乘一个矩阵就是把矩阵的 x轴上的列向量（第一列）给取出来；类似地，y 轴上的单位向量 $\boldsymbol{j}$ 右乘一个矩阵就是把矩阵的 y 轴上的列向量（第二列）给取出来；z 轴上的单位向量 $\boldsymbol{k}$ 右乘一个矩阵就是把矩阵的 z 轴上的列向量（第三列）给取出来。

另外，从**矩阵对**向量的作用上，我们可以从几何图形上这样理解其操作意义上的几何内涵：

(3) 一个矩阵右乘 x 轴上的单位向量$\boldsymbol{i}$ 就是把向量$\boldsymbol{i}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 x 轴上的行向量（第一行）；类似地，一个矩阵右乘 y 轴上的单位向量 $\boldsymbol{j}$ 就是把向量 $\boldsymbol{j}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 y 轴上的行向量（第二行）；一个矩阵右乘 z 轴上的单位向量 $\boldsymbol{k}$ 就是把向量 $\boldsymbol{k}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 z 轴上的行向量（第三行）。

(4) 一个矩阵左乘 x 轴上的单位向量$\boldsymbol{i}$ 就是把向量$\boldsymbol{i}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 x 轴上的列向量（第一列）；类似地，一个矩阵左乘 y 轴上的单位向量 $\boldsymbol{j}$ 就是把向量 $\boldsymbol{j}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 y 轴上的列向量（第二列）；一个矩阵左乘 z 轴上的单位向量 $\boldsymbol{k}$ 就是把向量 $\boldsymbol{k}$ 的图形进行缩放旋转变换，变换后的向量就是这个矩阵的 z 轴上的列向量（第三列）。

任何一个矩阵和单位向量 $\boldsymbol{i}$ 相乘得到矩阵的第一列向量。类似地，任何一个向量和单位向量 $\boldsymbol{i}$ 相乘（内积），是这个向量的第一个分量。

图 5-2 中，我们给出了三阶矩阵 $\boldsymbol{A}$ 分别右乘 x、y、z 轴上的单位向量$\boldsymbol{i}^{\mathrm{T}}$、$\boldsymbol{j}^{\mathrm{T}}$、$\boldsymbol{k}^{\mathrm{T}}$ 后的变化几何图形，单位向量$\boldsymbol{i}^{\mathrm{T}}$、$\boldsymbol{j}^{\mathrm{T}}$、$\boldsymbol{k}^{\mathrm{T}}$ 分别**缩放旋转**变化成了向量$(a_1,b_1,c_1)^{\mathrm{T}}$、向量$(a_2,b_2,c_2)^{\mathrm{T}}$ 和向量$(a_3,b_3,c_3)^{\mathrm{T}}$，其中虚线向量表示一种变化的过程。

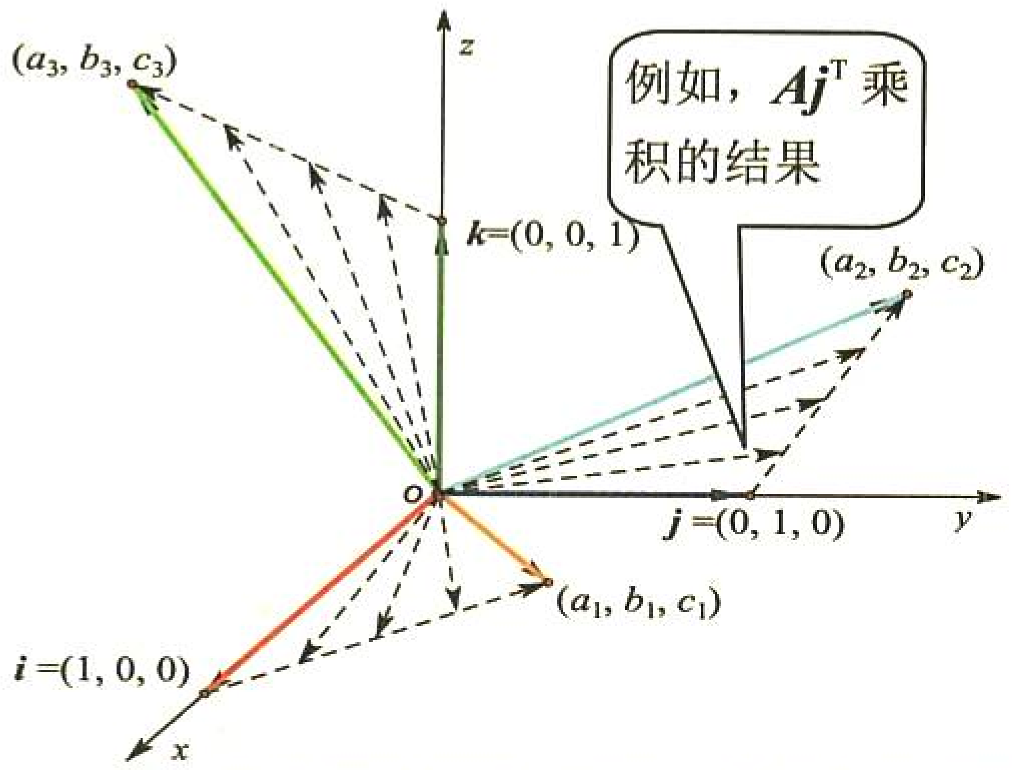


图 5-2 三阶矩阵对单位向量的变换过程

2. 矩阵与任意向量的乘积的几何解释

在前面的章节中我们讲过，一个向量可以拆分为单位坐标向量的线性表示，或者是单位坐标向量的伸缩变换后的和。我们再次列出这个表达式：

$$\boldsymbol{d}=(d_1,d_2,d_3)=d_1\boldsymbol{i}+d_2\boldsymbol{j}+d_3\boldsymbol{k}=d_1(1,0,0)+d_2(0,1,0)+d_3(0,0,1)$$

或

$$\boldsymbol{d}=\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1\boldsymbol{i}^{\mathrm{T}}+d_2\boldsymbol{j}^{\mathrm{T}}+d_3\boldsymbol{k}^{\mathrm{T}}=d_1\begin{pmatrix}1\\0\\0\end{pmatrix}+d_2\begin{pmatrix}0\\1\\0\end{pmatrix}+d_3\begin{pmatrix}0\\0\\1\end{pmatrix}$$

所以，矩阵 $\boldsymbol{A}$ 对任意向量 $\boldsymbol{d}$ 的乘积式就是

$$\begin{aligned}\boldsymbol{A}\cdot\boldsymbol{d} &= \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\left(d_1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + d_2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + d_3\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) \\ &= d_1\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\cdot\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + d_2\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\cdot\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + d_3\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\cdot\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= d_1\boldsymbol{A}\cdot\boldsymbol{i}^{\mathrm{T}} + d_2\boldsymbol{A}\cdot\boldsymbol{j}^{\mathrm{T}} + d_3\boldsymbol{A}\cdot\boldsymbol{k}^{\mathrm{T}} \end{aligned} \tag{5-4}$$

式（5-4）揭示了矩阵 $\boldsymbol{A}$ 对任意向量 $\boldsymbol{d}$ 的乘积的几何解释可以分解为几个操作过程：首先，矩阵 $\boldsymbol{A}$ 分别对单位向量 $\boldsymbol{i}^{\mathrm{T}}$、$\boldsymbol{j}^{\mathrm{T}}$、$\boldsymbol{k}^{\mathrm{T}}$ 进行伸缩旋转变换，得到三个向量（列向量，上节的内容）；然后对这三个列向量 $\boldsymbol{Ai}^{\mathrm{T}}$、$\boldsymbol{Aj}^{\mathrm{T}}$、$\boldsymbol{Ak}^{\mathrm{T}}$ 分别进行伸缩变换，得到 $d_1\boldsymbol{Ai}^{\mathrm{T}}$、$d_2\boldsymbol{Aj}^{\mathrm{T}}$、$d_3\boldsymbol{Ak}^{\mathrm{T}}$；再把此三向量相加，得到的和向量就是 $\boldsymbol{A}\cdot\boldsymbol{d}$ 的乘积。

绘出式(5-4)的几何图像，如图 5-3 所示。图中给出了三个向量 $d_1\boldsymbol{Ai}^{\mathrm{T}} = d_1(a_1, b_1, c_1)$，$d_2\boldsymbol{Aj}^{\mathrm{T}} = d_2(a_2, b_2, c_2)$，$d_3\boldsymbol{Ak}^{\mathrm{T}} = d_3(a_3, b_3, c_3)$ 的几何图形。

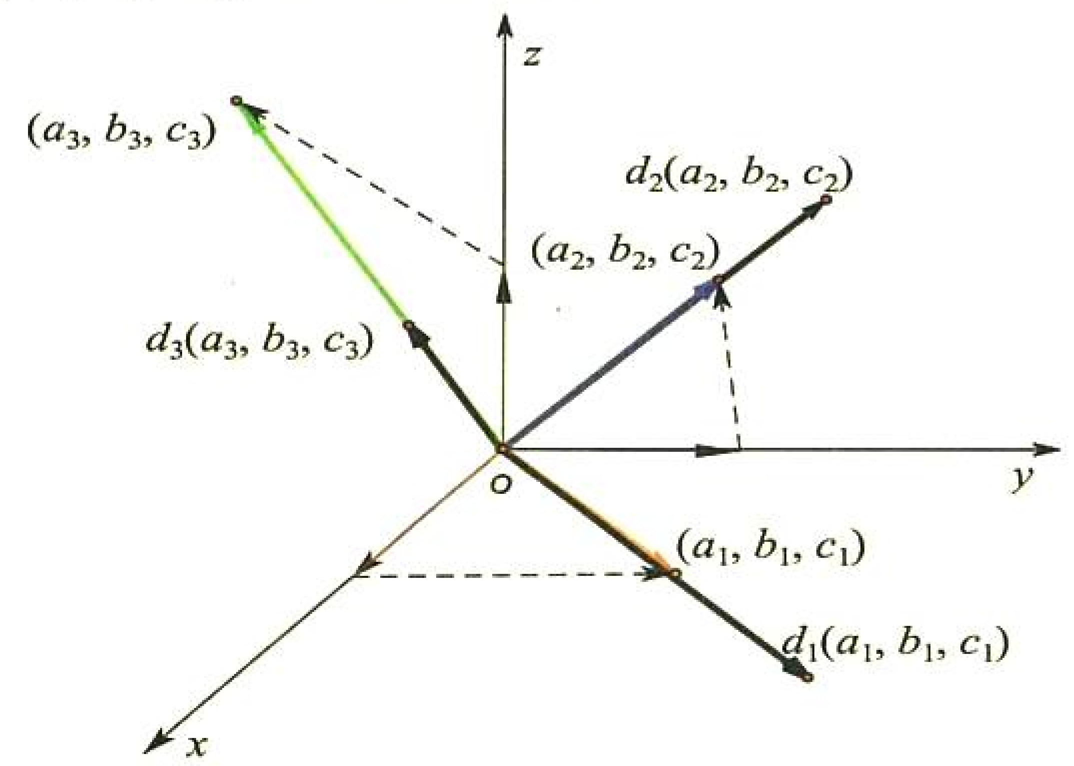


图 5-3 三阶矩阵对任意向量的变换过程之一

图 5-4 把中间结果三个向量 $d_1(a_1, b_1, c_1)$、$d_2(a_2, b_2, c_2)$ 和 $d_3(a_3, b_3, c_3)$ 依照平行四边形法则或多边形法则连加，得到最终结果：

$$\boldsymbol{A}\cdot\boldsymbol{d} = (a_1d_1 + a_2d_2 + a_3d_3, b_1d_1 + b_2d_2 + b_3d_3, c_1d_1 + c_2d_2 + c_3d_3)。$$

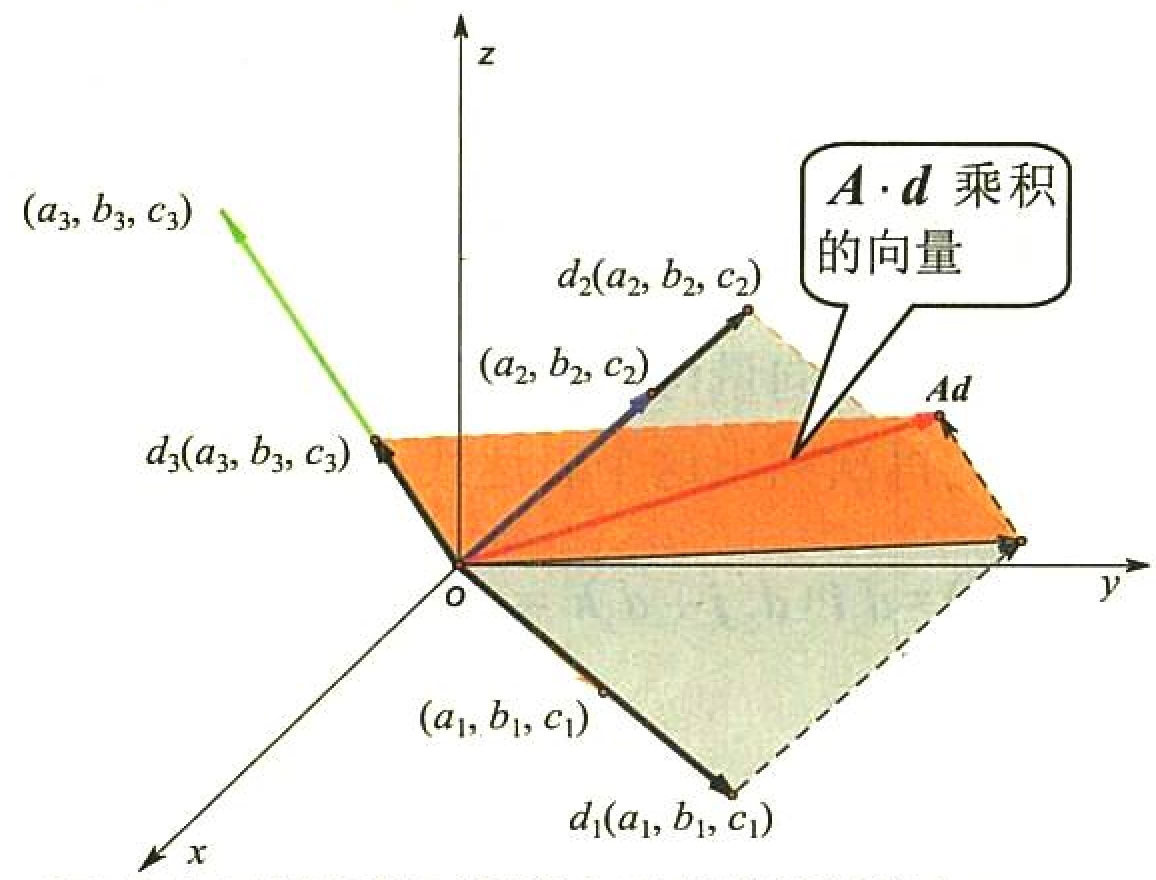


图 5-4 三阶矩阵对任意向量的变换过程之二

呵，不太直观是吧。好，把上式简化一下，利用线性表示的概念（在 5.3.1 节中讲过），我们可以得到一个更直接的易于理解的几何图像。

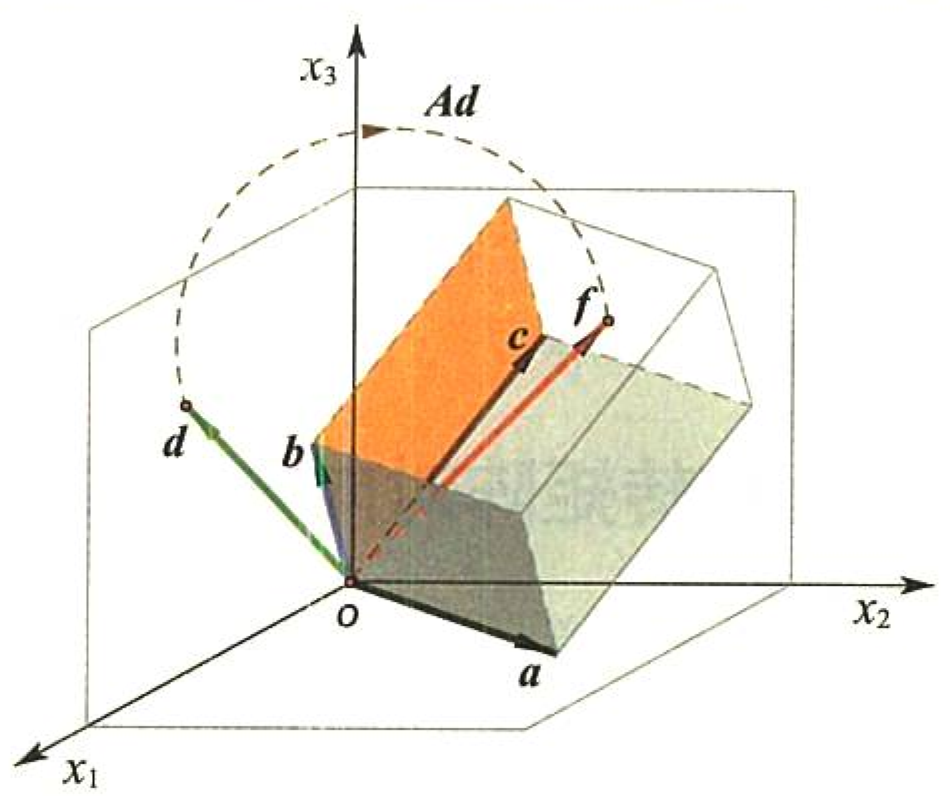
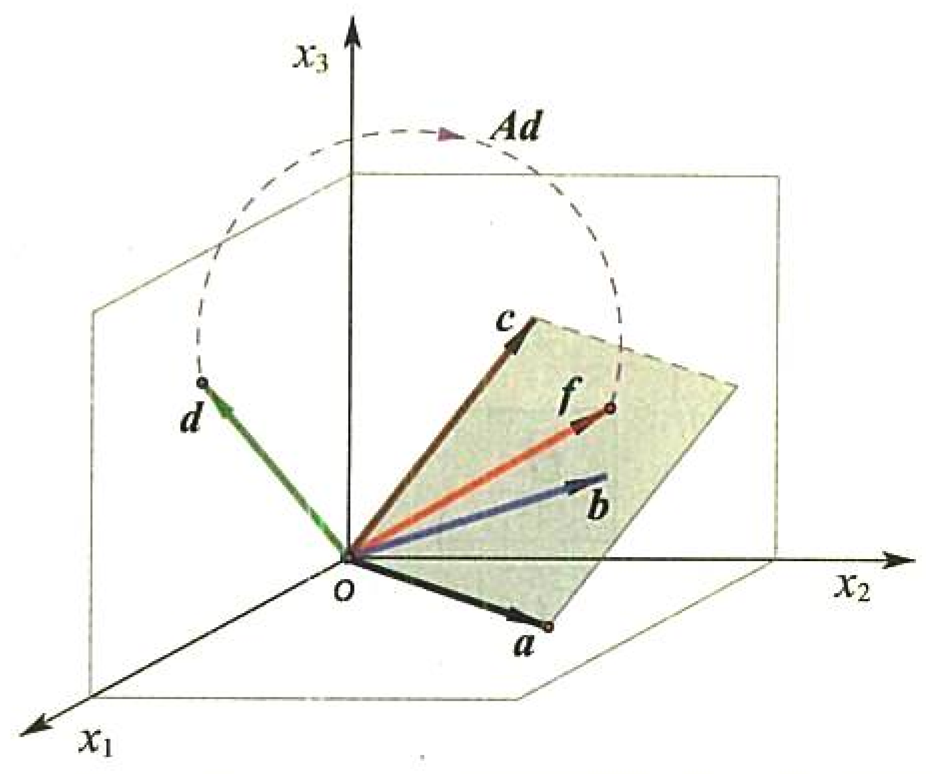
重新调整乘式如下：

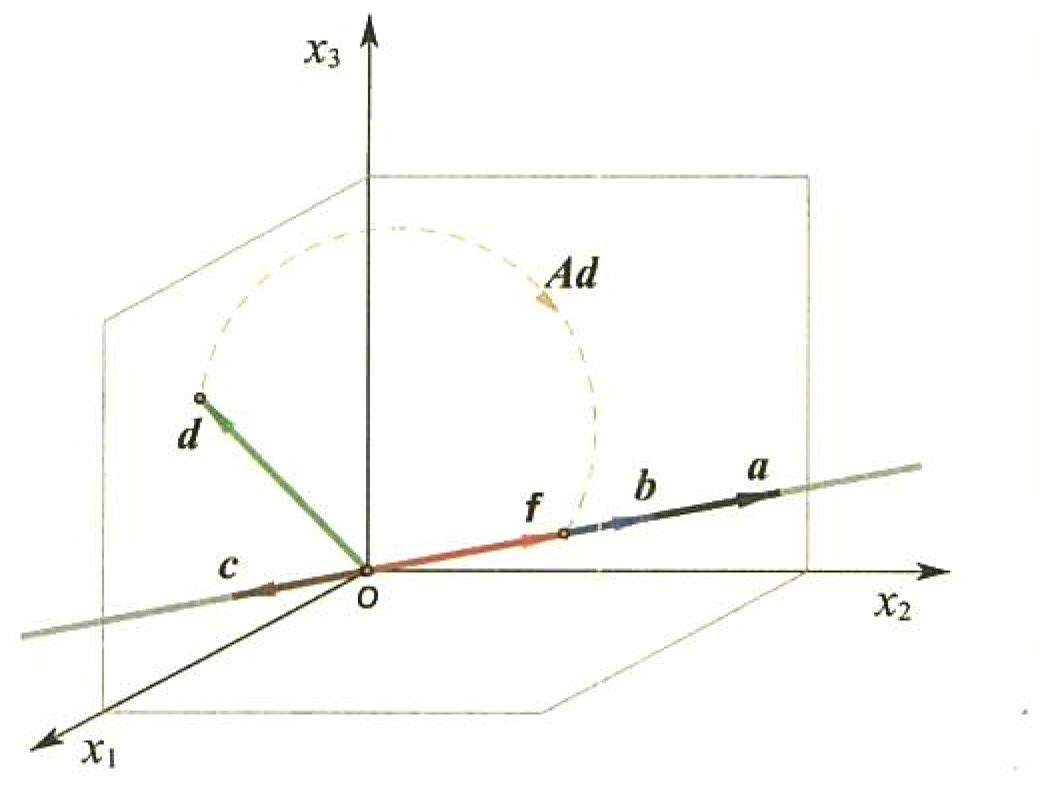
$$\boldsymbol{A}\cdot\boldsymbol{d}=\begin{bmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\\c_1&c_2&c_3\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=\begin{bmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\\c_1&c_2&c_3\end{bmatrix}\left(d_1\begin{pmatrix}1\\0\\0\end{pmatrix}+d_2\begin{pmatrix}0\\1\\0\end{pmatrix}+d_3\begin{pmatrix}0\\0\\1\end{pmatrix}\right)=d_1\begin{pmatrix}a_1\\b_1\\c_1\end{pmatrix}+d_2\begin{pmatrix}a_2\\b_2\\c_2\end{pmatrix}+d_3\begin{pmatrix}a_3\\b_3\\c_3\end{pmatrix}\quad(5\text{-}5)$$

式（5-5）的几何意义就是向量组的线性表示的几何意义（后面的章节中还要讨论），几何解释就是把矩阵 $\boldsymbol{A}$ 的列向量进行伸缩变换（比例变换，也可能改变方向）后首尾相连（即向量的和），得到一个新向量，这个向量就是 $\boldsymbol{A}\cdot\boldsymbol{d}$ 。这个几何解释较前面的解释更简洁一点，但没有大的差别。如果我们知道矩阵的列空间的概念，则可以得到一个新的几何意义上的理解：矩阵 $\boldsymbol{A}$ 的列向量空间是 $\boldsymbol{A}$ 的所有列向量所张成 $\mathbf{R}^n$ 中的一个子空间。尽管这些列向量可能是线性相关的，那么 $\boldsymbol{A}\cdot\boldsymbol{d}$ 的表达式（见式（5-5））则说明，矩阵 $\boldsymbol{A}$ 把 $\mathbf{R}^n$ 中的向量 $\boldsymbol{d}$ 映射到列向量子空间里的一个向量上去了。一个 m 行 n 列的实矩阵 $\boldsymbol{A}_{m\times n}$ 就是一个 $\mathbf{R}^n\to\mathbf{R}^m$ 上的线性映射，或者说，矩阵 $\boldsymbol{A}_{m\times n}$ 把一个 n 维空间的 n 维向量映射为一个 m 维空间的 m 维向量。

看一个综合性的表 5-3。

表 5-3　$m\times n$ 矩阵对任意向量变换的结果

顺序	图形解释	对应的几何图形
1	$\boldsymbol{Ad}=\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}+d_2\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}+d_3\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}=\begin{pmatrix}f_1\\f_2\\f_3\end{pmatrix}$ 3×3 可逆矩阵 $\boldsymbol{A}$ 把一个三维向量 $\boldsymbol{d}$ 映射成一个三维向量 $\boldsymbol{f}$ 。$\boldsymbol{f}$ 存在于矩阵列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 所张成的三维空间内	
2	$\boldsymbol{Ad}=\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1\begin{pmatrix}a_1\\a_2\end{pmatrix}+d_2\begin{pmatrix}b_1\\b_2\end{pmatrix}+d_3\begin{pmatrix}c_1\\c_2\end{pmatrix}=\begin{pmatrix}f_1\\f_2\end{pmatrix}$ 2×3 矩阵 $\boldsymbol{A}$ 把一个三维向量 $\boldsymbol{d}$ 映射成一个二维向量 $\boldsymbol{f}$ 。$\boldsymbol{f}$ 存在于矩阵列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 所张成的二维空间内	

3	$\boldsymbol{Ad}=\begin{bmatrix}a_1 & b_1 & c_1\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}=d_1a_1+d_2b_1+d_3c_1$ $=f_1$ 1×3 矩阵 $\boldsymbol{A}$ 把一个三维向量 $\boldsymbol{d}$ 映射成一个一维向量 $\boldsymbol{f}$。$\boldsymbol{f}$ 存在于矩阵列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 所张成的一维空间内	

把一个向量映射到列空间中的另一个向量，这个说法很简单，但理解起来可能仍然困难。干脆俺再换个说法加点物理意义在里面，就是把一个向量看成一根铁箭，一根真正的刚体，放在那里**没动**，列空间是一个新的坐标系下的空间（坐标基就是矩阵 $\boldsymbol{A}$ 的列向量），这根铁箭在两个坐标系下的投影（就是坐标）$(f_1,f_2,f_3)^{\mathrm{T}}$ 和 $(d_1,d_2,d_3)^{\mathrm{T}}$ 的数学关系就是：

$$\begin{pmatrix}f_1\\f_2\\f_3\end{pmatrix}=\begin{bmatrix}a_1 & b_1 & c_1\\a_2 & b_2 & c_2\\a_3 & b_3 & c_3\end{bmatrix}\begin{pmatrix}d_1\\d_2\\d_3\end{pmatrix}$$

换句话说，在标准坐标系下观察/测量，这个铁箭的投影是 $(f_1,f_2,f_3)^{\mathrm{T}}$；在新坐标系下观察/测量，这个铁箭的投影是 $(d_1,d_2,d_3)^{\mathrm{T}}$。矩阵在两个坐标系之间就是个过渡和转换媒介。呵呵，这下明白了吧。

下节咱讨论一个特殊的矩阵——旋转矩阵。因为通过进一步了解旋转矩阵会深化大家对矩阵几何意义的理解。

3. 旋转矩阵对向量的乘积的几何解释

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。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 $\boldsymbol{A}$：

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

显然，所有的某一特定角度的二阶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，见表 5-4。

表 5-4　二维旋转矩阵的例子（行向量）

θ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$
$\boldsymbol{A}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}\frac{\sqrt2}{2} & -\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2} & \frac{\sqrt2}{2}\end{bmatrix}$	$\begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$	$\begin{bmatrix}-\frac{\sqrt2}{2} & -\frac{\sqrt2}{2}\\ \frac{\sqrt2}{2} & -\frac{\sqrt2}{2}\end{bmatrix}$

比如表 5-4 中的单位矩阵 $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ 实际上就是对一个向量 $\boldsymbol{c}$ 旋转的角度是 0，也就是 $\boldsymbol{Ac}=\boldsymbol{c}$；

而矩阵$\begin{bmatrix}\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{bmatrix}$就是对一个向量 $\boldsymbol{c}$ 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 $\boldsymbol{A}$ 对单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$ 的作用效果：

$$\boldsymbol{Ai}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}$$

$$\boldsymbol{Aj}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}=\begin{pmatrix}\cos(\theta+\pi/2)\\ \sin(\theta+\pi/2)\end{pmatrix}$$

再结合图 5-5，可以看出，旋转矩阵对单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$ 确实分别逆时针旋转了一个 θ 角度。旋转后的两个向量 $\boldsymbol{Ai}$ 和 $\boldsymbol{Aj}$ 保持长度不变和夹角不变。或者说，向量 $\boldsymbol{i}\to\boldsymbol{Ai}$ 长度不变，角度逆时针增加了 θ 度；向量 $\boldsymbol{j}\to\boldsymbol{Aj}$ 长度不变，同时同向角度增加了 θ 度。

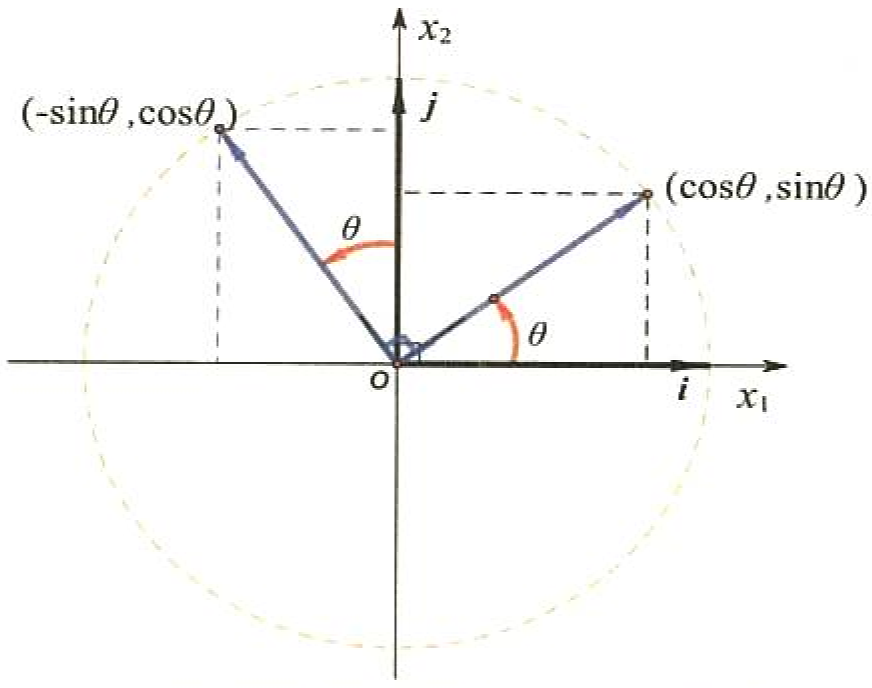


图 5-5 二维旋转矩阵 $\boldsymbol{A}$ 对单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$ 的作用

然后，我们考察旋转矩阵 $\boldsymbol{A}$ 把任意向量 $\boldsymbol{c}$ 变换到 $\boldsymbol{Ac}$ 的情形。对于任意向量 $\boldsymbol{c}$，我们知道，可以分解为单位向量的线性表示：

$$\boldsymbol{c}=\begin{pmatrix}c_1\\c_2\end{pmatrix}=c_1\begin{pmatrix}1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\end{pmatrix}=c_1\boldsymbol{i}+c_2\boldsymbol{j} \tag{5-5}$$

那么，旋转矩阵作用于向量 $\boldsymbol{c}$ 的式子为

$$\boldsymbol{Ac}=\boldsymbol{A}(c_1\begin{pmatrix}1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\end{pmatrix})=c_1\boldsymbol{A}\begin{pmatrix}1\\0\end{pmatrix}+c_2\boldsymbol{A}\begin{pmatrix}0\\1\end{pmatrix}=c_1\boldsymbol{Ai}+c_2\boldsymbol{Aj} \tag{5-6}$$

对比上式（5-5）和（5-6），分向量 $c_1\boldsymbol{i}\to c_1\boldsymbol{Ai}$ 长度不变，角度逆时针增加了 θ 度；分向量 $c_2\boldsymbol{j}\to c_2\boldsymbol{Aj}$ 长度不变，同时同向角度增加了 θ 度。那么，向量的和 $c_1\boldsymbol{i}+c_2\boldsymbol{j}\to c_1\boldsymbol{Ai}+c_2\boldsymbol{Aj}$ 长度不变，角度逆时针增加了 θ 度，即 $\boldsymbol{c}\to\boldsymbol{Ac}$ 长度不变，角度逆时针增加了 θ 度，如图 5-6 所示。

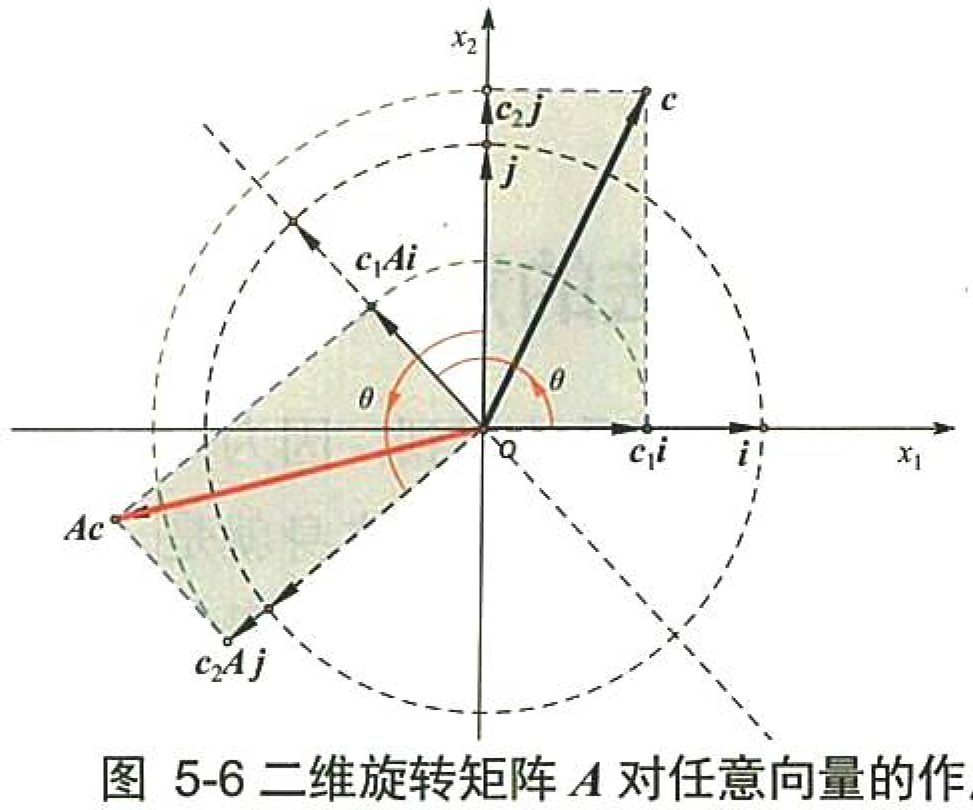


图 5-6 二维旋转矩阵 $\boldsymbol{A}$ 对任意向量的作用

另外，从行向量和列向量两个角度看旋转矩阵，有一个有趣的规律。

首先，我们看表 5-4，观察其**行向量**的图形的变化。由图 5-7（a）的图形知道，当 $\theta=0$

时，旋转矩阵的行向量就是坐标轴上的单位向量 $\boldsymbol{i}$ 和 $\boldsymbol{j}$，随着 θ 角度的增大，矩阵的行向量（两个单位向量）在单位圆上同时进行顺时针等角度旋转。

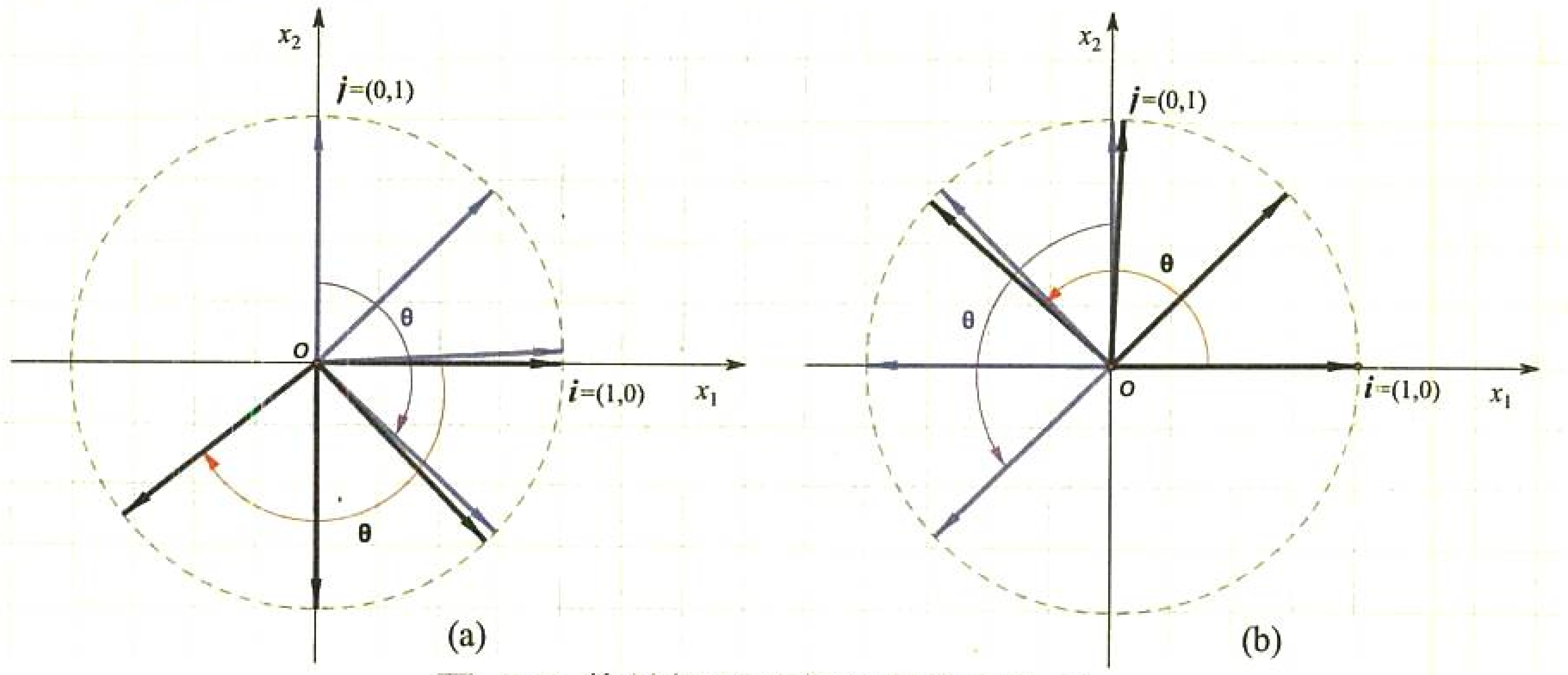


图 5-7 旋转矩阵的行向量和列向量

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？把表 5-4 具体的数值重新列在表 5-5 中，来观察其**列向量**(用虚线按列划分矩阵)的图形的变化。

表 5-5 二维旋转矩阵的例子（列向量）

θ	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$
$A=\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$	$\left[\begin{array}{c:c}1 & 0 \\ 0 & 1\end{array}\right]$	$\left[\begin{array}{c:c}\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{array}\right]$	$\left[\begin{array}{c:c}0 & -1 \\ 1 & 0\end{array}\right]$	$\left[\begin{array}{c:c}-\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\end{array}\right]$

显然，从图 5-7（b）的图形知道，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。因为，旋转矩阵就是正交矩阵。

5.4 矩阵与矩阵乘法的几何意义

两个矩阵相乘，如 $\boldsymbol{AB}$ 的几何意义可以从多个角度来了解。如果把矩阵 $\boldsymbol{A}$ 看做一个几何图形，那么乘以 $\boldsymbol{B}$ 就是把 $\boldsymbol{A}$ 的图形进行了有规律的变换，这个规律的变换就是线性变换（这里实际上把矩阵 $\boldsymbol{A}$ 看成了多个向量的组合）。如果把两个矩阵看做等同的，那么 $\boldsymbol{AB}$ 的结果是把两个线性变换进行了叠加或复合（嗨，至于矩阵乘法的代数意义，您可以先看 6.8.1 节）。下面咱具体讨论一下。

5.4.1 矩阵与矩阵乘法的意义

矩阵与矩阵的乘法可以从矩阵与向量的乘法得到，因为一个矩阵与多个向量相乘，这多个向量就可以组成一个矩阵(会有些限制)。或者说，矩阵本身就是一个有排列顺序要求的向量组，所以矩阵与矩阵相乘可以看做矩阵乘以列向量（或者行向量乘以矩阵）的组合。例如：

$$\boldsymbol{Ac}=\begin{bmatrix}a_1 & a_2 \\ b_1 & b_2\end{bmatrix}\begin{pmatrix}c_1 \\ c_2\end{pmatrix},\quad \boldsymbol{Ad}=\begin{bmatrix}a_1 & a_2 \\ b_1 & b_2\end{bmatrix}\begin{pmatrix}d_1 \\ d_2\end{pmatrix}$$

如果把列向量 $\boldsymbol{c}$ 和 $\boldsymbol{d}$ 组合成一个矩阵 $\boldsymbol{B}=(\boldsymbol{c},\boldsymbol{d})$，则上述的乘积可用两个矩阵的乘积来表示，即

$$AB = A(c,d) = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}\begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix}$$

当然，如果有更多的向量组合起来，可以形成这样的矩阵乘式：

$$AC = A(c,d,e,f,g) = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}\begin{bmatrix} c_1 & d_1 & e_1 & f_1 & g_1 \\ c_2 & d_2 & e_2 & f_2 & g_2 \end{bmatrix}$$

类似地，通过向量乘以矩阵的定义，我们也可以定义反顺序的矩阵乘式（这里向量为行向量）：

$$CA = \begin{pmatrix} c \\ d \\ e \\ f \\ g \end{pmatrix} A = \begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \\ e_1 & e_2 \\ f_1 & f_2 \\ g_1 & g_2 \end{bmatrix}\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

因此，矩阵与矩阵的相乘的几何意义，可以从矩阵与多个向量相乘的几何意义得到，只是多个向量被按照顺序组合成了另一个矩阵。

$AB = C$ 的一般几何意义，就是矩阵 A 把矩阵 B 的数个行向量（或列向量）构成的几何图形进行旋转、缩放、镜像等变换（另一个起到的作用），得到数个新向量，这些新向量作为行向量（或者列向量）组成一个新的矩阵 C，这个新矩阵 C 会构成新的几何图形。对于下面的乘式：

$$C = AB = A(c,\ d,\ e,\ f,\ g) = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}\begin{bmatrix} c_1 & d_1 & e_1 & f_1 & g_1 \\ c_2 & d_2 & e_2 & f_2 & g_2 \end{bmatrix} = \begin{bmatrix} h_1 & i_1 & j_1 & k_1 & l_1 \\ h_2 & i_2 & j_2 & k_2 & l_2 \end{bmatrix} = (h,\ i,\ j,\ k,\ l)$$

我们给出具体的数据例子，被变换图形是矩阵 B 的列向量作为骨架的图形，变换后的图形是矩阵 C 的列向量作为骨架的图形。矩阵 A 是作用矩阵：

$$C = AB = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 3 & 5 & 3 & 2 \\ -3 & -1 & 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 2 & 6 & 5 & 4 \\ 5 & 4 & 4 & 1 & 0 \end{bmatrix}$$

画出这个变换的图形，如图 5-8 所示。一片绿色的小枫叶，经过夏天几多风雨之洗礼的自然变换，终于成了一片红彤彤的大枫叶。

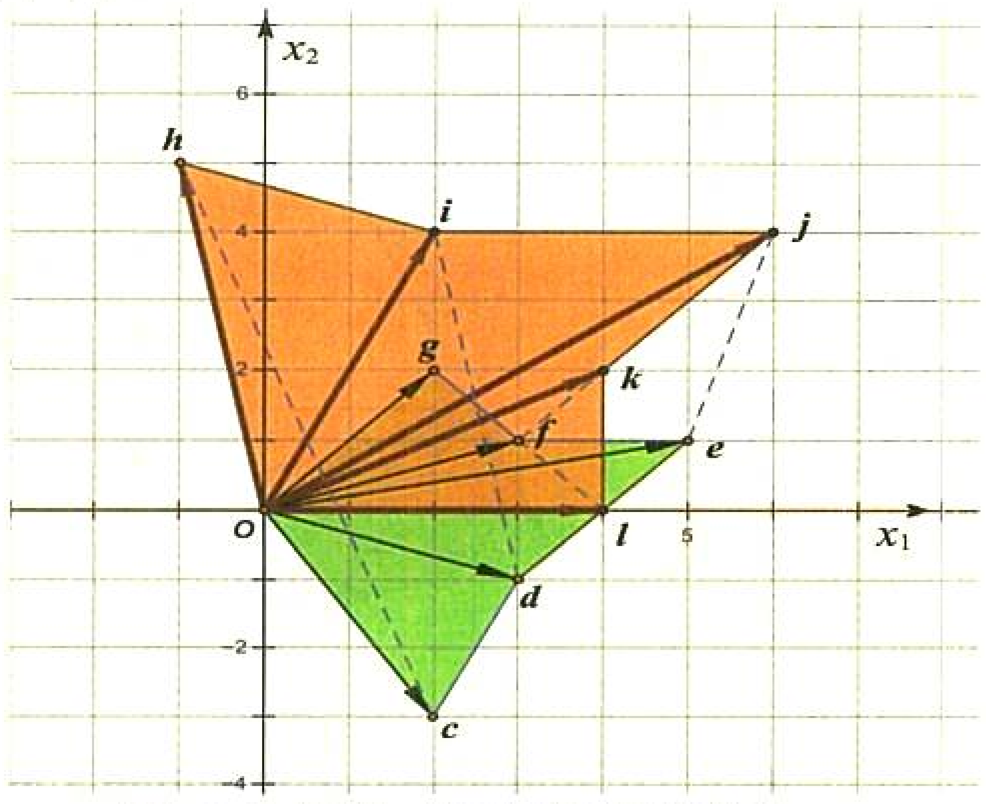


图 5-8 矩阵对图形矩阵的变换

两个矩阵相乘的这样一种理解主要是考察一个矩阵对另一个矩阵所起的变换作用。其作用的矩阵看做动作矩阵，被作用的矩阵看做由行或列向量构成的几何图形。这个理解就是上面给出的几何解释。

另外一个理解就是两个矩阵都被看做作用矩阵，两个矩阵的乘积被看做两个矩阵作用的叠加。一连串的矩阵相乘如 $S = ABCDEF$，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 $\boldsymbol{S}$。工业上的例子就是机器人的手臂（见图 5-9），机械臂上的每个关节就是一个旋转矩阵（比如可以是一个 4×4 矩阵），机械臂末端手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。在 5.6.1 节中，我们将给出几何图形的例子。

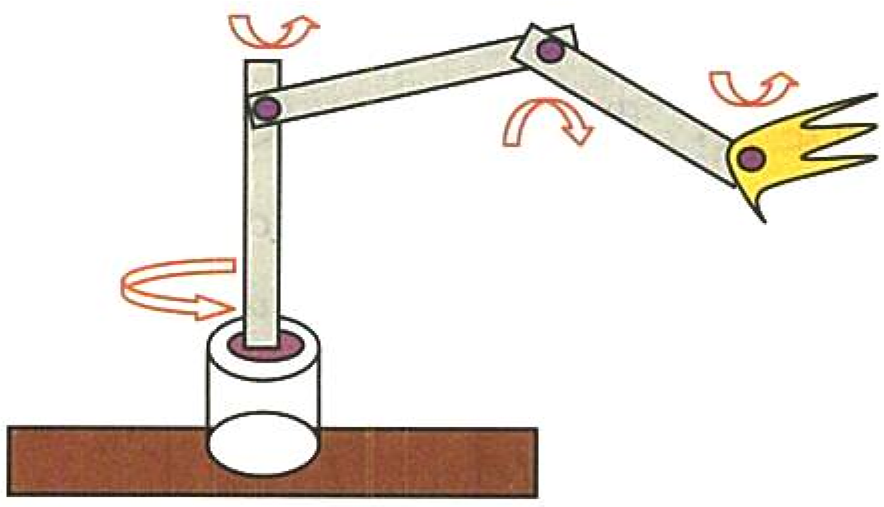


图 5-9 矩阵相乘的传递作用

矩阵相乘可以把矩阵的作用传递下去的原因和意义是什么呢？真正理解它需要熟知线性变换的知识。我想把它的解释放在后面的章节中（比如 5.10 节），目的是希望各位看客能逐步地深入下去。

5.4.2 矩阵左乘与右乘的不同

一个矩阵左乘或右乘另一个矩阵的几何图形，其变化不会相同，因此，矩阵相乘不能满足交换律，也即 $\boldsymbol{AC}\neq\boldsymbol{CA}$。

下面大家看看矩阵的乘积交换后它们的列向量所构成的三角图形的变化（见图 5-10）：

$$\boldsymbol{AC}=\begin{bmatrix}1&1\\1&-1\end{bmatrix}\left[\begin{array}{c:c}2&3\\-3&-1\end{array}\right]=\left[\begin{array}{c:c}-1&2\\5&4\end{array}\right],\quad \boldsymbol{CA}=\begin{bmatrix}2&3\\-3&-1\end{bmatrix}\left[\begin{array}{c:c}1&1\\1&-1\end{array}\right]=\left[\begin{array}{c:c}5&-1\\-4&-2\end{array}\right]$$

矩阵 $\boldsymbol{C}$ 的图形表示为三角形 $\boldsymbol{C}$，矩阵 $\boldsymbol{C}$ 在矩阵 $\boldsymbol{A}$ 的左乘和右乘的作用下分别得到了三角形 $\boldsymbol{AC}$ 和 $\boldsymbol{CA}$ 两个图形，显然，$\boldsymbol{AC}$ 和 $\boldsymbol{CA}$ 是不同的。

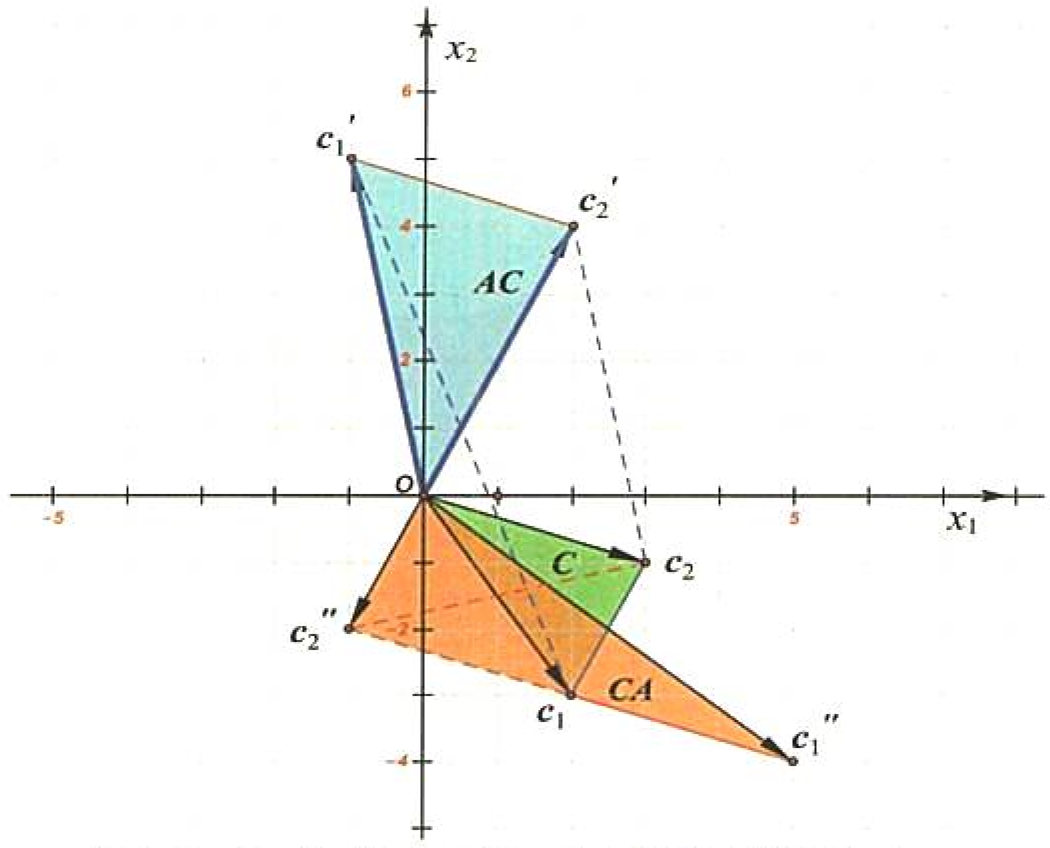


图 5-10 矩阵左乘和右乘的结果不同

矩阵相乘 $\boldsymbol{AB}\neq\boldsymbol{BA}$ 不能满足交换律，是由矩阵相乘的定义决定的，$\boldsymbol{A}$ 乘以 $\boldsymbol{B}$ 定义为 $\boldsymbol{A}$ 的行向量逐个点乘 $\boldsymbol{B}$ 的列向量； $\boldsymbol{B}$ 乘以 $\boldsymbol{A}$ 定义为 $\boldsymbol{B}$ 的行向量逐个点乘 $\boldsymbol{A}$ 的列向量；一个矩阵的行向量与其列向量是不同的，因此 $\boldsymbol{AB}\neq\boldsymbol{BA}$。

一般情况下，矩阵乘法不能交换，但有两个例外要注意：一是单位矩阵可以交换，即 $\boldsymbol{EA}=\boldsymbol{AE}=\boldsymbol{A}$；另一个是逆矩阵可以交换，即 $\boldsymbol{AA}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{E}$。可以交换的主要原因之一是单位矩阵的行向量等于列向量。

5.4.3 矩阵乘幂的几何及物理解释

矩阵乘幂运算就是对一个矩阵 $\boldsymbol{A}$ 进行一定次数的连乘运算 $\boldsymbol{A}^n$。有实际物理意义的可乘幂矩阵常见的有 Markov(马尔科夫)链中的转移概率矩阵、旋转矩阵和图的邻接矩阵等。在天气的马尔科夫链中，天气转移概率矩阵的 n 次幂 $\boldsymbol{A}^n$ 表示第 n 天后天气是晴、阴或雨的概率。逆时针旋转 θ 角的矩阵的 n 次幂 $\boldsymbol{A}^n$ 表示对一个对象旋转了 n 个 θ 角度。邻接矩阵可以表明一个电路网络、电话网络、道路网络以及机构网络的连接关系，而邻接矩阵的乘幂可以得到任意两个网络节点之间的连接长度及是否有连接的情况。比如邻接矩阵的 2 次幂 $\boldsymbol{A}^2$ 的结果告诉我们任意两个节点之间是否存在长度为 2 的道路，而且还告诉我们每一对连接长度为 2 的道路有多少。

矩阵的乘幂的实际意义还表现在具有反馈环路的物理网络中，如一个具有反馈的线性电路。更具体的例子是一个电子或机械的振荡器（见图 5-11）。电子振荡器的选频和放大的电路网络可以抽象为一个矩阵 $\boldsymbol{A}$，电路网络的输出又返回到输入；这样二次输入又经过矩阵的作用得到二次输出……进出进出，反反复复，恰如矩阵 $\boldsymbol{A}$ 的 n 次幂。具体的分析例子见 5.8.2 节的“电子振荡器的特征向量”。

电子振荡器的特征向量

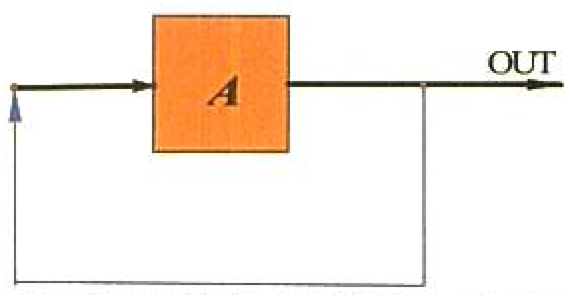


图 5-11 振荡器的数学模型是矩阵的乘幂

矩阵幂的物理意义先说这么多，下面给出旋转矩阵乘幂的几何解释。

前面讲过，二阶逆时针旋转矩阵表示为 $\boldsymbol{A}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$。那么，两个至 n 个旋转矩阵的乘积表达式可以由三角公式得到：

$$\boldsymbol{A}^2=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}=\begin{bmatrix}\cos 2\theta & -\sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{bmatrix}$$

$$\boldsymbol{A}^3=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos 2\theta & -\sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{bmatrix}=\begin{bmatrix}\cos 3\theta & -\sin 3\theta\\ \sin 3\theta & \cos 3\theta\end{bmatrix}$$

$$\vdots$$

$$\boldsymbol{A}^n=\begin{bmatrix}\cos n\theta & -\sin n\theta\\ \sin n\theta & \cos n\theta\end{bmatrix}$$

由此可以看到，旋转矩阵的连乘积等效于旋转角的连加和。图 5-12 给出了旋转矩阵 $\boldsymbol{A}^1$、$\boldsymbol{A}^2$、$\boldsymbol{A}^3$ 对单位向量 $\boldsymbol{i}$ 的旋转效果。矩阵的乘法变成了角度的加法。

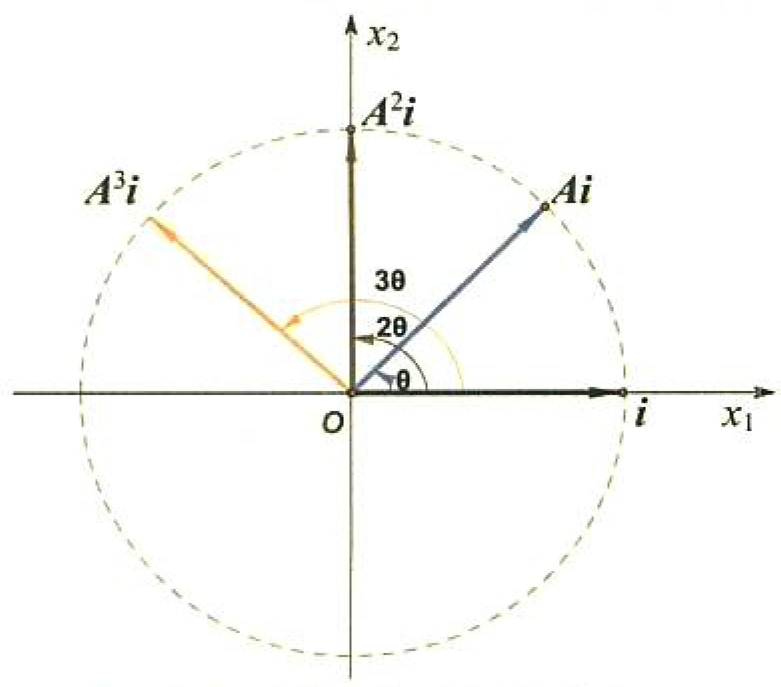


图 5-12 旋转矩阵的乘幂

矩阵乘幂的次数越高，越显露出矩阵的特征值出来，这个方法用于数值线性代数中求解特征值。这个问题我们在下面特征值和特征向量的几何意义一节（见 5.8.1 节）中也有深入讨论。

5.5 矩阵与线性变换关系的几何意义

其实矩阵一开始就与线性变换纠结在一起了，但俺总觉得线性变换概念比较高深哈，故在介绍了矩阵的基本运算后才推出矩阵的线性变换意义出来，希望专业人士笑谅（笑着原谅）啊。

我们在前言里面区分了线性映射和线性变换的概念，其实无论是映射还是变换都可以用矩阵来表示。简单地说，$m \times n$ 阶矩阵可表示把一个 n 维空间的向量映射到 m 维空间的向量的线性映射（详见 5.3.2 节），而一个 n 阶方阵是把一个 n 维空间的向量映射到自身空间另外一个向量的线性变换。因此，多数情况下不必刻意区分映射和变换。

5.5.1 线性变换如何用矩阵表示

从前面矩阵的乘法可以知道，任意一个矩阵其本身蕴含着一个变换。这个变换我们可以称为一个矩阵变换。前面章节里我们也讲过线性变换的概念：数域 F 上线性空间 V 中的变换 T 称为 V 中的**线性变换**，那么当且仅当满足条件：

$$\begin{aligned} &T(\boldsymbol{\alpha}+\boldsymbol{\beta})=T(\boldsymbol{\alpha})+T(\boldsymbol{\beta}) \quad &(\boldsymbol{\alpha},\boldsymbol{\beta}\in V) \\ &T(k\boldsymbol{\alpha})=kT(\boldsymbol{\alpha}) \quad &(k\in F,\boldsymbol{\alpha}\in V) \end{aligned}$$

这里，我们把变换 T 的符号替换为矩阵 $\boldsymbol{A}$，线性变换就变成了矩阵变换。

实际上，从 $\mathbf{R}^n \to \mathbf{R}^m$ 上的线性变换都可以表述为一个矩阵变换；反过来，**一个矩阵变换也必然是一个线性变换。两者具有一一对应的关系**。这个对应关系笼统地表述如下：

- 线性变换的和对应着矩阵的和；
- 线性变换的乘积对应着矩阵的乘积；
- 线性变换的数量乘积对应着矩阵的数量乘积；
- 线性变换的逆对应着矩阵的逆。

因此，对于一个变换我们需要弄懂两个问题：这个变换是否为线性变换？如果是线性变换，应如何求对应的矩阵？

判断是否为线性变换只要确定这个变换是否满足加法和数乘法则就可以了。线性变换一般可以用文字描述出来，如“一个旋转角度为 $\pi/8$ 的旋转变换”，或者“一个关于 $x_2=-x_1$ 的镜像变换”等，但如何知道旋转变换所对应的矩阵呢？这个问题有幸由一个定理给出了回答。作为一个例子，下面的定理给出了如何把一个 $\mathbf{R}^2$ 空间上的线性变换转换成一个对应的二阶矩阵的办法（对 n 阶空间也适用）。

定理：设 $T:\mathbf{R}^2 \to \mathbf{R}^2$ 是任意一个线性变换，那么 T 变换对应的矩阵的列向量为 $T(e_1)$ 和 $T(e_2)$。

为了以后方便引用，我们不妨称此定理为“**线性变换的矩阵定理**”。这个定理的意思是，只要求解空间坐标系轴上的标准正交基在这个线性变换下的映像，就可以用像组成对应的矩阵。

亲们，千万记住这个定理！这可是线性代数中基本的定理，因为它是线性变换和矩阵关联的纽带。

$R^2 \to R^2$

OK，我们验证两个例子加深一下印象。先看一个在平面上的旋转变换的例子（呵，这个例子举了 n 多次了）。

设变换 $T:\mathbf{R}^2 \to \mathbf{R}^2$ 是 $\mathbf{R}^2$ 中把每个向量按逆时针方向绕原点 o 转动 θ 角的旋转变换。这里我们要弄清两个问题：一是旋转变换是否是线性变换；二是旋转变换对应的矩阵如何求（注：这个矩阵就是前面多次介绍的平面旋转矩阵）。

判断是否为线性变换就是确认这个变换是否满足加法和数乘。我们看图 5-13。

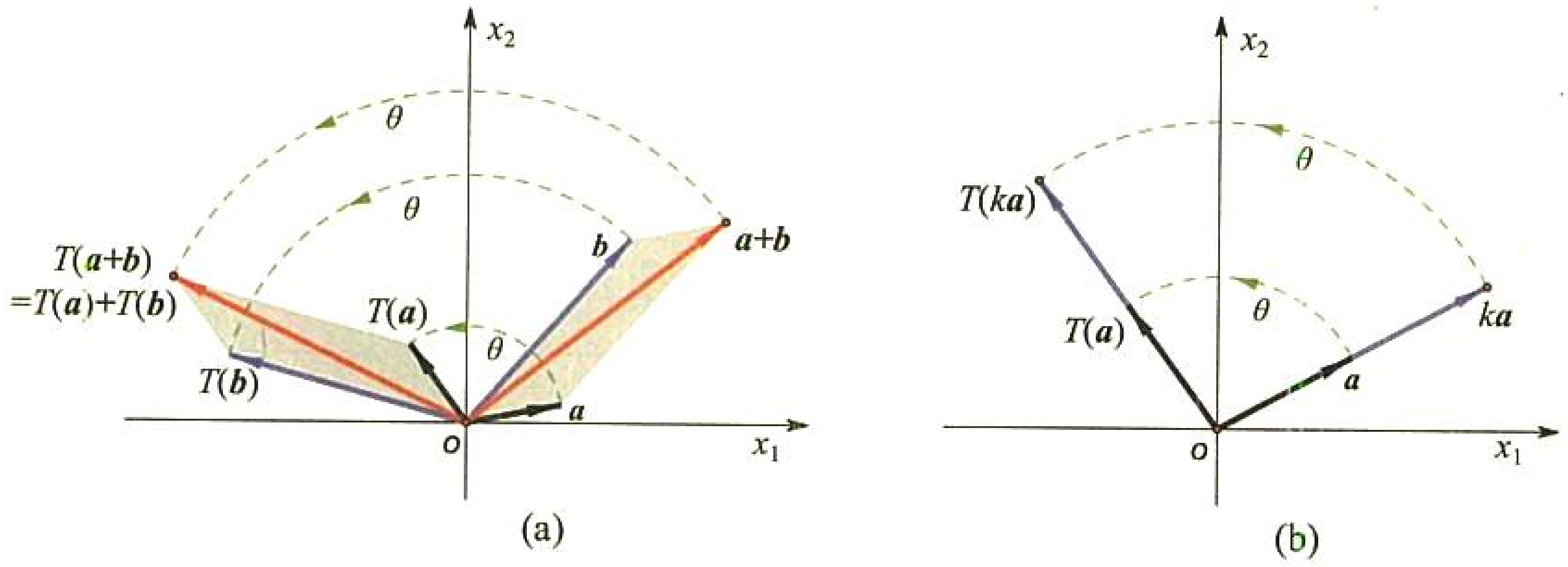


图 5-13 旋转变换是线性变换

图 5-13（a）中向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 同时逆时针旋转了一个 θ 角，因此两个向量之间的夹角没变，长度也没变，显然，构成的平行四边形的外形没变；或者说，平行四边形也同时逆时针旋转了同样的角度。因此，旋转前后的平行四边形的对角线长度没变，同时也旋转了同样的角度。换句话说，旋转前的平行四边形对角线向量 $\boldsymbol{a}+\boldsymbol{b}$ 经过旋转变换 $T()$ 后等于旋转后平行四边形的对角线 $T(\boldsymbol{a})+T(\boldsymbol{b})$，亦即 $T(\boldsymbol{a}+\boldsymbol{b})=T(\boldsymbol{a})+T(\boldsymbol{b})$。这说明满足加法法则。

图 5-13（b）是满足数乘 $T(k\boldsymbol{a})=kT(\boldsymbol{a})$ 的图示，解释从略。

我们已经知道了平面空间中的逆时针旋转的通用矩阵是 $\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$，如何由定理得到这个旋转矩阵呢？实际上咱在前面多处章节中已给了回答。这里只是套一下定理，给予重新确认：

如图 5-14 所示（也可参考图 5-5），把坐标系 x_1 轴的单位向量 $\boldsymbol{e}_1$（有时写做 $\boldsymbol{i}$）按照线性变换的要求逆时针旋转 θ 角度，得到基向量的像，即 $T(\boldsymbol{e}_1)=T(\boldsymbol{i})=\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}$；同理，把 x_2 轴上的单位向量 $\boldsymbol{e}_2$（有时写做 $\boldsymbol{j}$）逆时针旋转 θ 角度，即有 $T(\boldsymbol{e}_2)=T(\boldsymbol{j})=\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}$。根据定理，这两个新向量就是旋转矩阵的两个列向量，把它们按顺序组合起来就得到旋转矩阵 $\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$。

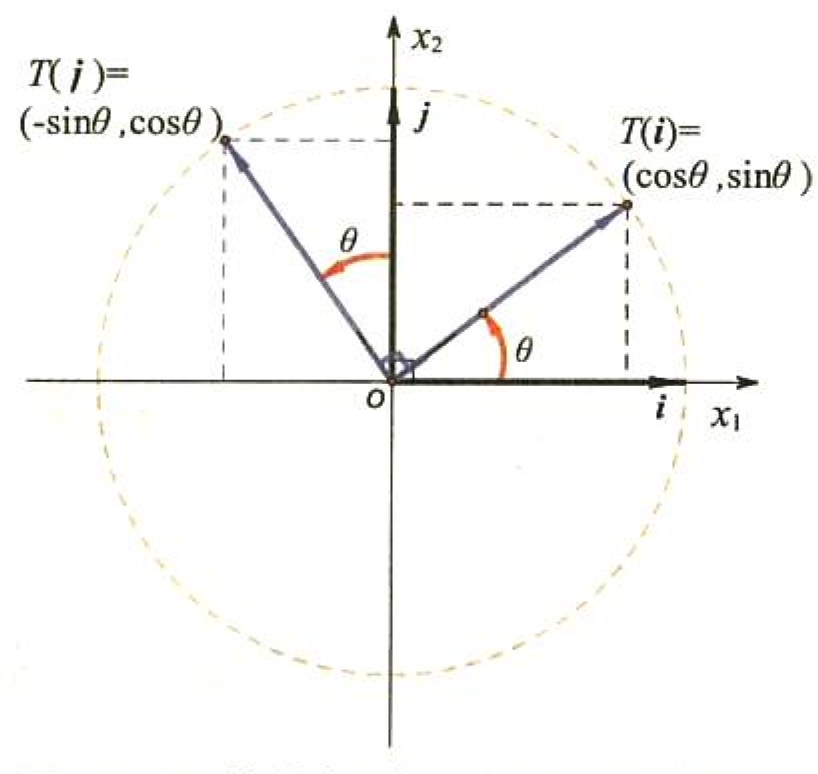


图 5-14 旋转矩阵列向量的求解

因为这个求矩阵的定理很实用，咱再举一个镜像变换的例子。

例 5.2 设$T:\mathbf{R}^2\rightarrow\mathbf{R}^2$是把向量空间$\mathbf{R}^2$中的每个向量对直线$x_2=-x_1$作反射的变换。容易看出，这个变换是保持了向量的加法和数乘的原则。如图 5-15(a)所示，向量$\boldsymbol{a}$和$\boldsymbol{b}$及其和$\boldsymbol{a}+\boldsymbol{b}$所构成的平行四边形在镜像变换$T()$（对于直线$x_2=-x_1$的反射）后同样是平行四边形，它是反射向量$T(\boldsymbol{a})$、$T(\boldsymbol{b})$和$T(\boldsymbol{a})+T(\boldsymbol{b})$所构成的平行四边形，这说明加法法则被保持。图 5-15(b)是向量的数乘运算，变换前后的比例k没有改变，这说明数乘法则被保持。

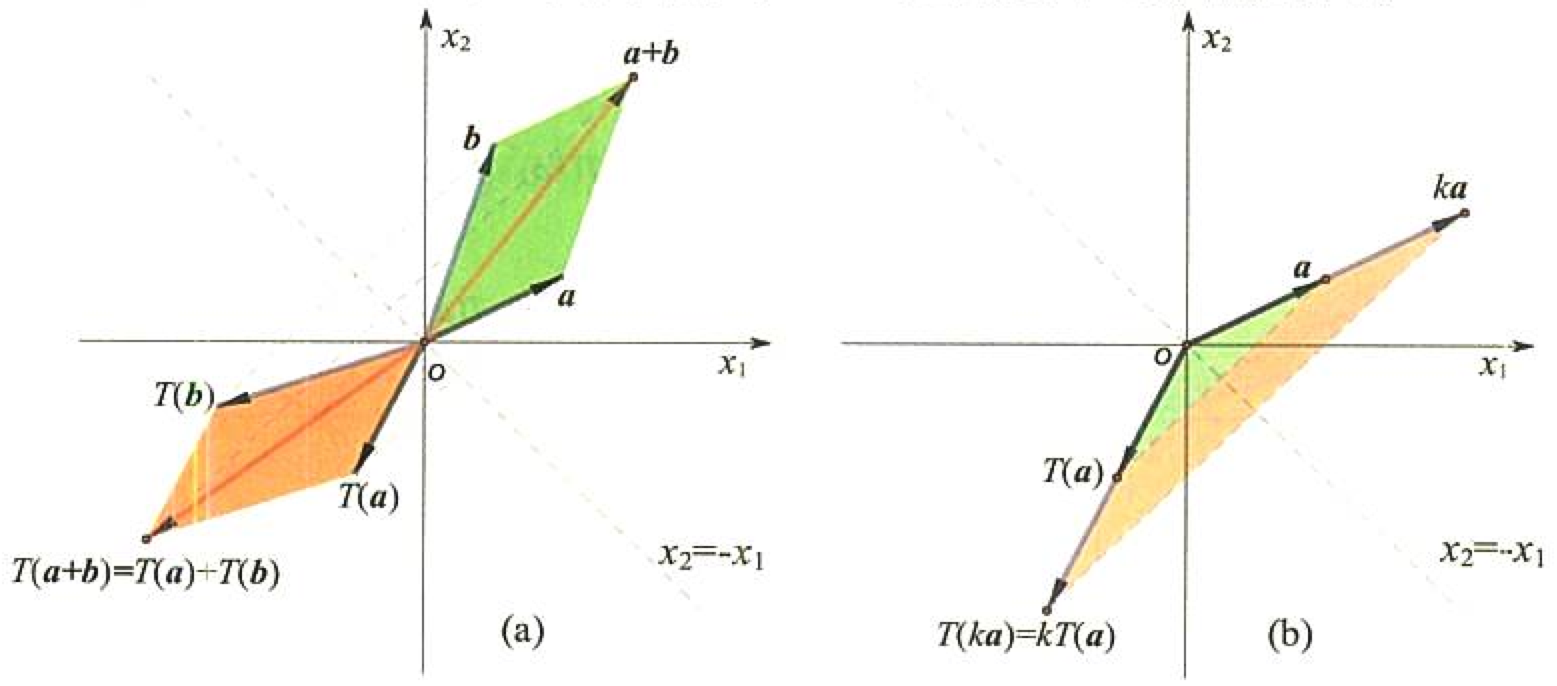


图 5-15 镜像变换是一个线性变换

因此镜像变换是一个线性变换。下面求解对应的镜像矩阵。

对应镜像变换的矩阵的第一个列向量是$T(\boldsymbol{e}_1)=T(\boldsymbol{i})=-\boldsymbol{j}=\begin{pmatrix}0\\-1\end{pmatrix}$，第二个列向量是$T(\boldsymbol{e}_2)=T(\boldsymbol{j})=-\boldsymbol{i}=\begin{pmatrix}-1\\0\end{pmatrix}$，见图 5-16。根据线性变换的矩阵定理，把它们组合起来，所以其镜像矩阵为$\begin{bmatrix}0&-1\\-1&0\end{bmatrix}$。

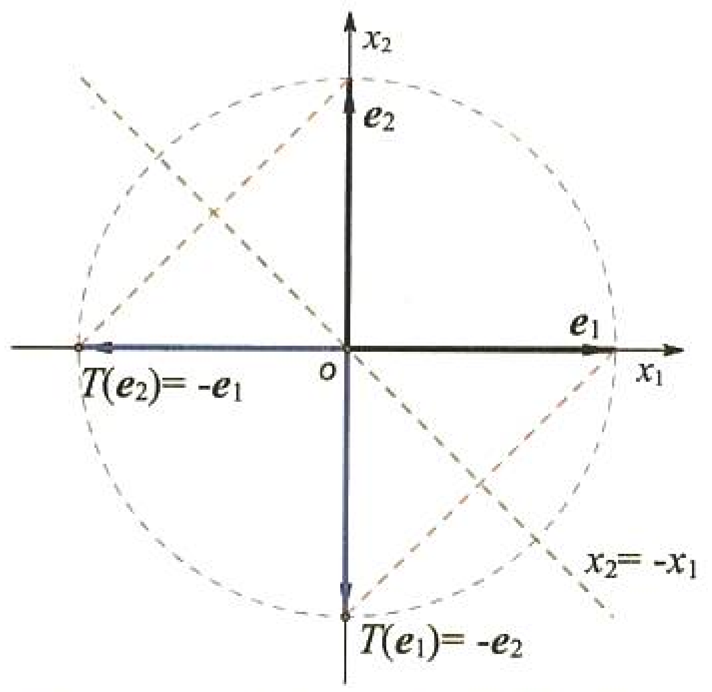


图 5-16 镜像矩阵列向量的求解

所以啊，由于线性变换与矩阵之间有一一对应（在给定基的前提下），而且保持线性运算关系不变（线性变换的加法和数乘分别对应于在某一个基下的矩阵的加法和数乘），因此，可以用矩阵来研究线性变换，也可以用线性变换来研究矩阵。

5.5.2 线性变换矩阵定理的几何及物理意义

一般的线性变换的矩阵定理可以这样给出：

设$T:\mathbf{R}^n\rightarrow\mathbf{R}^n$是任意一个线性变换，那么$T$变换对应的矩阵为$[T(\boldsymbol{e}_1),T(\boldsymbol{e}_2),\cdots,T(\boldsymbol{e}_n)]$。 其中$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$是自然正交坐标系轴$x_1,x_2,\cdots,x_n$上的单位基向量。

这个定理确实很奇妙，只要把单位基向量被线性变换映射的像向量按序排列就得到了此变换矩阵。其实这个定理的几何意义在前面的例子里已给出了。这里咱给出一般意义上的几何解释。

对于二维空间上的线性变换，一个线性变换会把标准正交基的两个基向量变换成另外两个向量，如果把它们张成的四边形勾画出来，就会发现，线性变换把第一象限的正方形变成了平行四边形。

如图 5-17，标准正交基的两个基向量 e_1、e_2 张成的是正方形。在一个线性变换下，两个基向量的像 $T(e_1)$、$T(e_2)$ 张成的一般是一个平行四边形。平行四边形的两棱（向量）构成了变换矩阵。

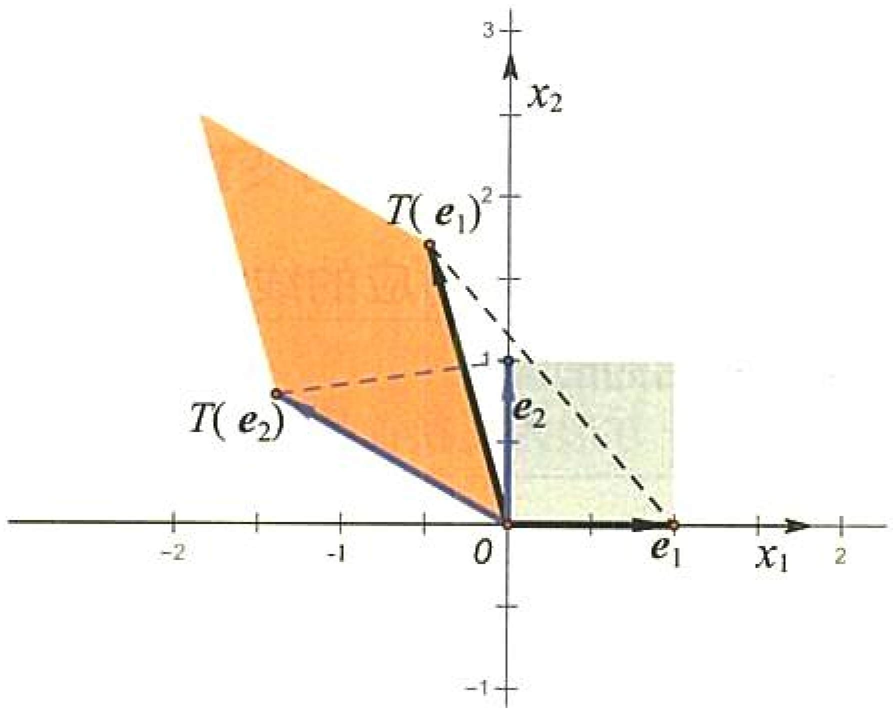


图 5-17 线性变换把正方形映射成平行四边形

类似地，三维空间里，一个线性变换会把一个标准正交基所张成的立方体映射成一个平行六面体。平行六面体的三棱（向量）构成了变换矩阵。

以上是线性变换矩阵定理的几何意义。如果要理解线性变换矩阵构成的物理意义，我们可以用前面的点石成金矩阵（参见 5.1.2 节）的构成来类比理解。

点石成金是个线性变换。$\begin{pmatrix}1\\0\end{pmatrix}$ 吨矿石被变换为 $\begin{pmatrix}a\\b\end{pmatrix}$ 克金银；$\begin{pmatrix}0\\1\end{pmatrix}$ 吨矿石被变换为 $\begin{pmatrix}c\\d\end{pmatrix}$ 克金银。这个刚好是把坐标轴上的标准基进行了变换，因此可以由此组合成变换矩阵 $\begin{bmatrix}a & c\\ b & d\end{bmatrix}$。

常见的线性变换有初等变换、等价变换、正交变换、相似变换、合同变换等。下面或以后的章节我们将讨论对应着各种线性变换的不同类型矩阵的几何意义。首先看看对应着初等变换的称为初等矩阵的几何意义。

5.5.3 矩阵及其对应线性变换的几何图形

如果给你一个矩阵，如何马上理解其几何意义上的线性变换呢？其实很简单，你只要把前面已知变换求矩阵的矩阵定理反着用就可以了。也就是说，把一个矩阵的列向量所张成的平行多面体和同阶的单位矩阵的单位向量所张成的正多面体进行对比，就可以发现其对应的变换关系。一般情况下，我们只能对二阶和三阶矩阵的几何图形进行明显的对比。从前面的例子我们知道，二阶和三阶矩阵的正多面体就是正方形和正立方体。

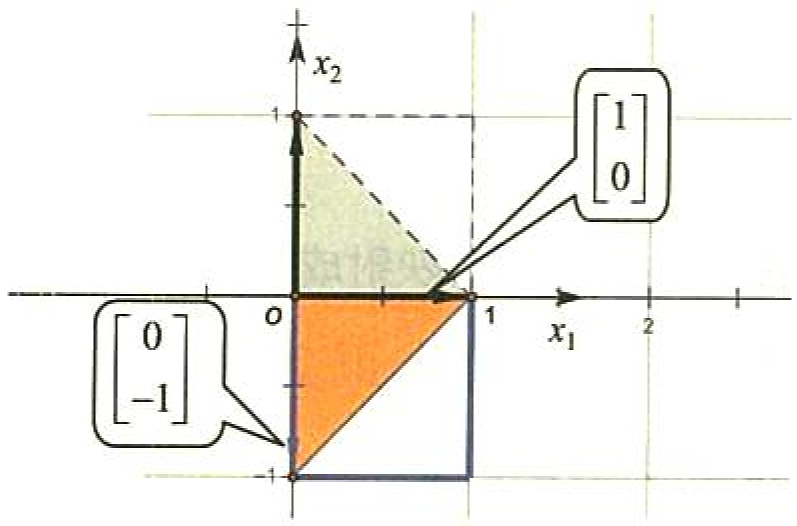
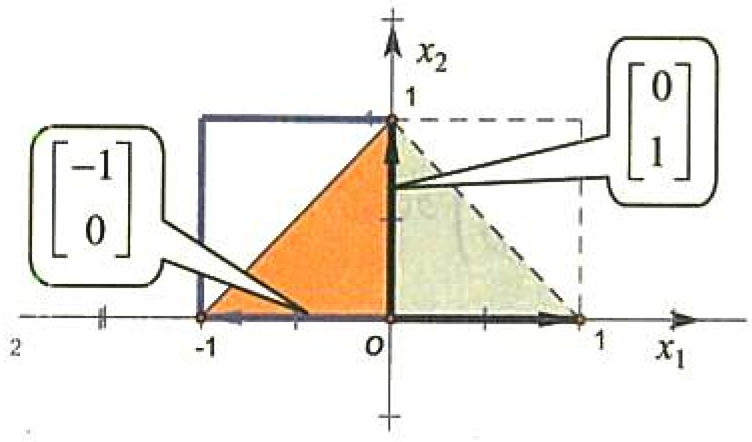
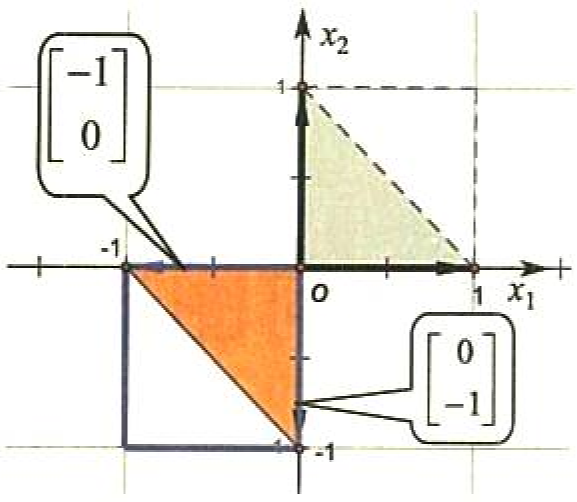
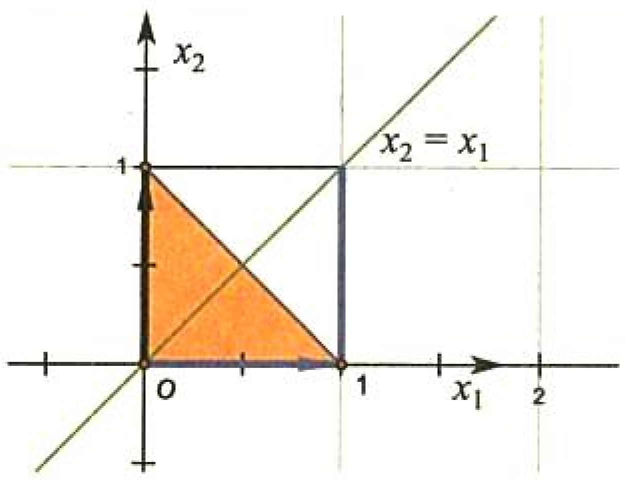
一个几何图形乘以一个矩阵，就会对这个图形进行一个线性变换或矩阵变换。几何图形可以是一个向量、一个三角形、一个四边形或者一个圆，这些几何图形都被认为是一个向量集合

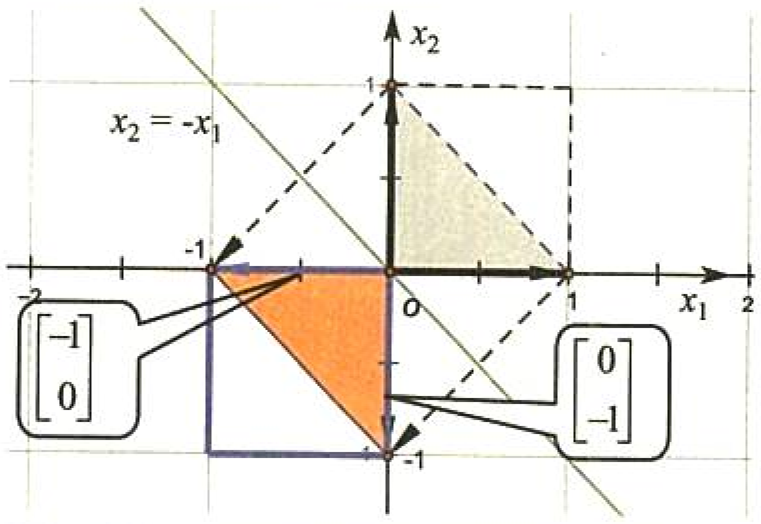
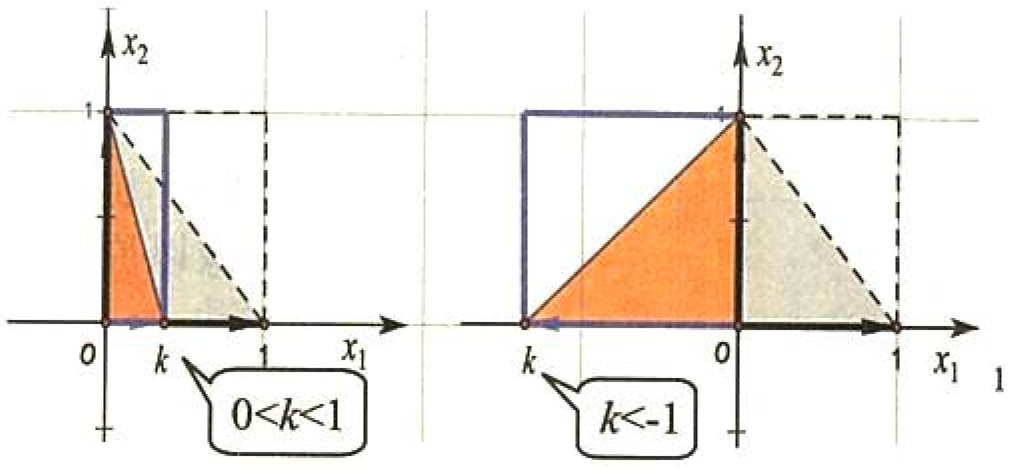
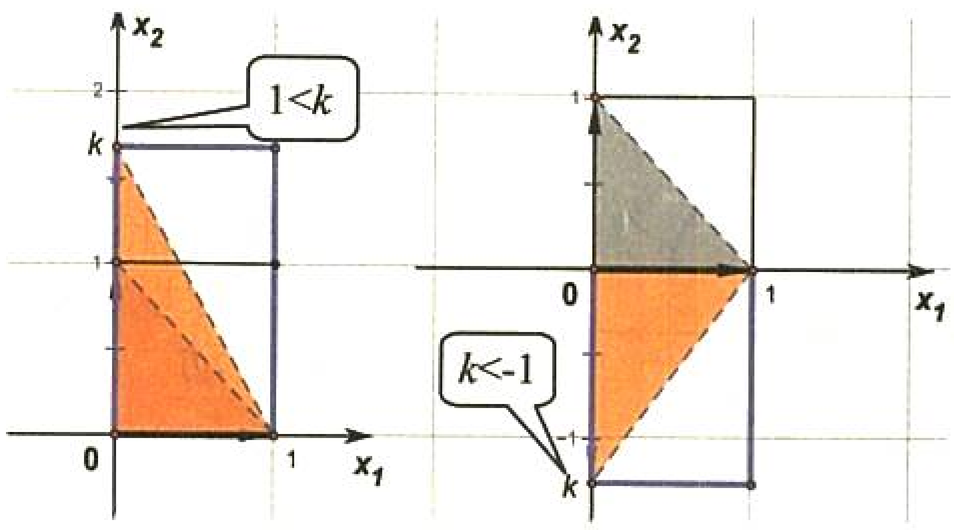
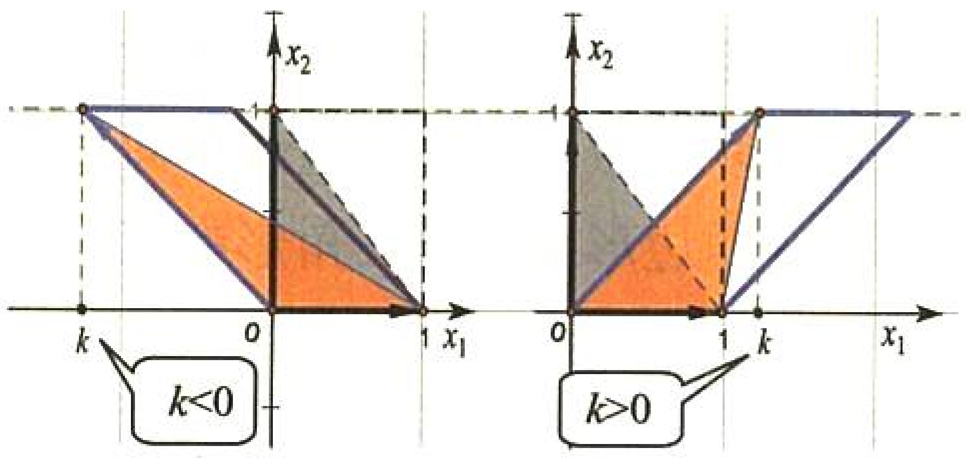
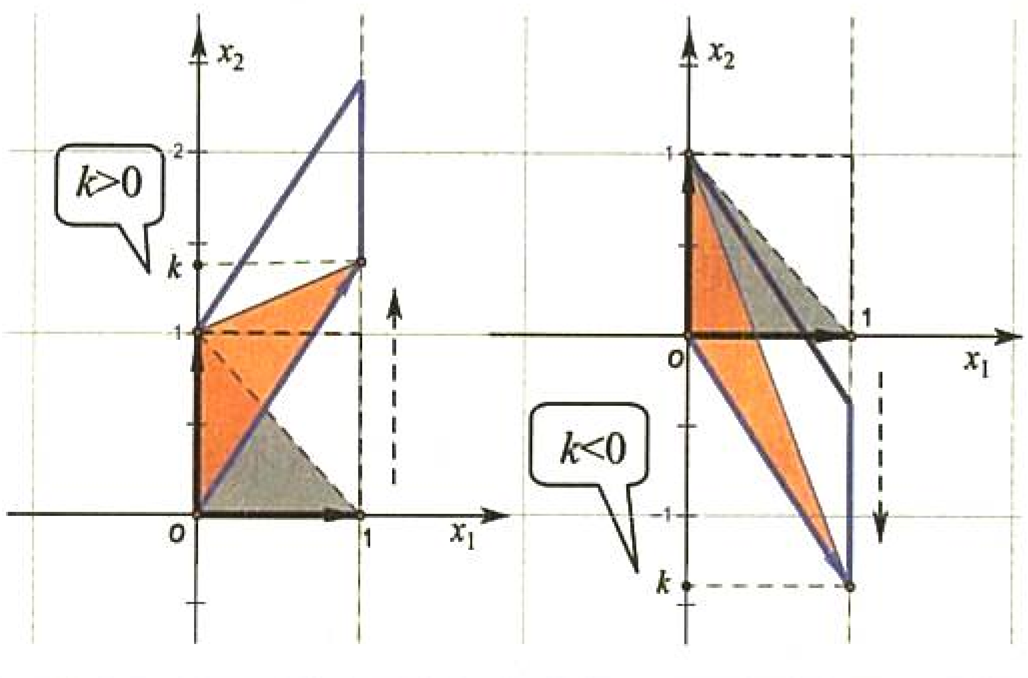
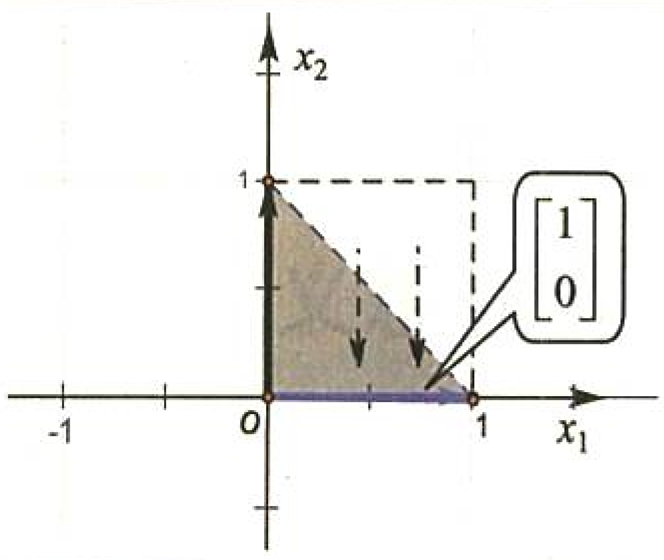
所构成的。当一个矩阵对众多不同的几何图形进行变换时就会显现不同的变换特征，同时也会加深我们对一个具体矩阵意义的了解。比如，我们常常见到二阶矩阵对一个单位正方形的图形进行变换，实际上就是把单位矩阵 $\boldsymbol{E}$ 的行向量或列向量所张成的几何图形进行的一个变换。

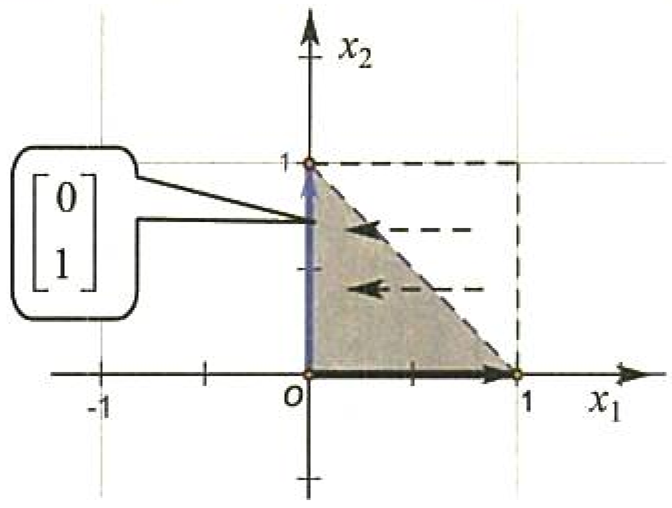
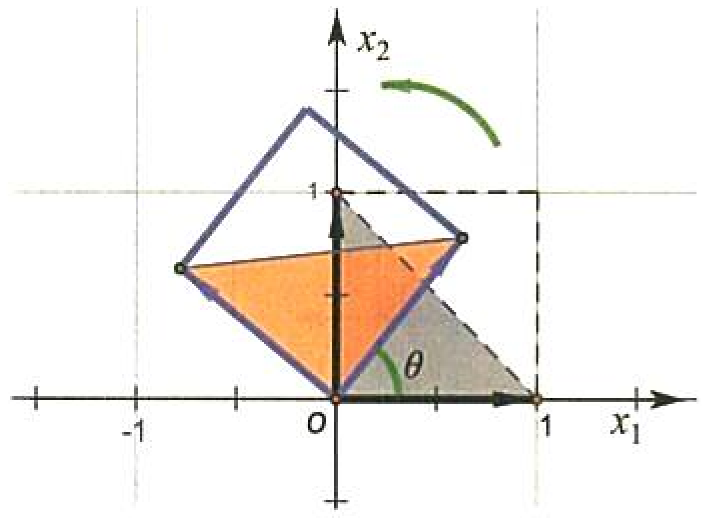
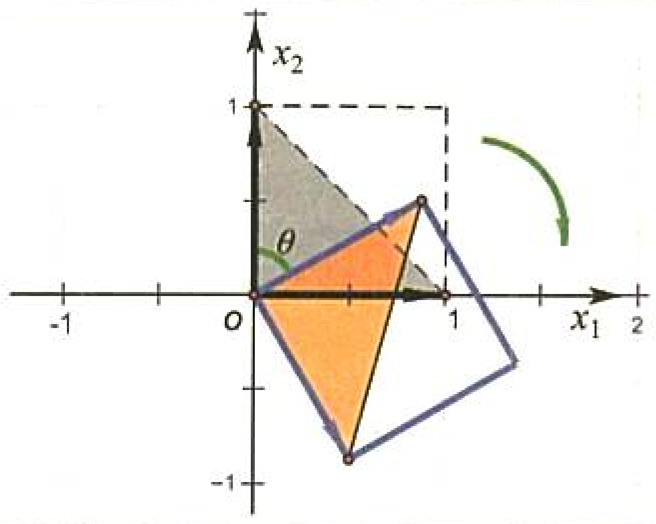
另外，如果被变换的几何图形是一个向量圆，那么就会容易发现矩阵的特征值和特征向量的特质，如被变换的几何图形是一个正方形，其特征向量的几何特点就不太明显。因此在后面的矩阵的特征向量和秩的几何意义的讨论中，我们总把被变换的几何图形设为单位圆或单位球。

下面汇总了各种二阶矩阵所对应的线性变换的平面几何图形（见表 5-6）。这里被变换的图形是一个位于第一象限内的三角形或正方形，它们由单位矩阵$\begin{bmatrix}1&0\\0&1\end{bmatrix}$所生成，变换后的像一般位于其他象限内。这个汇总可帮助大家有个系统的认识。另外，在表 5-6 的最右列给出了线性变换对应矩阵的行列式值，由行列式的值我们可以看出变换前后平面图形面积的变化比率。

表 5-6 典型二阶矩阵所对应的线性变换的几何图形

序号	矩阵	变换前后的图像（第一象限的三角形和正方形为变换前原图形）	对应的线性变换	矩阵的行列式值
1	$\left[\begin{array}{c:c}1&0\\0&-1\end{array}\right]$		关于 x_1 轴的镜像对称变换；图像（三角形和正方形）被翻转到 x_1 轴下	-1，面积变号
2	$\left[\begin{array}{c:c}-1&0\\0&1\end{array}\right]$		关于 x_2 轴的镜像对称变换；图像被翻转到 x_2 轴左侧	-1，面积变号
3	$\left[\begin{array}{c:c}-1&0\\0&-1\end{array}\right]$		关于原点的对称变换；图像被翻转到原点的对面	1，面积不变
4	$\left[\begin{array}{c:c}0&1\\1&0\end{array}\right]$		关于直线 $x_2=x_1$ 的对称/镜像变换；图像与原图像重合，但方向相反	-1，面积变号

5	$\begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$		关于直线 $x_2=-x_1$ 的对称变换； 图像被翻转到原点的对面，但方向相反	-1，面积变号
6	$\begin{bmatrix}k & 0\\0 & 1\end{bmatrix}$		x_1 轴（水平）方向的伸缩变换； 图像在水平方向上被压缩或拉伸	k，面积改变 k 倍
7	$\begin{bmatrix}1 & 0\\0 & k\end{bmatrix}$		x_2 轴（垂直）方向的伸缩变换； 图像在竖直方向上被压缩或拉伸	k，面积改变 k 倍
8	$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$		x_1 轴（水平）方向的剪切变换； 图像被左剪切和右剪切	1，面积不变
9	$\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}$		x_2 轴（垂直）方向的剪切变换； 图像被上剪切和下剪切	1，面积不变
10	$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$		x_1 轴（水平）方向的投影变换； 图像被压缩为 x_1 轴上的单位向量线段	0，面积变为 0

11	$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$		x_2 轴（垂直）方向的投影变换； 图像被压缩为 x_2 轴上的单位向量线段	0，面积变为 0
12	$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$		逆时针旋转变换； 图像以原点为转轴逆时针转动	1，面积不变
13	$\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}$		顺时针旋转变换； 图像以原点为转轴顺时针转动	1，面积不变

由第三章可知，二阶行列式的几何意义就是两个列向量或行向量所构成的平行四边形的有向面积。实际上，二阶行列式的几何意义也可以看做两个列向量或行向量所构成三角形的有向面积的 2 倍（如表 5-6 中的图形）。看做平行四边形或者是三角形都可以，这不会影响我们对矩阵的理解。

读表提示

在读上表中的矩阵和变换图形时，请随时应用线性变换的矩阵定理进行印证，这样就会使你真正地掌握它们。

由表 5-6，我们得到与行列式有关的线性变换的一些结论作为总结：

- 行列式等于 1，表示变换后图形的面积乘以 1，面积大小不会变化，图形的有向性也不会变化（通俗地讲，图面的正反面不会反转）。
- 行列式等于-1，表示变换后图形的面积乘以-1，面积大小不会变化，但图形的有向性会变化（通俗地讲，图面的正反面发生反转）。
- 行列式等于 0，表示变换后图形的面积乘以 0，面积收缩为 0，图形缩变为一维图形。
- 行列式等于 k，表示变换后图形的面积会变化 k 倍。k 大于 1，图形面积会放大；k 小于 1，图形面积会缩小。图形的有向性也会因 k 是否为负值而发生变化。

5.5.4 初等矩阵/初等变换的几何意义

线性变换与矩阵之间有一一对应关系，因此就有初等矩阵所对应的初等变换。一组向量的线性相关性必然会体现在这组向量的数组之间的关系上，因此我们定能通过向量之间的线性运

算把这个关系揭示出来。由于线性运算是在同一分量上进行的，人们把一组向量并列在一起，便于对同一分量进行运算，这就是矩阵的初等变换。由此看来，初等变换只对表示向量组的矩阵才能看到其几何意义。在数学中，数域 **R** 上的矩阵的基本初等变换有三种：

（1）交换某两行 $\boldsymbol{i}$、$\boldsymbol{j}$ 的位置，记为 $T(\boldsymbol{i},\boldsymbol{j})$；

（2）把某一行 $\boldsymbol{i}$ 乘以一个非零数 $k\ (k\in\mathbf{R})$，记为 $T(\boldsymbol{i}(k))$；

（3）把某一行 $\boldsymbol{j}$ 的 $k\ (k\in\mathbf{R})$ 倍加到另一行 $\boldsymbol{i}$ 上，记为 $T(\boldsymbol{i},\boldsymbol{j}(k))$。

对于数域 **R** 上的一个单位矩阵分别实施上述三种基本初等变换，所得矩阵分别称为基本初等矩阵（1）、（2）、（3）。所以对于二阶单位矩阵 $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$：

基本初等矩阵（1）是 $\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$；

基本初等矩阵（2）分别是 $\begin{bmatrix}k & 0\\ 0 & 1\end{bmatrix}$ 或 $\begin{bmatrix}1 & 0\\ 0 & k\end{bmatrix}$；

基本初等矩阵（3）分别是 $\begin{bmatrix}1 & 0\\ k & 1\end{bmatrix}$ 或 $\begin{bmatrix}1 & k\\ 0 & 1\end{bmatrix}$。

基本初等矩阵与矩阵的基本初等变换也是一一对应的；任何一个可逆矩阵都可以分解成基本初等矩阵（1）、（2）、（3）的乘积。同时基本初等矩阵（1）可以由（2）和（3）导出。所以，任何一个可逆矩阵都可以分解成（2）和（3）这两种基本初等矩阵的乘积。从变换的角度来说，一个可逆的线性变换是连续实施若干次（2）和（3）两种基本初等变换的结果。

注意： 初等变换（2）就是线性变换的数乘变换；初等变换（3）就是线性变换的数乘后的加法，或者讲是线性组合。它们之所以基础就是因为它们是线性变换的数乘和加法的法则。

一个二阶矩阵作用在一个二维向量上得到一个新的向量，一个向量对应着一个点。因此，一个二阶矩阵把平面上的每一个点都变成唯一的点，从而它是平面到平面的映射（即变换），可以刻画平面上的几何变换。同理，任意一个三阶矩阵可以刻画空间中的几何变换。由于任何一个可逆矩阵都可以分解成基本初等矩阵（1）、（2）、（3）的乘积，因此，基本初等矩阵（1）、（2）、（3）的几何意义是理解一般矩阵所表示的变换的几何意义的基础。

基本初等矩阵（1）的几何意义是：关于某一“标准轴（面）”的镜像反射（对称）变换；基本初等矩阵（2）的几何意义是：在某一坐标轴方向的伸缩变换；基本初等矩阵（3）的几何意义是：在某一坐标轴方向的切变变换。

1. 基本初等矩阵（1）的几何意义

在二维和三维几何空间中，基本初等矩阵（1）所表示的几何变换的意义可以从线性变换矩阵的逆定理推得。实际上，表 5-6 的序号 4 中已经给出了二阶基本初等矩阵（1）的几何意义：以直线 $x_2=x_1$ 为对称轴的对称/镜像变换。

下面我们给出三维空间中基本初等矩阵（1）的几何意义。

如图 5-18 所示，先给出三阶单位矩阵 $\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$ 行向量张成的图形，就是第一象限的立方体；为了图像比对效果明显，特意画了单位行向量不等的图形（见图 5-18（a），把立方体故意画成

了长方体）。在 1，2 互换的换行矩阵$\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}$作用下，立方体仍然在第一象限，只是图形以平面$x_2=x_1$为对称面进行了镜像变换（见图 5-18（b））。

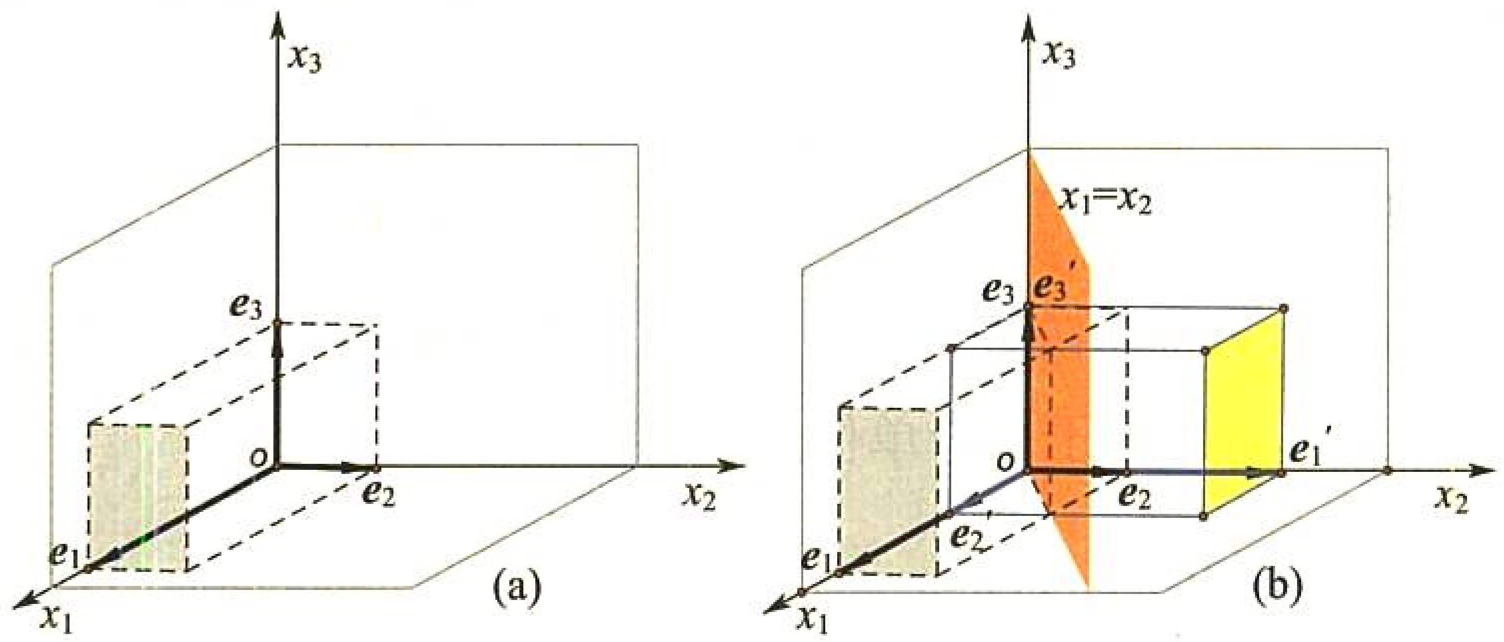


图 5-18 三阶初等矩阵（1）的几何意义

下面列举平面及空间里基本初等矩阵（1）对线段及点的变换的例题。

[平面线段变换例]：设$P(a,b)$是二维几何平面x_1ox_2上的任意一点（等同于任意向量），点P经基本初等矩阵$T(1,2)=\begin{pmatrix}0&1\\1&0\end{pmatrix}$变换后的结果为$P_1=T(1,2)P=(b,a)$（见图 5-19）。

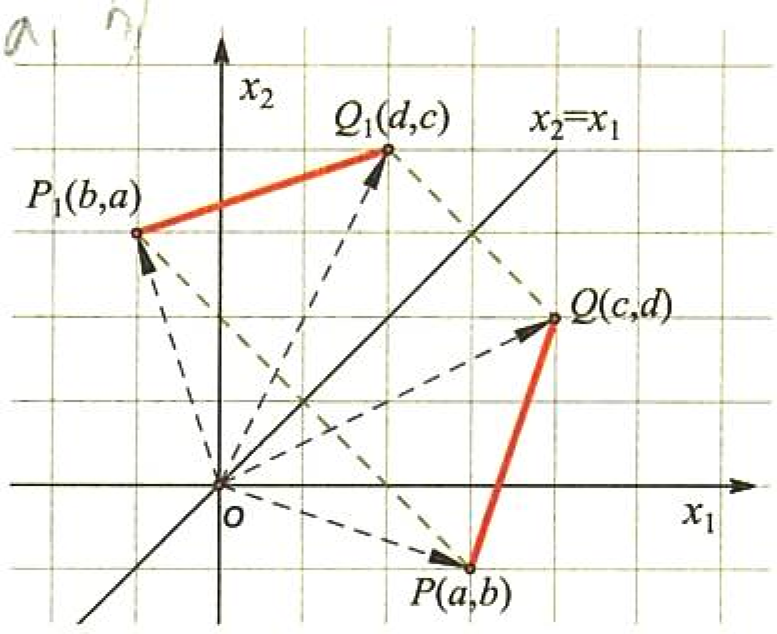


图 5-19 平面线段的初等变换 1

P_1是点P关于直线$x_2=x_1$的对称点。同样，Q_1是点Q关于直线$x_2=x_1$的对称点，由此可类推知线段P_1Q_1 是PQ关于直线$x_2=x_1$的对称图形。

[空间点变换例]：设$P(a,b,c)$是三维几何空间$ox_1x_2x_3$中任意一点，点P经基本初等矩阵$T(1,2)$、$T(2,3)$、$T(1,3)$分别变换后，所得结果为$P_1=T(1,2)P=(b,a,c)$，$P_2=T(2,3)P=(a,c,b)$，$P_3=T(1,3)P=(c,b,a)$。点P_1是点P关于平面$x_2=x_1$的对称点，点P_2是点P关于平面$x_2=x_3$的对称点，点P_3是点P关于平面$x_3=x_1$的对称点。

图 5-20 绘出了点P_1是点P关于平面$x\ =x$的对称点。

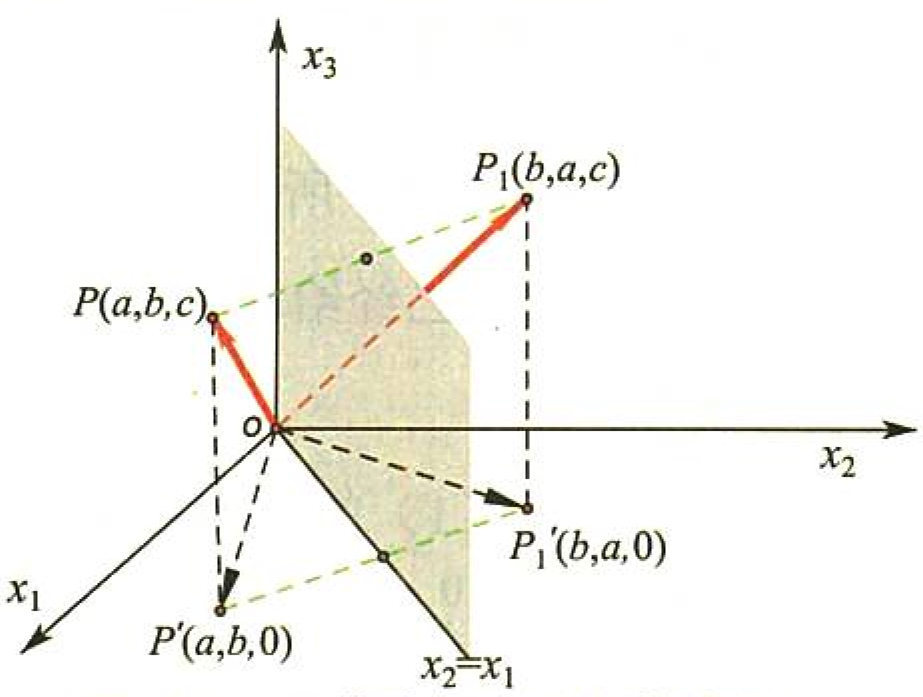


图 5-20 三维空间点的初等变换 1

这个结论可从二维的例子演推到，点 $P'(a,b,0)$ 和 $P_1'(b,a,0)$ 在二维平面 x_1ox_2 上关于直线 $x_2=x_1$ 镜像对称；扩展到三维空间后，就得到了 $P(a,b,c)$ 和 $P_1(b,a,c)$ 关于平面 $x_2=x_1$ 镜像对称。

2. 基本初等矩阵（2）的几何意义

在二维和三维几何空间中，基本初等矩阵（2）所表示的几何变换是：对图形实施的在某一坐标轴方向的伸缩变换。这个几何变换（2）的意义也可以从线性变换矩阵的逆定理推得。实际上，在表 5-6 的序号 6、7 中已经给出了二阶基本初等矩阵（2）的几何意义。

三维空间中基本初等矩阵（2）的几何意义我们容易想到，就是对立方体在某一坐标轴方向的伸缩变换。

下面列举平面及空间里基本初等矩阵（2）对线段及点的变换。

例如：设 $P(a,b)$ 是平面 x_1ox_2 上的任意一点，点 P 经基本初等矩阵 $T(1(k))=\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$ 和 $T(2(k))=\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$ 分别实施变换后，所得结果为 $P_1=T(1(k))P=(ka,b)$ 和 $P_2=T(2(k))P=(a,kb)$。P_1 是把点 P 在 x_1 轴方向的坐标伸缩 k 倍而在 x_2 轴方向的坐标不变而得到的，P_2 是把点 P 在 x_2 轴方向的坐标伸缩 k 倍而在 x_1 轴方向的坐标不变而得到的。

图 5-21 给出了两点 $P(a,b)$, $Q(c,d)$ 在 $T(1(2))$ 作用下沿 x_1 轴伸缩的情况。同时，线段 PQ 在 $T(1(2))$ 的作用下变化为 P_1Q_1。

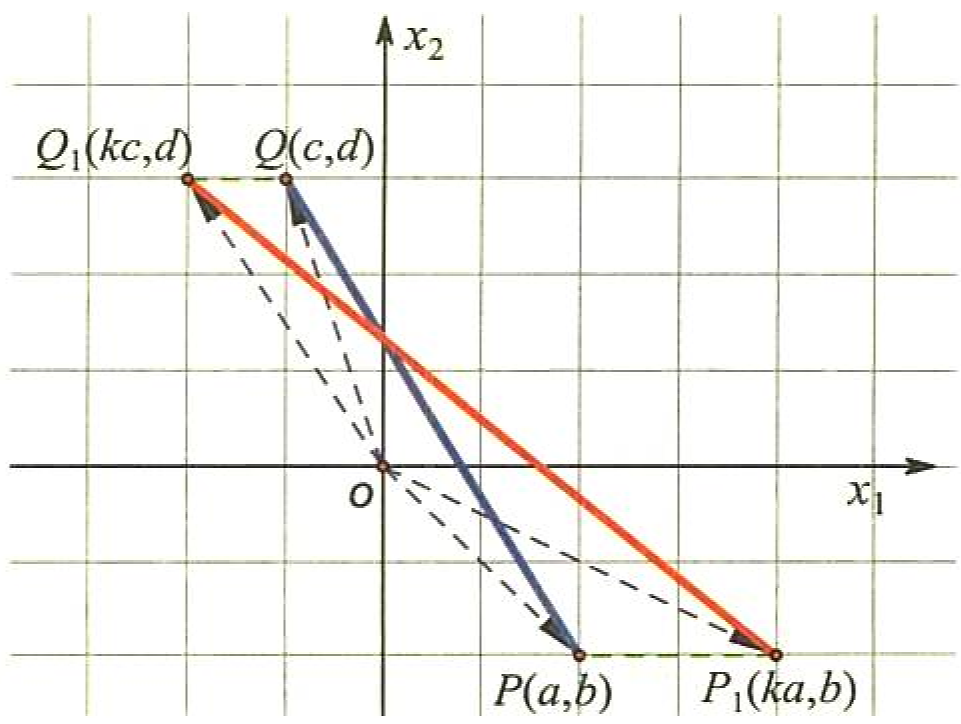


图 5-21 平面线段的初等变换 2

设 $P(a,b,c)$ 是三维几何空间中的任意一点，点 P 经三阶基本初等矩阵 $T(1(k))$、$T(2(k))$、$T(3(k))$（其中 $k\neq 0$）实施变换后，所得结果分别为 $P_1=T(1(k))P=(ka,b,c)$，$P_2=T(2(k))P=(a,kb,c)$，$P_3=T(3(k))P=(a,b,kc)$。P_1 是把点 P 在 x_1 轴方向的坐标伸缩 k 倍而在 x_2、x_3 轴方向的坐标不变而得到的；P_2 是把点 P 在 x_2 轴方向的坐标伸缩 k 倍而在 x_1、x_3 轴方向的坐标不变而得到的；P_3 是把点 P 在 x_3 轴方向的坐标伸缩 k 倍而在 x_1、x_2 轴方向的坐标不变而得到的。若 $k>0$，其伸缩方向与原坐标方向相同；若 $k<0$，其伸缩方向与原坐标方向相反。

作为特例，当 $k=0$ 时，基本初等矩阵（2）表示的几何变换是对图形实施的在某一坐标轴（面）上的投影变换；当 $k=-1$ 时，基本初等矩阵（2）表示的几何变换是对图形实施的关于某

一坐标轴（面）的镜像反射（对称）变换。

例如，矩阵 $T(1(0))=\begin{pmatrix}0&0\\0&1\end{pmatrix}$ 和 $T(1(-1))=\begin{pmatrix}-1&0\\0&1\end{pmatrix}$ 对点 $P(a,b)$ 实施变换后结果分别为：$P_1=T(1(0))P=(0,b)$，是点 $P(a,b)$ 在 x_2 轴上的投影点；$P_2=T(1(-1))P=(-a,b)$ 是点 $P(a,b)$ 关于 x_2 轴的反射点。

三阶矩阵 $T(1(0))$、$T(1(-1))$ 对点 $P(a,b,c)$ 实施变换后结果分别为：$P_1=T(1(0))P=(0,b,c)$，是点 $P(a,b,c)$ 在 x_2ox_3 平面上的投影点；$P_2=T(1(-1))P=(-a,b,c)$ 是点 $P(a,b,c)$ 关于坐标平面 x_2ox_3 的反射点。

3. 基本初等矩阵（3）的物理及几何意义

在二维和三维几何空间中，基本初等矩阵（3）所表示的几何变换是：对图形实施的在某一坐标轴方向的切变变换。这个几何变换（3）的意义也可以从线性变换矩阵的逆定理推得。实际上，在表 5-6 的序号 8、9 中已经给出了二阶基本初等矩阵（3）的几何意义。

切变在物理学中属于物体最基本的形变之一，在数学中很少使用切变这一概念，因此，对于基本初等矩阵（3）的几何意义需要结合物理中的切变来加以说明。

切变，在物理学中指由剪应力作用造成的扭曲（或变形）。如图 5-22 所示，当物体受到力偶作用使物体两个平行截面间发生相对平行移动时，在弹性力学中把这种形变叫做剪切形变，简称切变。切变的主要特征为：平行截面间相对滑移；只有纯粹的形状变化，而没有大小（面积或体积）的变化。

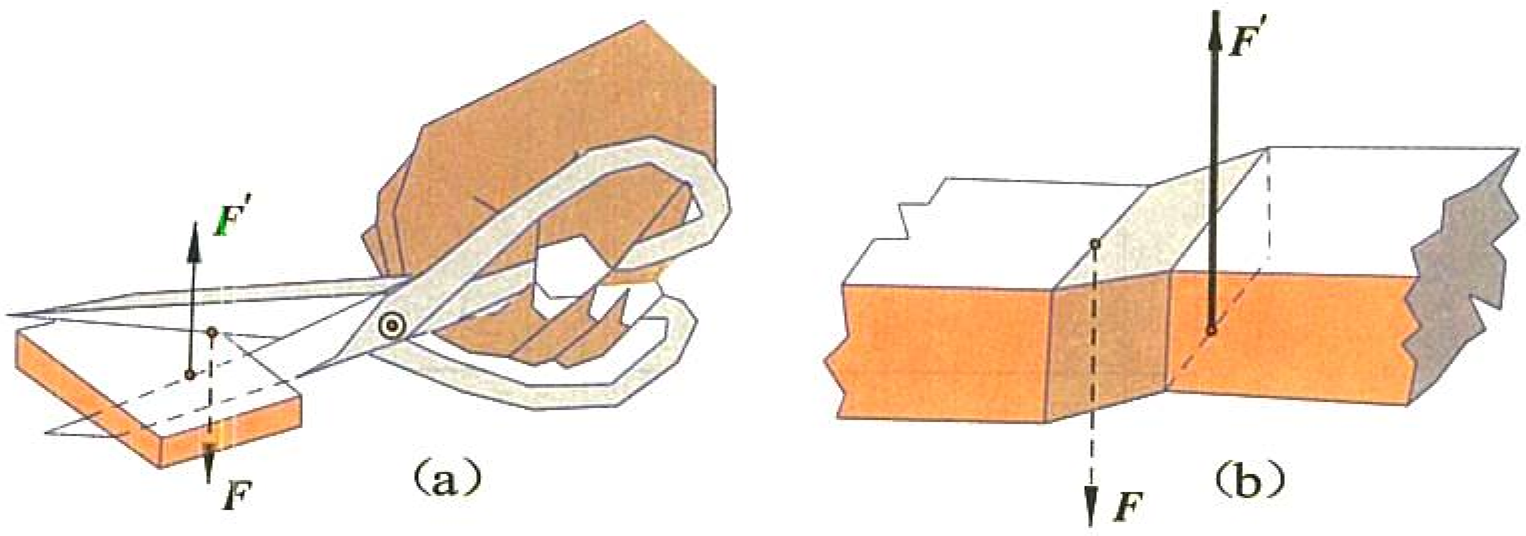


图 5 22 剪切形变图

切变的事例在日常生活中有很多。如图 5-23，将一本厚书放在桌面上，推动它的封面，使书页发生滑动，这时在书页的两头边缘上画出的一个矩形变成了平行四边形，这本书受到的就是切变。其形状发生了变化，而体积不变（这本书的宽度和厚度均没有发生变化）。

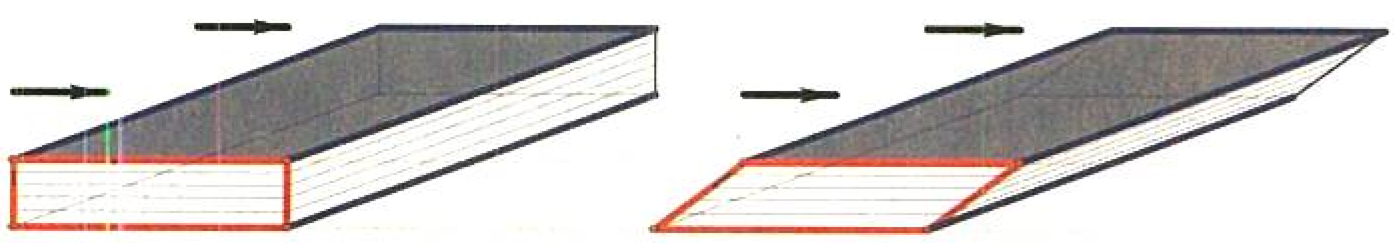


图 5-23 厚书切变图

本节咱们从切变变换推导出平面和空间里的切变矩阵，是为加强印象，理解线性变换与矩阵的互推关系。

1) 二维几何空间中的切变变换矩阵

如图 5-24（a），在平面上，一个正方形受到了一个 x_1 轴方向的水平切变力，切变表现为，其力偶作用由正方形变成一个等底、等高的平行四边形。其中，点 $P_1(x_1,x_2)$ 为正方形的任意一点，点 $P_2(x_1',x_2')$ 是切变后 P_1 点所对应的平移点。

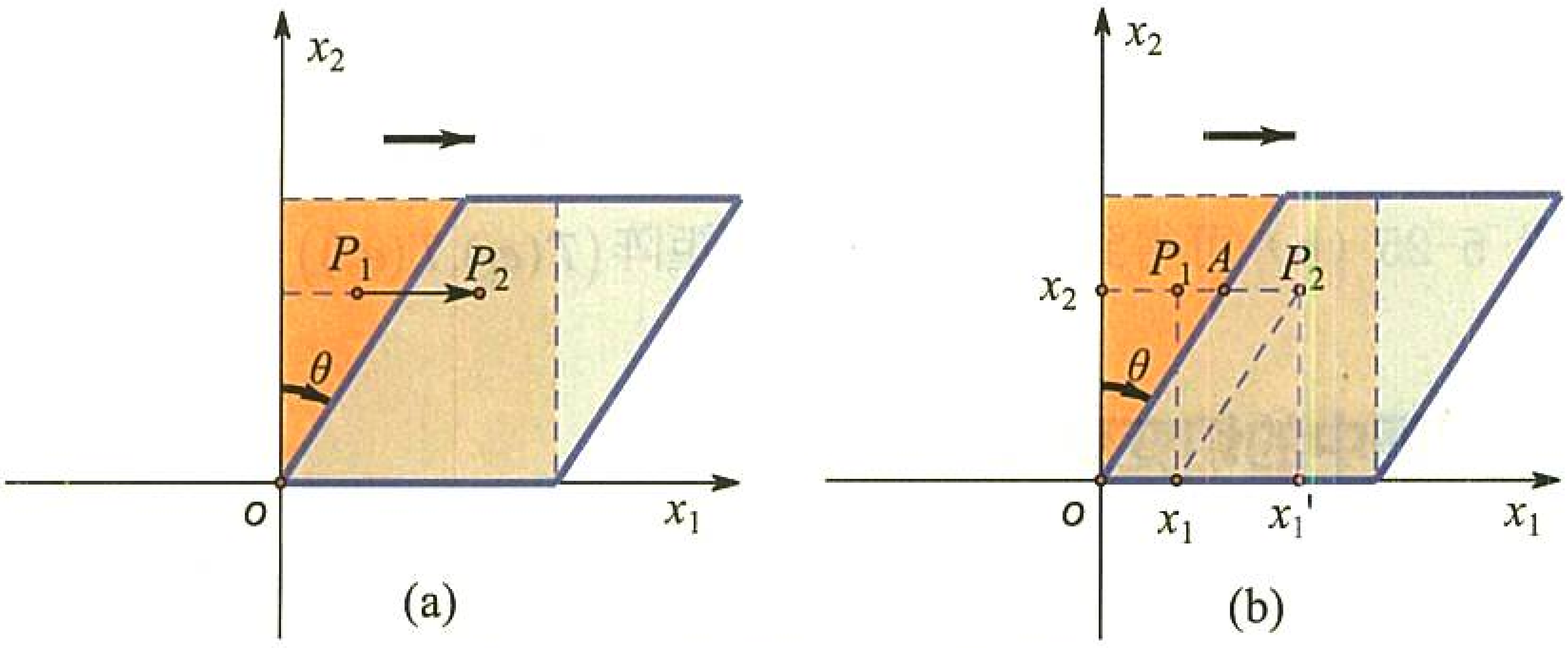


图 5-24 求解平面上的切变矩阵方法 1

由切变的意义，在从点 P_1 到点 P_2 的切变过程中，纵坐标 x_2（高度）不会发生变化，即 $x_2' = x_2$；横坐标 x_1 会向右（力的方向）平移一段距离 P_1P_2。如果把切变的变化量用一个变化的夹角来度量，就可以得到 $P_1P_2 = x_2\tan\theta$（见图 5-24（b），因为两个三角形全等：$\Delta oAx_2 \cong \Delta x_1P_2P_1$），因此横坐标的变化关系为 $x_1' = x_1 + x_2\tan\theta$。把 $\tan\theta$ 简写为一个变量 $k(k\in\mathbf{R})$，则有 $\begin{cases}x_1' = x_1 + kx_2\\ x_2' = x_2\end{cases}$，将其改写成矩阵/向量的表达式

$$\begin{pmatrix}x_1'\\x_2'\end{pmatrix}=\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

类似地，如果正方形切变的方向是竖直 x_2 轴的方向，那么变化前后的坐标表达式为

$$\begin{cases}x_1' = x_1\\ x_2' = kx_1 + x_2\end{cases} \text{或} \begin{pmatrix}x_1'\\x_2'\end{pmatrix}=\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

因此，矩阵 $\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$ 和 $\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}$ 是刻画平面上切变的数学模型，我们称其为平面上的切变变换矩阵。它们是把二阶单位矩阵的某一行的 k 倍加到另一行的基本初等矩阵 $T(\boldsymbol{i},\boldsymbol{j}(k))$。

实际上，利用线性变换的矩阵定理来快速推算这个切变矩阵，可能让你的思路更清晰简单。下面咱推一下：

先看图 5-25（a），设切变前的正方形为单位长度的，其两个边分别是单位向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$，则切变变换后，单位向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$ 分别变换为新向量 $T(\boldsymbol{e}_1)$ 和 $T(\boldsymbol{e}_2)$。由图易看出：向量 $\boldsymbol{e}_1$ 没有变化，所以 $T(\boldsymbol{e}_1)=\boldsymbol{e}_1=\begin{pmatrix}1\\0\end{pmatrix}$；向量 $\boldsymbol{e}_2$ 的高度没有变化，但在 x_1 方向移动了距离 k，从而有 $T(\boldsymbol{e}_2)=\begin{pmatrix}k\\1\end{pmatrix}$。

因此，有 x_1 方向的切变矩阵 $(T(\boldsymbol{e}_1),T(\boldsymbol{e}_2))=\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$。

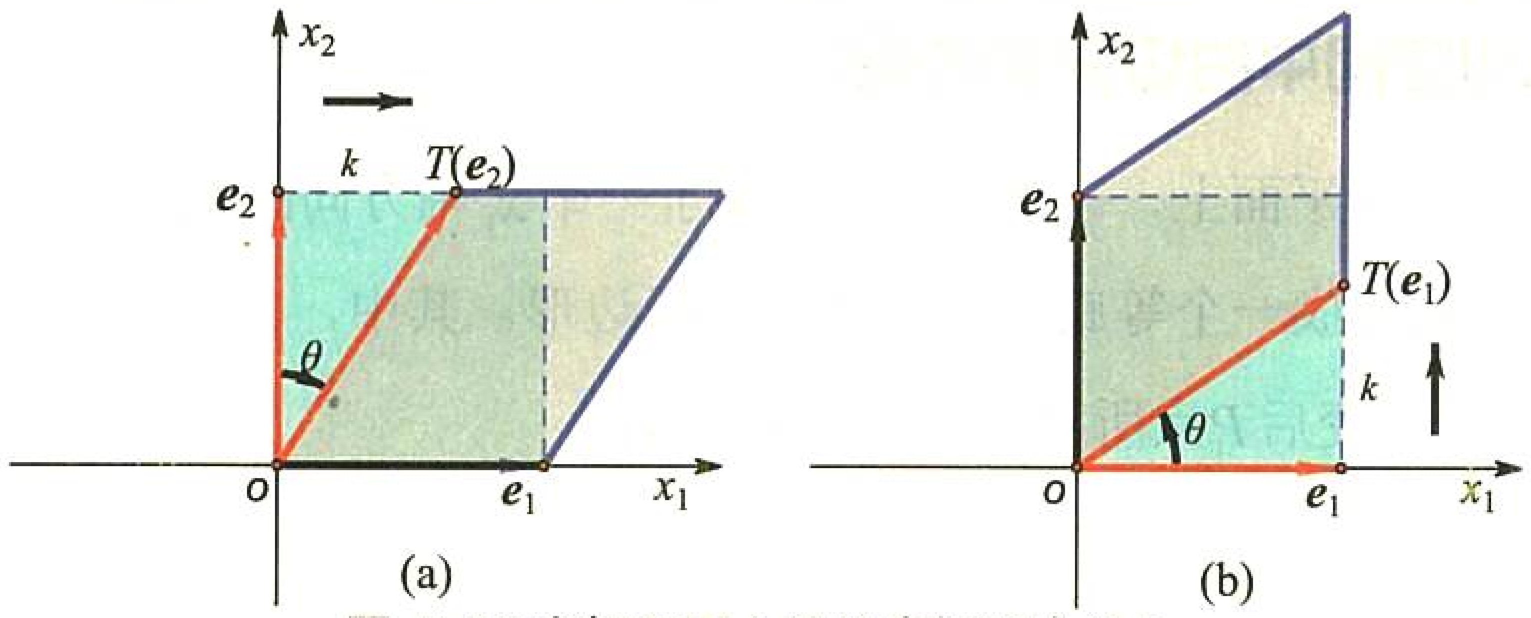


图 5-25 求解平面上的切变矩阵方法 2

类似地，由图 5-25（b）可得到 x_2 方向的切变矩阵 $(T(e_1),T(e_2))=\begin{bmatrix}1 & 0\\ k & 1\end{bmatrix}$。

2）三维几何空间中的切变变换矩阵

在三维几何空间中，一个正方体的切变表现为，其受到力偶作用由正方体变成一个等底、等高的平行六面体。在下面的例子中，我们假设正方体受到了平行于 x_1 或者 x_2 轴方向的切变作用力两种情况（见图 5-26）。

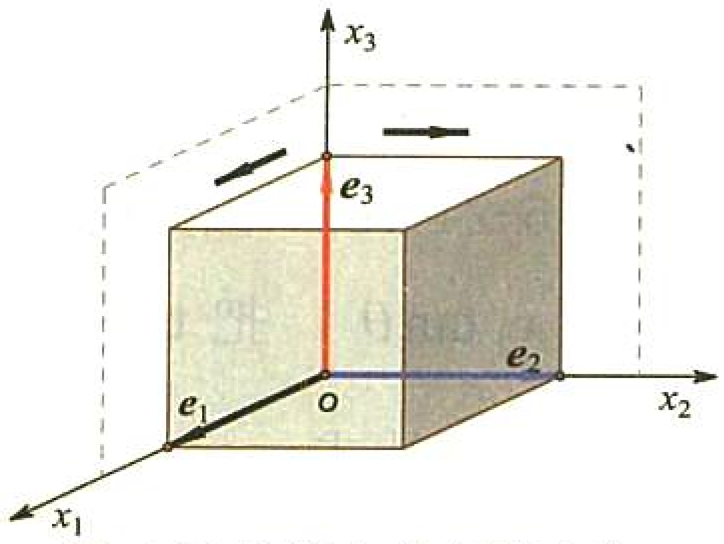


图 5-26 空间上的切变方向

当受到平行于 x_1 轴方向的切变作用力时（见图 5-27（a）），正方体内部的任意一点的坐标中 x_2 和 x_3 不会变，只有 x_1 坐标会发生变化。x_1 坐标的变化情况与二维平面上的切变类同。

当受到平行于 x_2 轴方向的切变作用力时（见图 5-27（b）），正方体内部的任意一点的坐标中 x_1 和 x_3 不会变，只有 x_2 坐标会发生变化。x_2 坐标的变化情况与二维平面上的切变类同。

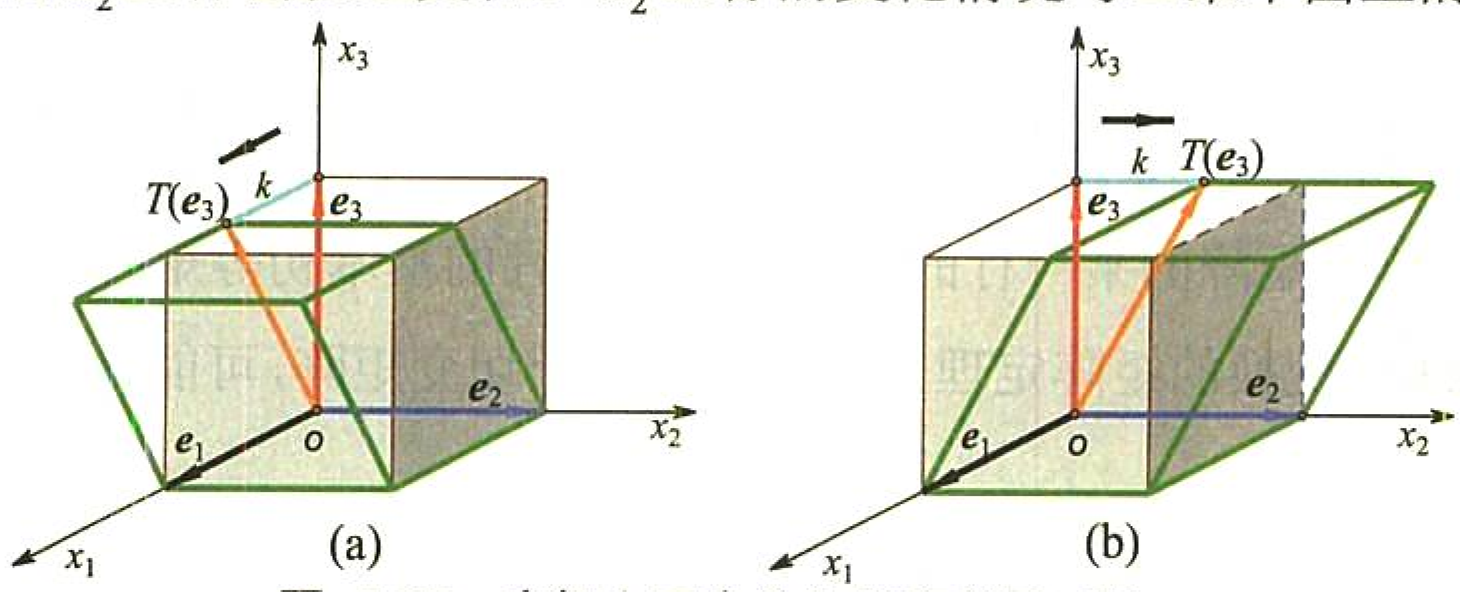


图 5-27 空间上正方体的切变前后对比

当我们应用线性变换的矩阵定理求解其矩阵时，就会发现这两种情况都是单位向量 $\boldsymbol{e}_3$ 发生了变化，而 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$ 没有变化。或者说，单位向量 $\boldsymbol{e}_3$ 只有两个坐标方向可以切变，要么 $\boldsymbol{e}_1$ 方向要么 $\boldsymbol{e}_2$ 方向。因此，不难得到图 5-27（a）所示的切变变换矩阵：

$$[T(\boldsymbol{e}_1),T(\boldsymbol{e}_2),T(\boldsymbol{e}_3)]=\left[\boldsymbol{e}_1,\boldsymbol{e}_2,\begin{pmatrix}k\\0\\1\end{pmatrix}\right]=\begin{bmatrix}1 & 0 & k\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$$

其中，k 是单位向量 $\boldsymbol{e}_3$ 在 x_1ox_3 平面上沿 x_1 轴方向的移动距离。

同理，图 5-27（b）所示的切变换矩阵为

$$[T(\boldsymbol{e}_1),T(\boldsymbol{e}_2),T(\boldsymbol{e}_3)]=\left[\boldsymbol{e}_1,\boldsymbol{e}_2,\begin{pmatrix}0\\k\\1\end{pmatrix}\right]=\begin{bmatrix}1&0&0\\0&1&k\\0&0&1\end{bmatrix}$$

其中，k 是单位向量 $\boldsymbol{e}_3$ 在 x_2ox_3 平面上沿 x_2 轴方向的移动距离。

至此，还有四种切变的情况没有得到矩阵。因为单位向量 $\boldsymbol{e}_1$ 也有两个坐标方向可以切变，要么 $\boldsymbol{e}_2$ 方向要么 $\boldsymbol{e}_3$ 方向；单位向量 $\boldsymbol{e}_2$ 亦有两个坐标方向可以切变，要么 $\boldsymbol{e}_1$ 方向要么 $\boldsymbol{e}_3$ 方向。这四种切变的情况如下：

$\boldsymbol{e}_1$ 沿 $\boldsymbol{e}_2$ 方向切变：

$$[T(\boldsymbol{e}_1),\boldsymbol{e}_2,\boldsymbol{e}_3]=\left[\begin{pmatrix}1\\k\\0\end{pmatrix},\boldsymbol{e}_1,\boldsymbol{e}_2\right]=\begin{bmatrix}1&0&0\\k&1&0\\0&0&1\end{bmatrix}$$

$\boldsymbol{e}_1$ 沿 $\boldsymbol{e}_3$ 方向切变：

$$[T(\boldsymbol{e}_1),\boldsymbol{e}_2,\boldsymbol{e}_3]=\left[\begin{pmatrix}1\\0\\k\end{pmatrix},\boldsymbol{e}_1,\boldsymbol{e}_2\right]=\begin{bmatrix}1&0&0\\0&1&0\\k&0&1\end{bmatrix}$$

$\boldsymbol{e}_2$ 沿 $\boldsymbol{e}_1$ 方向切变：

$$[\boldsymbol{e}_1,T(\boldsymbol{e}_2),\boldsymbol{e}_3]=\left[\boldsymbol{e}_1,\begin{pmatrix}k\\1\\0\end{pmatrix},\boldsymbol{e}_2\right]=\begin{bmatrix}1&k&0\\0&1&0\\0&0&1\end{bmatrix}$$

$\boldsymbol{e}_2$ 沿 $\boldsymbol{e}_3$ 方向切变：

$$[\boldsymbol{e}_1,T(\boldsymbol{e}_2),\boldsymbol{e}_3]=\left[\boldsymbol{e}_1,\begin{pmatrix}0\\1\\k\end{pmatrix},\boldsymbol{e}_2\right]=\begin{bmatrix}1&0&0\\0&1&0\\0&k&1\end{bmatrix}$$

因此，以上六个变换矩阵是把三阶单位矩阵的某一行的 k 倍加到另一行的基本初等矩阵 $T(\boldsymbol{i},\boldsymbol{j}(k))$。作为特例，若 $k=0$，则基本初等矩阵（3）为单位矩阵，其几何意义是图形到其自身的恒等变换。

5.6　矩阵乘法运算律的几何意义

与前面的章节比较起来，本节利用了线性变换的概念来再次回顾矩阵乘积运算的几何意义，由此可更深入地认识矩阵运算的性质。

5.6.1　两个矩阵相乘是两个线性变换的复合

例如，对于平面坐标系 x_1ox_2 上由单位矩阵列向量所张成的第一象限的正方形实施关于 x_1 轴的反射变换 $\boldsymbol{A}=\begin{bmatrix}1&0\\0&-1\end{bmatrix}$（见图 5-28（a））再实施关于标准轴 $x_2=x_1$ 的反射变换 $\boldsymbol{B}=\begin{bmatrix}0&1\\1&0\end{bmatrix}$，得到的图形正好是将原来的正方形逆时针旋转 90° 所得到的（见图 5-28（b））。

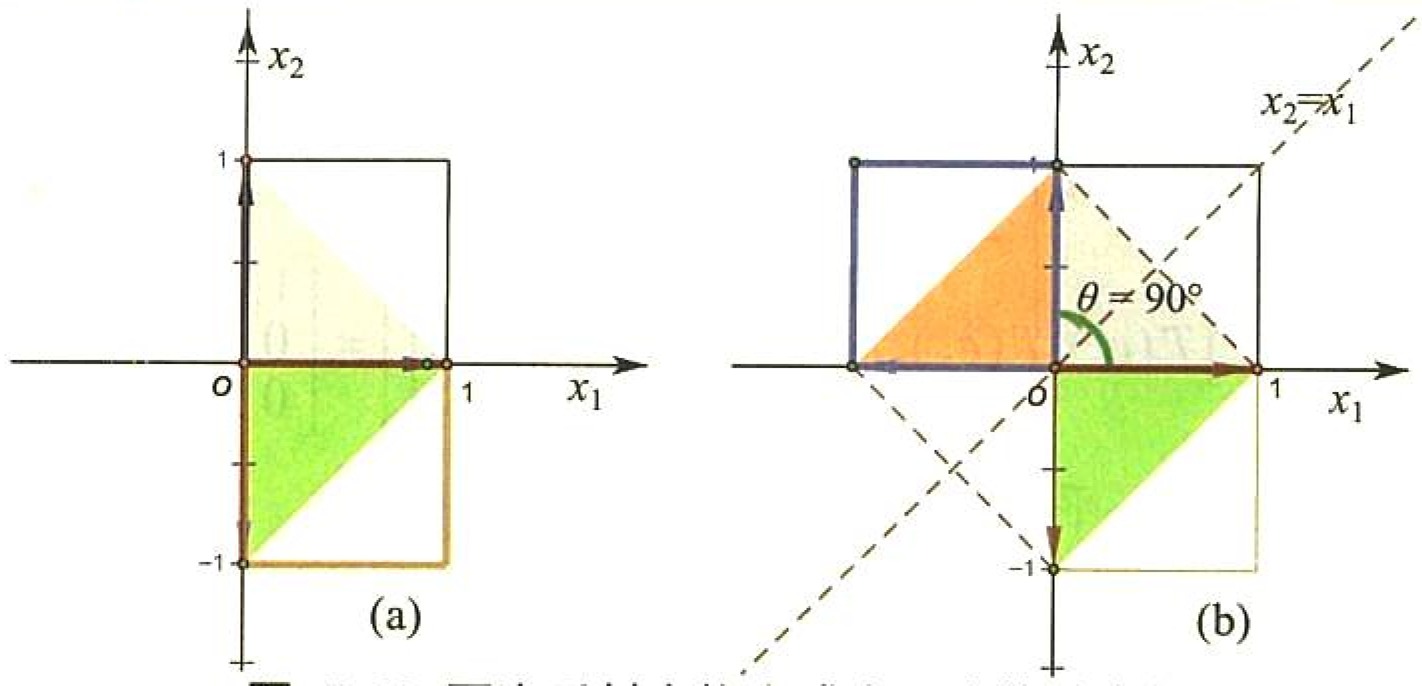


图 5-28 两次反射变换合成为一次旋转变换

这一结果用矩阵运算表示为 $\boldsymbol{BA}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}1&0\\0&-1\end{bmatrix}=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$，即两个矩阵相乘得到一个新矩阵，新矩阵所表示的变换正是原来两个矩阵表示的变换的复合。这正是矩阵乘法的意义，即两个矩阵相乘表示连续两次实施变换。这一特例同时也可以帮助大家认识“旋转变换可以由连续两次反射变换来实现”这一性质。

5.6.2 矩阵的乘法不满足交换律

根据上述对矩阵乘法的理解，容易直观得出矩阵运算满足结合律。是否满足交换律呢？仍然看上面的例子。

如果先对第一象限的正方形实施关于标准轴 $x_2=x_1$ 的反射变换 $\boldsymbol{B}$，再实施关于 x_1 轴的反射变换 $\boldsymbol{A}$，得到 $\boldsymbol{AB}=\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$，矩阵 $\boldsymbol{AB}$ 的变换是将原来的正方形顺时针旋转 90°。这与前面的 $\boldsymbol{BA}$ 变换的结果不同，因而 $\boldsymbol{BA}\neq\boldsymbol{AB}$。

我们再看个例子。同样地，对在第一象限由单位矩阵列向量所张成的正方形图形，先向靠近 x_2 轴的方向压缩一半再逆时针旋转 90° 得到的结果（见图 5-29（a））与先逆时针旋转 90° 再向靠近 x_2 轴的方向压缩一半得到的结果（见图 5-29（b））的比较，大家应确信交换线性变换的顺序得到的结果一般来说是不同的。即

$$\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}0.5&0\\0&1\end{bmatrix}\neq\begin{bmatrix}0.5&0\\0&1\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix}$$

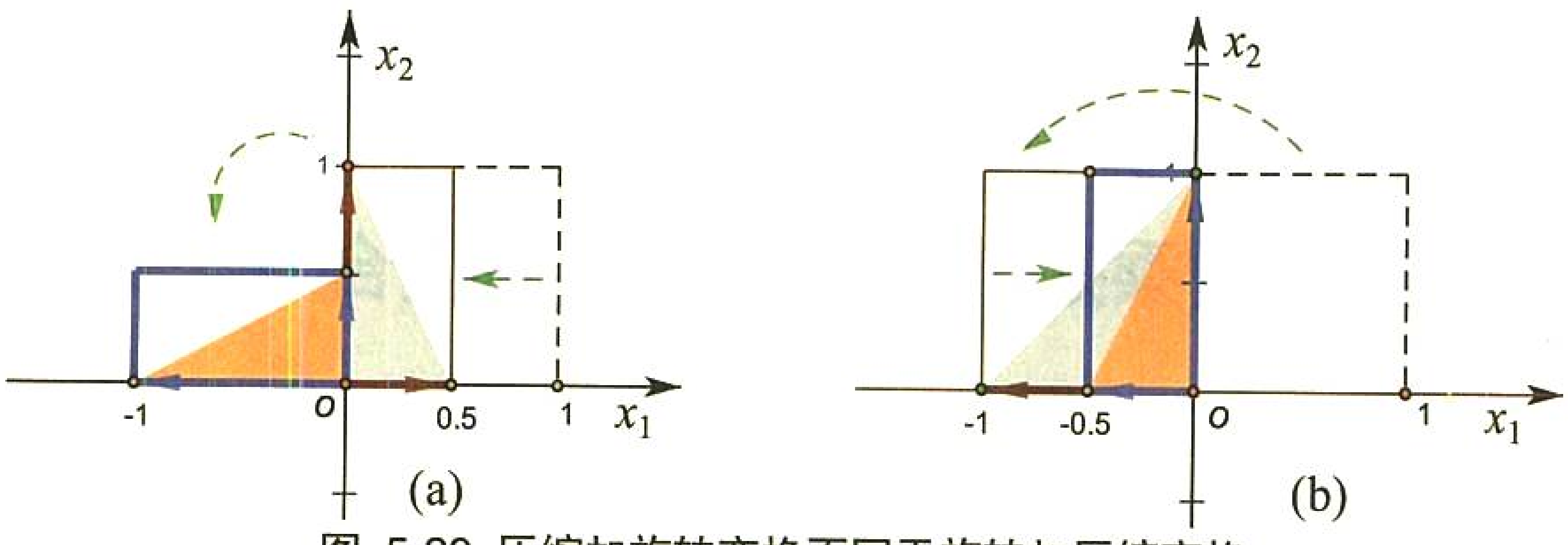


图 5-29 压缩加旋转变换不同于旋转加压缩变换

因此，矩阵的乘法一般不满足交换律。但需要注意的是，有些情形矩阵的乘法可以满足交换律，例如连续多次旋转或连续多次压缩变换是可以交换顺序的。

5.6.3 矩阵的乘法不满足消去律

根据投影变换的特点，把一个图形先作关于 x_2 轴的反射变换，再向 x_1 轴作投影变换的结果（见图 5-30（a）），与先作关于坐标原点的对称变换，再向 x_1 轴作投影变换得到的结果（见图 5-30（b））是一样的。

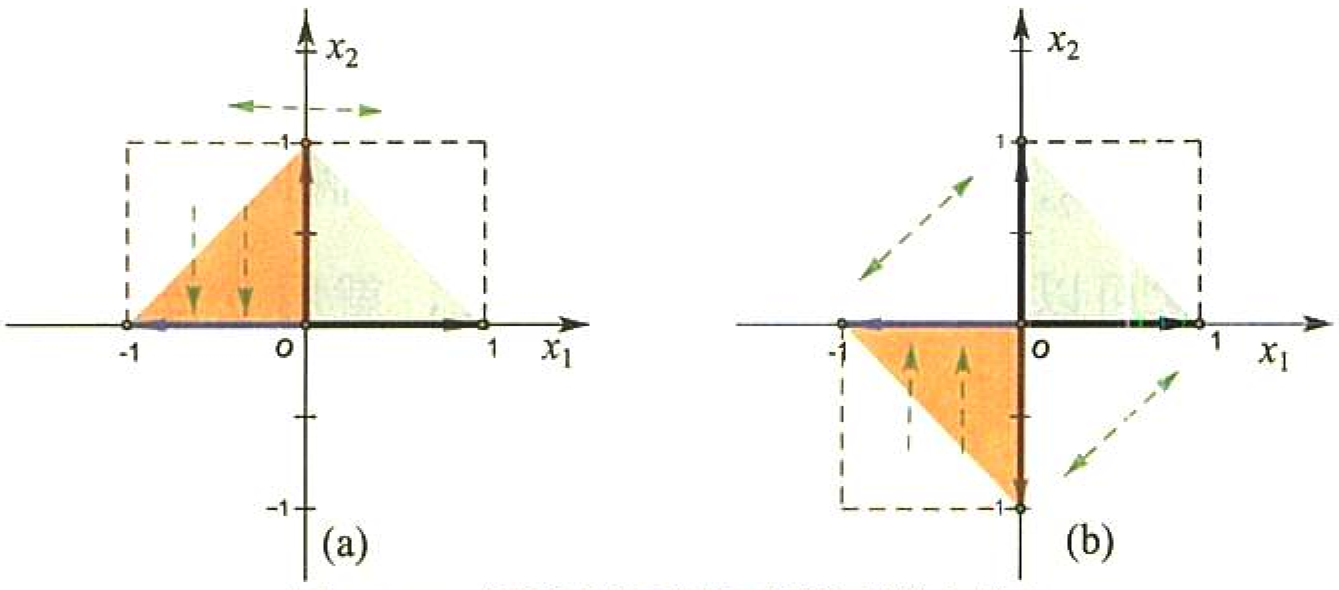


图 5-30 矩阵的乘法不满足消去率

上述的变换矩阵等式为$\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix}=\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$。如果消去等式两边的共同项，会得到错误的等式$\begin{bmatrix}-1&0\\0&1\end{bmatrix}=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$。因此，矩阵的乘法不满足消去律。

5.7 矩阵秩的几何意义

5.7.1 矩阵秩的几何意义

$\boldsymbol{A}$ 的秩一般记为 $R(\boldsymbol{A})$。

矩阵秩的实质就是矩阵行向量组的秩和列向量组的秩的问题。行向量组张成了行空间，列向量组张成了列空间。这里有个可以称为矩阵秩的几何意义的结论：

任何矩阵行空间的维数等于列空间的维数等于这个矩阵的秩。

如果是一个满秩的方阵，大家都同意上面的说法，因为方阵的行数等于列数，秩都相等啊。

如果是一个不满秩的方阵呢？

在回答这个问题之前，先看一个“不方”的矩阵（俺的意思是行数不等于列数的矩阵）：

$$\begin{bmatrix}a_1&b_1\\a_2&b_2\\a_3&b_3\end{bmatrix}$$

看起来，这个矩阵的行空间是由三个行向量(a_1, b_1), (a_2, b_2), (a_3, b_3)所张成的。等等，怎么向量是二维向量啊，三个二维向量肯定是线性相关的。二维向量最多能张成一个二维的空间。

Ok，假设是行向量可以张成二维空间，就是行空间的维数是 2。

再看列向量，列向量有两个：$(a_1, a_2, a_3)^{\mathrm{T}}$和$(b_1, b_2, b_3)^{\mathrm{T}}$。两个三维的向量能张成几维的空间？最多只能张成两维的空间。注意兄弟，这个二维的空间是三维空间中的子空间。还不明白的话就回到向量组、向量空间的章节中复习复习吧。

好吧，列向量也是张成一个维数是 2 的列空间（必然，因为前面有个行空间维数是 2 的假

设，这句话暂不明白没关系，继续往下看)，两个空间的维数相等，因此 2 就是这个矩阵的秩了。

现在回答前面关于一个不满秩的方阵的秩的问题就容易了。举个三阶的方阵：

$$\left[\begin{array}{c:c:c} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{array}\right]$$

不满秩，意思不是行向量组线性相关就是列向量组线性相关。假设列向量组线性相关，列向量 $(c_1,c_2,c_3)^{\mathrm{T}}$ 可以被列向量 $(a_1,a_2,a_3)^{\mathrm{T}}$ 和 $(b_1,b_2,b_3)^{\mathrm{T}}$（假设这俩向量线性无关）线性表示。列向量组里面的向量 $(c_1,c_2,c_3)^{\mathrm{T}}$ 就可以删除了，在方阵里面呢，就用 0 替代它，即

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \sim \begin{bmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & 0 \end{bmatrix}$$

看后面的矩阵，这实际上又回到了前面不方的矩阵的秩的问题上了。三个行向量 $(a_1, b_1, 0)$、$(a_2, b_2, 0)$、$(a_3, b_3, 0)$ 在一个二维平面上，因此哥仨也肯定线性相关了。假设行向量 $(a_3, b_3, 0)$ 可以被行向量 $(a_1, b_1, 0)$ 和 $(a_2, b_2, 0)$ 线性表示，那么矩阵继续等价下去：

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \sim \begin{bmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & 0 \end{bmatrix} \sim \begin{bmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Ok，看到没，非零的行等于非零的列，行秩等于列秩。

有哥们看出问题来了，要是这俩行向量 $(a_1, b_1, 0)$ 和 $(a_2, b_2, 0)$ 继续线性相关呢?我说这是不可能的。反证法：假设这两位线性相关，好，再删去一个行向量，矩阵就变成

$$\begin{bmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} a_1 & b_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

问题来了，看看列向量，第一列和第二列线性相关了，这和前面的假设“向量 $(a_1,a_2,a_3)^{\mathrm{T}}$ 和 $(b_1,b_2,b_3)^{\mathrm{T}}$ 线性无关”矛盾。

一个结论：方阵里面，有几个行向量是多余的，就会有几个列向量是多余的。对于所有的矩阵，不相关的行的数等于不相关的列的数，就是秩数。

好了，上述的说法，是把列和行都看做向量，因此有张成行空间和列空间的说法。实际上这种传统的几何解释的说法仍然让人不易理解。其实比较彻底的解释应该是这样的：对一个 n 阶方矩阵，仍然从传统上**把列看做列向量，则行是每个列向量在列空间各个坐标轴上的投影(坐标)，行的数量则是列空间坐标系的维数**。方阵有 m 个不相关的列向量，就应该张成 m 维的列子空间；反过来，m 维的子空间里的向量只需要 m 个坐标轴，m 个坐标轴就是矩阵 m 个行数。因此就有了前面的说法：不相关的列的个数等于不相关的行的数，就是秩数 m。

因此，矩阵的行向量组的秩、矩阵的列向量组的秩以及矩阵中不等于 0 的子式的最高阶数（可称为矩阵的行列式秩）是相等的，三秩合一，统称为矩阵的秩。

5.7.2 矩阵的秩对图形变换的影响

二阶矩阵对于平面上单位圆上的向量变换之几何图形我们已有概念，前面的例子中矩阵把单位圆变换成了椭圆。是不是所有的二阶矩阵都是这样的？非也。我们可以用矩阵的秩来区别。

前面咱谈过，矩阵乘以一个向量就是把这个向量映射到列空间中去，同时矩阵的秩也是其列空间的维数。如果矩阵乘以一个图形（向量的集合），就会把这个图形统统变换到列空间去。因此啊，变换后的图形维数最大不能超过列空间的维数。举几个例子：

对于二阶矩阵 $\boldsymbol{A}=\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$，其行列式 $\det(\boldsymbol{A})=\begin{vmatrix}1 & 1.5\\0 & 1\end{vmatrix}=-0.5\neq 0$，因此 $R(\boldsymbol{A})=2$；

对于二阶矩阵 $\boldsymbol{B}=\begin{bmatrix}1 & 2\\0.5 & 1\end{bmatrix}$，其行列式 $\det(\boldsymbol{B})=\begin{vmatrix}1 & 2\\0.5 & 1\end{vmatrix}=0$，故 $R(\boldsymbol{B})<2$，进而得到 $R(\boldsymbol{B})=1$；

矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 对单位圆上向量的变化图形如图 5-31 所示。

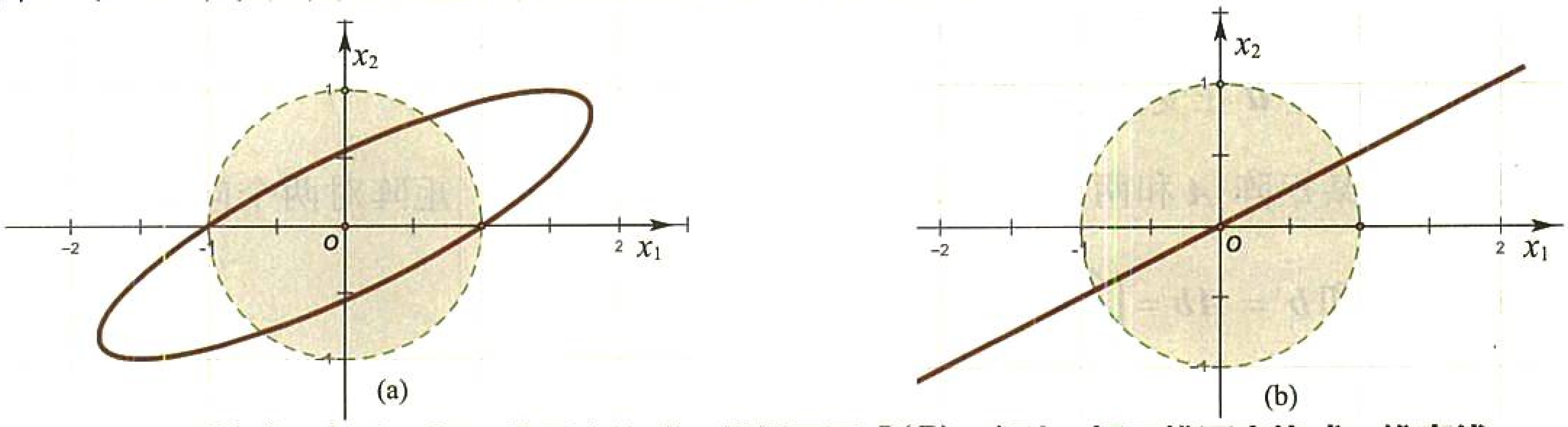


图 5-31　$R(A)=2$ 则 A 把二维圆变换成二维椭圆及 $R(B)=1$ 则 B 把二维圆变换成一维直线

显然，当二阶矩阵的秩等于 2 时，矩阵把二维的圆形仍然变换为二维的椭圆图形；当矩阵的秩降低为 1 时，矩阵变换后的图形退化为一维的线段。

同样地，三阶矩阵对于空间球面上的向量的变换图形是：

当矩阵的秩等于 3 时，单位球被变换成三维空间中的一个椭圆球（三维图形）；

当矩阵的秩等于 2 时，单位球被变换成三维空间中的一个椭圆面（二维图形）；

当矩阵的秩等于 1 时，单位球被变换成三维空间中的一个线段（一维图形）。

5.8 矩阵特征值和特征向量的几何及物理意义

当 $\boldsymbol{A}$ 是一个方阵，一个向量 $\boldsymbol{a}$ 被 $\boldsymbol{A}$ 相乘的结果和偶尔被一个数 λ 相乘的结果，恰好相等。即

$$\boldsymbol{A}\boldsymbol{a}=\lambda\boldsymbol{a} \tag{5-7}$$

那么向量 $\boldsymbol{a}$ 就称为 $\boldsymbol{A}$ 的特征向量，λ 称为 $\boldsymbol{A}$ 的特征值。这是定义。下面我们将分别讨论特征值和特征向量的几何及物理意义。

5.8.1 特征值和特征向量的几何意义

1. 特征值和特征向量的几何意义

方阵乘以一个向量的结果仍是一个同维向量，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。在这个变换的过程中，向量会发生旋转、伸缩或镜像的变化。矩阵不同，

向量变化的结果也会不同。如果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效果，那么这些向量就是这个矩阵的特征向量，伸缩的比例就是特征值；如果伸缩的比例值是负值，原向量的方向改变为反方向，原向量仍然是这个矩阵的特征向量。

对于前述的二维旋转矩阵，比如是逆时针旋转 $\pi/2$ 弧度的矩阵，这时我们可以问一个问题：在平面上的所有向量中，有没有向量在这个矩阵变换下只沿着某根直线缩放或振动而不旋转到其他方向呢？可以想一下，除了零向量，没有其他向量可以在平面上旋转 $\pi/2$ 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）。因此，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在某个直线上不转动，只是进行长度上的伸缩而已。

再返回来盯着特征向量的定义“ $\boldsymbol{Aa}=\lambda\boldsymbol{a}$ ”，目不转睛保持 3 分钟……没顿悟？棒喝：不要忘了 $\boldsymbol{A}$ 就是一个变换。再回来对着定义式目不转睛……，这时你豁然顿悟了：$\lambda\boldsymbol{a}$ 是方阵 $\boldsymbol{A}$ 对向量 $\boldsymbol{a}$ 进行变换后的结果，但显然 $\lambda\boldsymbol{a}$ 和 $\boldsymbol{a}$ 是在一条直线上（方向相同或相反）。特征值 λ 只不过反映了特征向量 $\boldsymbol{a}$ 在变换时的伸缩倍数而已。

例如，设某矩阵 $\boldsymbol{A}$ 和两个向量 $\boldsymbol{a}=\begin{pmatrix}-1\\1\end{pmatrix}$、$\boldsymbol{b}=\begin{pmatrix}2\\1\end{pmatrix}$，那么，矩阵对两个向量的变换结果为 $\boldsymbol{a}'=\boldsymbol{Aa}=\begin{pmatrix}-5\\-1\end{pmatrix}$ 和 $\boldsymbol{b}'=\boldsymbol{Ab}=\begin{pmatrix}4\\2\end{pmatrix}$，几何图形如图 5-32 所示。

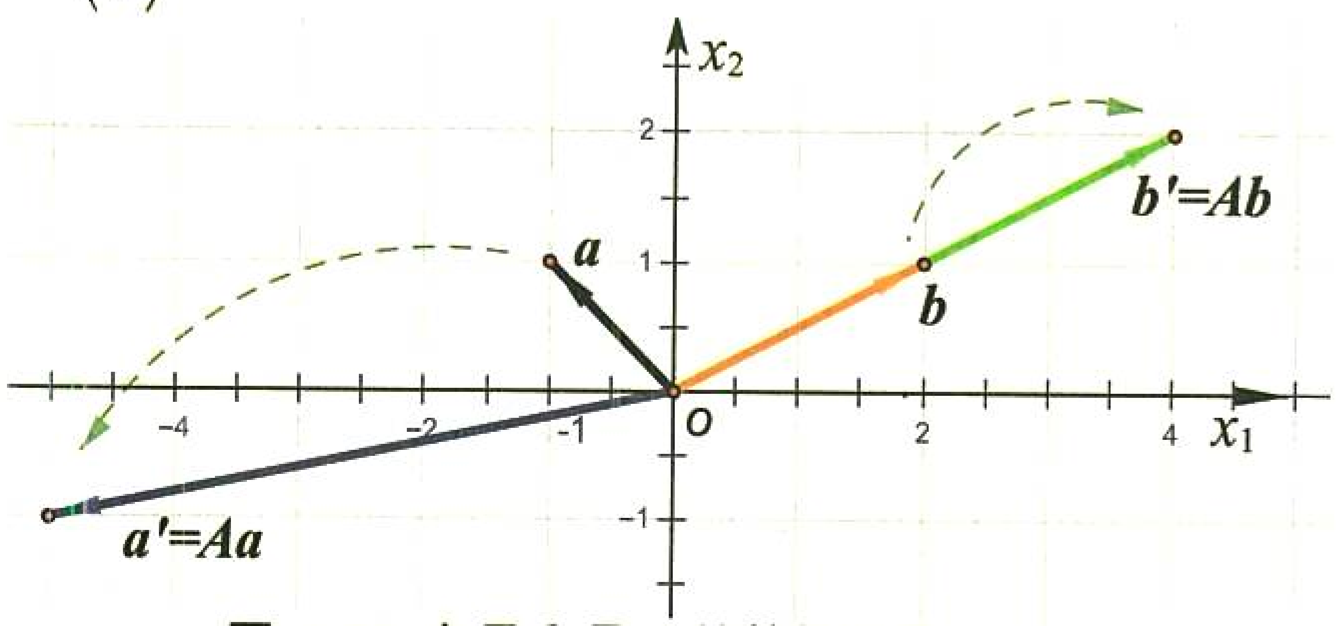


图 5-32 向量 b 是 A 的特征向量

显然，只有向量 $\boldsymbol{b}$ 被矩阵 $\boldsymbol{A}$ 同方向拉长了 2 倍，即 $\boldsymbol{Ab}=2\boldsymbol{b}$，因此向量 $\boldsymbol{b}$ 是矩阵 $\boldsymbol{A}$ 的特征向量，特征值为 2。

对于矩阵 $\boldsymbol{A}$，大多数向量 $\boldsymbol{a}$ 不满足 $\boldsymbol{Aa}=\lambda\boldsymbol{a}$ 这样的一个方程。因为当向量 $\boldsymbol{a}$ 被矩阵相乘时几乎都将改变 $\boldsymbol{a}$ 的方向，因此 $\boldsymbol{Aa}$ 和 $\boldsymbol{a}$ 常常不是倍数的关系。这意味着只有某些特殊的数 λ 是特征值，而且只有某些特殊的少数向量是特征向量(有的矩阵干脆一个没有)。当然有个例外，如 $\boldsymbol{A}$ 是单位矩阵或单位矩阵的倍数，那就没有向量被改变方向，从而所有的向量都是特征向量。

因此，从矩阵的几何意义来看，矩阵 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{a}$ 就是经过矩阵 $\boldsymbol{A}$ 变换后与自己平行(方向相同或者相反)的非零向量，矩阵 $\boldsymbol{A}$ 的特征值 λ 就是特征向量 $\boldsymbol{a}$ 经变换后的伸缩系数。

注意：

上面我们讲的是实特征值的几何意义。某一个实特征值使特征向量在一根直线上进行伸缩变换，而复特征值会使特征向量在复平面上进行旋转，但在实轴上仍然只是进行伸缩变换。为简单化，最后一节才讲复特征值的几何意义。下面不特别说明时，一般指的是实特征值的意义。

下面我们看一些矩阵对于一个几何图形的变换过程中关于特征值和特征向量的意义。

2. 特征向量与矩阵变换的关系

如前述平面的剪切矩阵$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$，当 k=1.5 时，矩阵为$\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$，这个矩阵对三角形向量群的变化如图 5-33 所示。

显然，直角三角形 ΔoAG 内部的无数向量被水平方向切变到 $\Delta oA'G'$ 钝角三角形内部的向量。图中，原向量与被变换的向量由一根根虚线段 ($AA',BB',\cdots,GG'$) 连接，因为是水平切变，因此这根虚线段所代表的向量差是水平的。

$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$

$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$

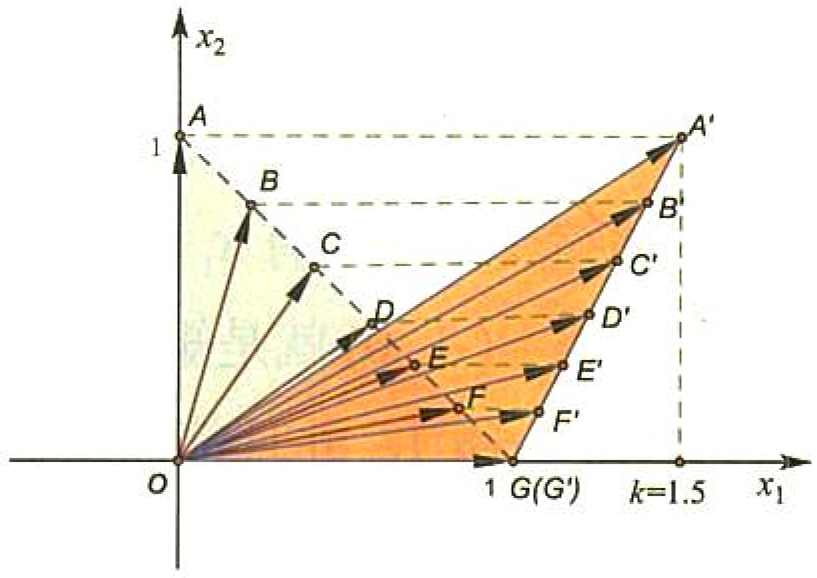


图 5-33 三角形里面的矩阵特征向量

显然，这些虚线段 $AA',BB',\cdots,GG'$ 所表示的向量变化量是由矩阵$\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$带来的；这些变换量的特点是 AA' 线段最长，BB' 线段长度次之，……直到 GG' 线段长度变为 0。这时原向量 $\overrightarrow{oG}$（x坐标单位向量）和变换向量 $\overrightarrow{oG'}$ 重合。由特征值和特征向量的定义可知，向量 $\overrightarrow{oG}=(1, 0)$ 是矩阵$\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$的特征向量，因为这个特征向量的变化量为 0，原向量与被变换向量相等，变换的比例为 1，所以特征值是 1。

显然上述对特征向量和特征值的讨论无论是三角形还是正方形都不全面，因为我们只是讨论了平面上的第一象限的向量，平面上还有四分之三的向量没有在图中画出来（见图 5-34(a)）。

后向量

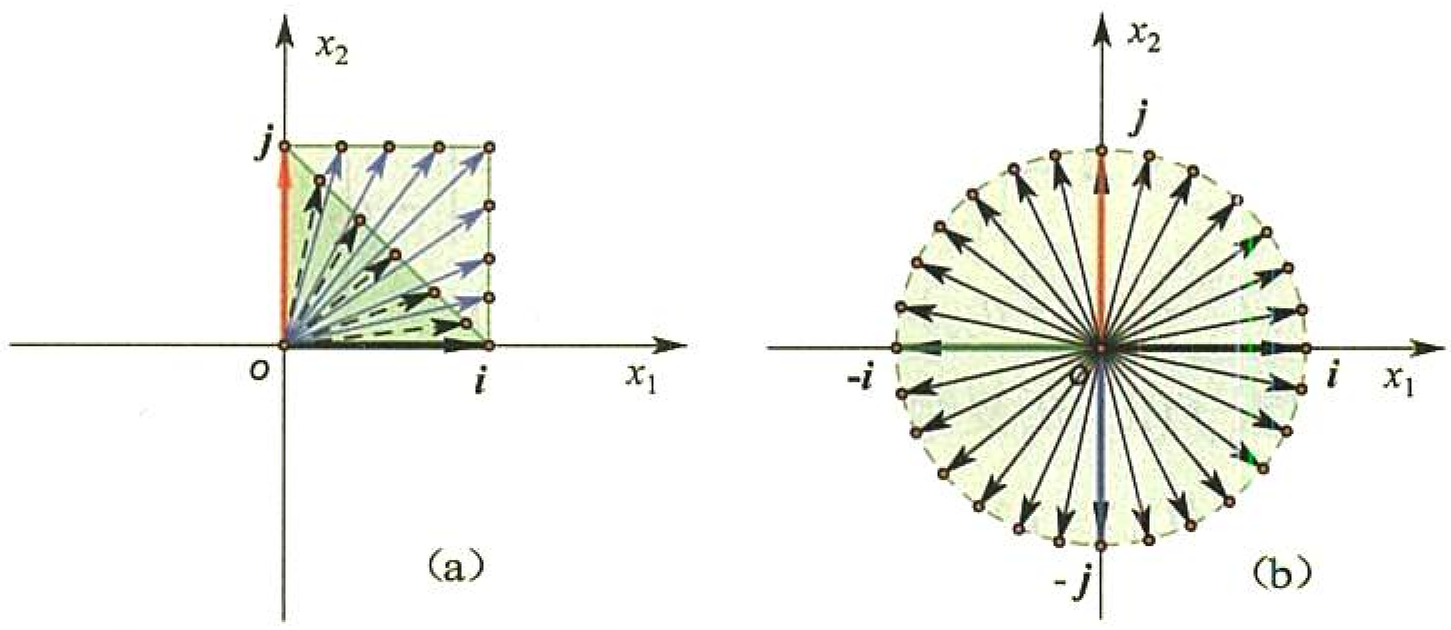


图 5-34 二阶矩阵的特征向量要讨论整个平面向量才能全面

一个矩阵对任意方向上的向量进行变换有什么特点呢？下面我们来讨论整个平面上的向量被矩阵变换的结果。

在这里，我们讨论平面上任意的向量被一个矩阵所变换的情况。我们把平面上任意的向量简化为由原点出发的任意方向的单位向量，这些无数个不同方向的单位长度向量终端形成一个单位圆（见图 5-34（b））。矩阵$\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$对这个单位圆内的向量进行变换的图形是如何的呢？仔细观察图 5-35。图中的圆及其上面的等弧度分布的向量是原图形，椭圆及其上面的向量是变换后的像。虚线连接了变换前后对应的向量关系。

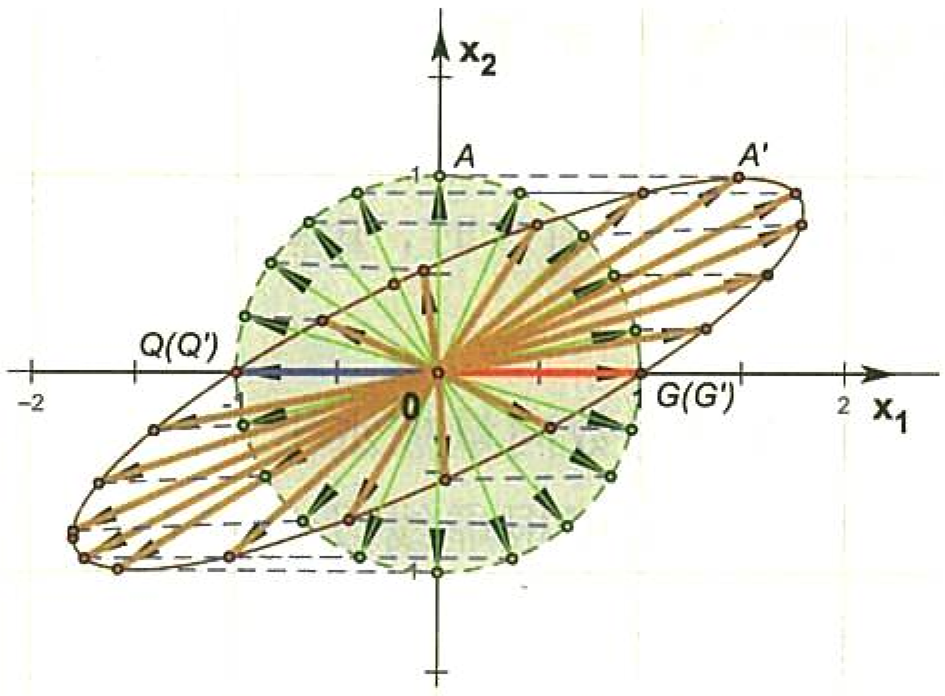


图 5-35 矩阵对单位圆线性变换中的特征向量保持了方向不变

从图中，我们可以看出：

(1) 大部分向量都发生了旋转和伸缩变换，只有 x_1 轴上的向量没有变换。

(2) 图中单位向量 $\overrightarrow{oG}=(1,0)$ 和 $\overrightarrow{oQ}=(-1,0)$ 就是矩阵的两个特征向量。显然，向量 $\overrightarrow{oG}$ 和 $\overrightarrow{oQ}$ 的倍数都是特征向量（因为满足 $\boldsymbol{Aa}=\lambda\boldsymbol{a}$），并且都在 x_1 轴上。

(3) 一个特征值 1，两个方向的特征向量。

用 MATLAB 动态观察特征值变化

数学计算软件 MATLAB 有一个工具，它可以把平面映射到自身的线性算子的作用可视化。该工具调用命令 eigshow。这个命令打开一个图形窗口，同时显示一个单位向量 x 和 $\boldsymbol{A}x$，即 x 在矩阵 $\boldsymbol{A}$ 下的像。矩阵 $\boldsymbol{A}$ 可以通过 eigshow 命令的输入参数给出，或者从图形窗口顶部的菜单中选择。为看到算子 $\boldsymbol{A}$ 在其单位向量上的作用，将鼠标指向向量 x 的端点，并拖动 x 沿着逆时针方向绕单位圆旋转。当 x 运动时，可看到像 $\boldsymbol{A}x$ 的变化。

3. 特征值和矩阵的迹

矩阵的迹，是矩阵 $\boldsymbol{A}$ 的主对角线元素的总和，也等于 $\boldsymbol{A}$ 的特征值的总和。

我们以二阶方阵为例看看特征多项式：

$$f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=\left|\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}-\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\right|=\begin{vmatrix}\lambda-a_{11} & -a_{12}\\ -a_{21} & \lambda-a_{22}\end{vmatrix} \tag{5-8}$$

$$=\lambda^2-(a_{11}+a_{22})\lambda+(a_{11}a_{22}-a_{12}a_{21})$$

另外，因为二次多项式 $f(\lambda)$ 在复数域必有两个根 λ_1、λ_2，因此这个多项式也可以表示为

$$f(\lambda)=(\lambda\text{-}\lambda_1)(\lambda\text{-}\lambda_2)=\lambda^2-(\lambda_1+\lambda_2)\lambda+(\lambda_1\lambda_2) \tag{5-9}$$

对比式（5-8）和式（5-9），我们可以知道：

$$\begin{cases}\lambda_1+\lambda_2=a_{11}+a_{22}\\ \lambda_1\lambda_2\quad=a_{11}a_{22}-a_{12}a_{21}=|\boldsymbol{A}|\end{cases} \tag{5-10}$$

式(5-10)是说，二阶**矩阵的特征值之和等于矩阵对角元素之和，特征值之积等于矩阵的行列式。**实际上这个总结对 n 阶矩阵同样成立。

这两个性质有用啊，比如常用于检验所求的特征值是否正确等。

特征值对于矩阵非常重要，矩阵的全部特征值直接关系到矩阵的许多性质；而我们又能通过矩阵的迹来有效检测有关特征值的问题，比如特征值的和是否等于迹来判断特征值是否求错；对比两个矩阵的迹是否相等来帮助判断俩矩阵是否是相似矩阵（迹不相等肯定不相似，迹相等

可能相似）等，迹就是矩阵特征值的蛛丝马迹啊，这可能是矩阵的迹名称的由来吧。

5.8.2 特征值和特征向量的物理意义

物理意义

前面说，矩阵乘法对应了一个变换，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，原向量主要发生旋转、伸缩的变化。如果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效果，那么这些向量就称为这个矩阵的特征向量，伸缩的比例就是特征值。

实际上，上述的一段话既讲了矩阵变换特征值及特征向量的几何意义（图形变换），也讲了其物理含义。物理含义就是运动的图景：特征向量在一个矩阵的作用下作伸缩运动，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。

特征向量不变，变的是特征值

常有教科书说特征向量是在矩阵变换下不改变方向的向量，实际上当特征值小于零时，矩阵就会把特征向量完全反方向改变，当然特征向量还是特征向量。我赞同特征向量不改变方向的说法：特征向量永远不改变方向，改变的只是特征值（方向反转特征值为负值了）。这有点像说冬天深圳的室外“温度”是 10℃，哈尔滨的室外“温度”是-30℃（称温度而不温）；也有点像说无人机在海拔“高度”100 米处飞行而核潜艇在海拔“高度”-500 米（称高度而不高）处游弋一样。

在后面的关于矩阵振动的谱讨论中，将遇到复特征值，复特征值将使特征向量在复平面上发生旋转，但在实平面看上同样只发生以原点为中心的线上伸缩。

详细的复特征值和复特征向量的意义请参考 5.8.5 小节。

关于特征值和特征向量，这里请注意两个亮点：一个是线性不变量的含义，一个是振动的谱含义。

1. 特征向量是线性不变量

有一个著名的布劳维尔（Brouwer）不动点定理，在经济分析中常用于求解市场均衡点的问题。不动点就是一种不变量。不动点的定义如下：

设 $\sigma: M \to M$ 是集合 M 到自身的一个变换。如果存在 $x^* \in M$，使得

$$\sigma(\boldsymbol{x}^*) = \boldsymbol{x}^* \tag{5-11}$$

则称 $\boldsymbol{x}^*$ 是变换 σ 的一个不动点。

我们现在讲的线性变换或矩阵变换同样具有不动点的概念。线性空间里的元素 $\boldsymbol{x}^*$ 是向量，不动点就成了不动向量或不变向量。进一步，我们把不变向量的概念扩大一下，不但包含不动向量 $\boldsymbol{x}^*$ 还包含共线的向量 $\lambda\boldsymbol{x}^*$ 作为研究对象，也就是把不动点的定义式改造为

$$\sigma(\boldsymbol{x}^*) = \lambda\boldsymbol{x}^* \qquad (\lambda \text{ 为常数})$$

这样就成了线性变换 σ 的特征向量的定义式了。

特征向量概念的亮点之一是不变量，这里叫线性不变量。因为我们常讲，线性变换啊线性变换，不就是把一根线（向量）变成另一根线（向量），线的变化的地方大多是方向和长度一块

变。而一种名叫“特征向量”的向量特殊，在矩阵作用下不变方向只变长度。不变方向的特性就被称为线性不变量。

如果有读者坚持认为负方向的特征向量就是改变了向量方向的话，我们不妨这样看线性不变量：特征向量的不变性是它们变成了与其自身共线的向量，它们所在的直线在线性变换下保持不变；特征向量和它的变换后的向量们在同一根直线上，变换后的向量们或伸长或缩短，或反向伸长或反向缩短，甚至变成零向量（特征值为零时），见图 5-36。图中，向量 $\boldsymbol{a}$ 是一个特征向量，因为在某一个矩阵变换 $\boldsymbol{A}$ 下，$\boldsymbol{Aa}$ 和 $\boldsymbol{a}$ 共线。

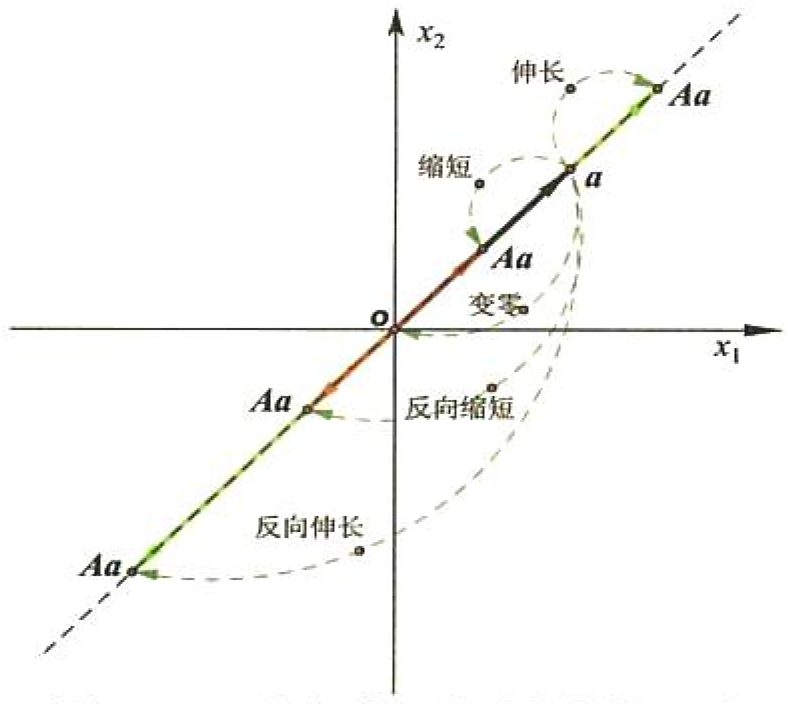


图 5-36 特征向量保持线性不变

特征向量是不变量的经典例子是地球旋转的自然现象。地球的全方位旋转矩阵就是三阶旋转阵。这个矩阵就是前面推导出的三维空间 $\mathbf{R}^3$ 中的标准正交基向量组：

$$\{(\cos\theta\cos\varphi, \cos\theta\sin\varphi, -\sin\theta)^{\mathrm{T}}, (-\sin\varphi, \cos\varphi, 0)^{\mathrm{T}}, (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^{\mathrm{T}}\}$$

（$0\leqslant\theta\leqslant\pi$，$0\leqslant\varphi\leqslant2\pi$）。把它们以 x、y、z 轴的顺序排列为列向量，所组成的三阶方阵为下面的方阵，这是一个任意方向旋转的矩阵：

$$\begin{bmatrix} \cos\theta\cos\varphi & -\sin\varphi & \sin\theta\cos\varphi \\ \cos\theta\sin\varphi & \cos\varphi & \sin\theta\sin\varphi \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

在图 5-37 中，如果我们取定地球自转的旋转轴为 z 轴，北（上）方向为 z 轴方向（同时 xoy 平面为赤道平面），那么 $\theta=0$，只有 φ 为变量。详细地说，x 轴上的 $\boldsymbol{i}$ 向量右旋了 φ 弧度，y 轴上的 $\boldsymbol{j}$ 向量右旋了 φ 弧度；z 轴上的 $\boldsymbol{k}$ 向量没有右旋，$\varphi=0$。因此旋转矩阵简化为下面的方阵 $\boldsymbol{A}$：

$$\boldsymbol{A}=\begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

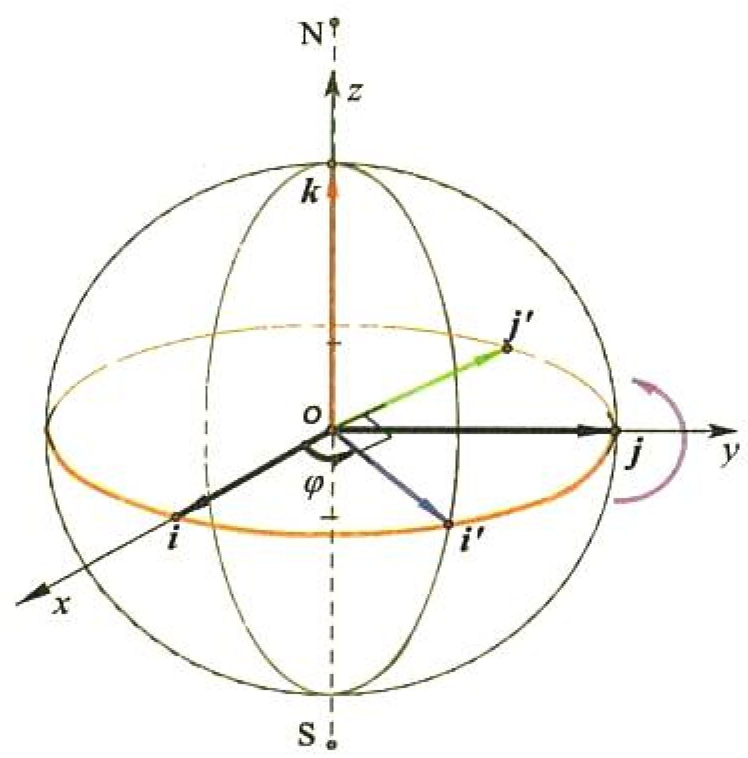


图 5-37 当地球绕着 NS 轴旋转时三阶旋转矩阵可以进一步化简

在计算 $\boldsymbol{A}$（注意这是一个变矩阵）的特征值和特征向量之前，我们不妨先大概地猜测一下。大家知道，方向不变或完全反转是特征向量的特性。那么，地球绕着南北极轴转动的时候，明显不变的就是极轴了。极轴上的所有向量包括南极点和北极点都是特征向量。因为极轴上的向量大小在旋转前后也不变，所以其对应的特征值是 1。一般情况下，其他的点都在动，没有更多的特征向量可以找到。

注意有几个例外。当 φ 取某些定值时还是有特征向量的。例如，当你旋转地球（你比那个试图用杠杆撬地球的阿基米德还牛！）到 180° 时，赤道平面上全都是特征向量，因为你把赤道平面上的所有的向量都掉转到反方向了。方向刚好相反但长度没有变化，因此特征值就是-1。旋转 180° 的矩阵是 $\boldsymbol{B}$：

$$\boldsymbol{B}=\left.\begin{bmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{bmatrix}\right|_{\varphi=\pi}=\begin{bmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

别走，还有特征向量没找完呢。

当你继续旋转地球（呵呵，再旋转下去估计牛顿就会认定你就是宇宙的第一推动力了）到 360° 时，奇迹发生了，地球上所有的点（所有的向量）都是特征向量了，因为旋转前和旋转后的点重合了。所有向量的方向和大小都没变，因此特征值是 1。旋转 360° 的矩阵 $\boldsymbol{C}$ 是：

$$\boldsymbol{C}=\left.\begin{bmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{bmatrix}\right|_{\varphi=2\pi}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}=\boldsymbol{E}$$

$A\alpha=\lambda E$

好啊，原来把所有向量整成特征向量的矩阵不是别人，正是大名鼎鼎的单位矩阵 $\boldsymbol{E}$。

前面的猜测毕竟是猜测，下面我们通过简单的计算来验证验证。

旋转矩阵的特征值和特征变量不难用特征多项式$|\boldsymbol{A}-\lambda\boldsymbol{E}|=0$求出，得到三个特征值：

$$\lambda_1=1,\ \lambda_2=\cos\varphi+\mathrm{i}\sin\varphi,\ \lambda_3=\cos\varphi-\mathrm{i}\sin\varphi$$

MY GOD！出现复数的特征值了。我们后面专门讨论它的意义（呵呵，先来实在的，一谈虚数我也有点发虚）。这里我们先看看地球的旋转运动中只取实数的特征值及其特征向量是什么东东。

显然，对于复数 λ_2、λ_3，只有当 φ=0, π, 2π 三种情况下其值为实数，分别为 $\lambda_2=\lambda_3=1,\ -1,\ 1$。总之，旋转矩阵的特征值只有两个取值 1 和-1（0，2π 相同）。由解方程组 $(\boldsymbol{A}-\lambda\boldsymbol{E})\boldsymbol{x}=\boldsymbol{0}$ 可知：

当 $\lambda_1=1$ 时，有两种情况。一种情况是对于所有的旋转矩阵（旋转角度 φ 为任意值）地球上的特征向量为 $k\begin{pmatrix}0\\0\\1\end{pmatrix}$，这正是地球的自转极轴（$z$ 轴），包括北极和南极点。不论旋转角度 φ 是多少，自转极轴一直是特征向量空间，空间里全是特征向量。另一种情况是对于旋转 0 或 360° 的矩阵，也就是单位矩阵 $\boldsymbol{E}$，求得特征向量为 $k_1\begin{pmatrix}1\\0\\0\end{pmatrix}+k_2\begin{pmatrix}0\\1\\0\end{pmatrix}+k_3\begin{pmatrix}0\\0\\1\end{pmatrix}$，这正是三维全空间。三维空间里的向量全部是特征向量，三维空间也就是特征向量空间。

当$\lambda_1=-1$时，旋转角度$\varphi=\pi$，地球上的特征向量为$k_1\begin{pmatrix}1\\0\\0\end{pmatrix}+k_2\begin{pmatrix}0\\1\\0\end{pmatrix}$，其张成的空间正是地球的赤道平面。赤道平面里的向量都是特征向量。

2. 特征值是振动的谱

除了线性不变量，另外一个亮点是关于振动方面的。戏说在宋朝时，我国就与发现矩阵特征值理论的机会擦肩而过。话说没出息的秦少游在往池塘里扔了一颗小石头后，刚得到一句"投石冲开水底天"的诗对之后，就猴急猴急地去洞房了，全然没有想到水波中隐含着矩阵的特征值及特征向量的科学大道理。大概地说，水面附近的任一点小水珠在原处上下振动（实际上在做近似圆周运动，因为水平方向也有运动分量），并没有随着波浪向外圈移动，同时这些上下振动的水珠的幅度在渐渐变小，直至趋于平静。在由某块有着特定质量和形状的石头被以某种角度和速度投入某个面积和深度特定的水池中所决定的某个矩阵中时，纹波荡漾中水珠的渐变过程的特征值起着决定性的作用，它决定着水珠振幅向量的频率和幅度减弱的衰退率。

最终，您面对着一个鱼跃鸟飞的风月荷塘，只看到一个巨型振荡器，满池的特征向量在那里上下舞动着，而潜水英雄特征值在默默的导演着这些……

在理解关于振动的特征值和特征向量的过程中，需要加入复特征值或复向量（参看 5.8.5 节）抑或复矩阵的概念，因为在实际应用中实特征值们是干不了多少事的。机械振动和电振动有频谱，振动的某个频率具有某个幅度，那么矩阵也有矩阵的谱，矩阵的谱就是矩阵特征值的概念，是矩阵所固有的特性，所有的特征值形成了矩阵的一个频谱，每个特征值是矩阵的一个"谐振频点"。

美国数学家斯特朗（G . Strang）在其经典教材《线性代数及其应用》中这样介绍了特征值作为频率的物理意义，他说：

大概最简单的例子（我从不相信其真实性，虽然据说 1831 年有一桥梁毁于此因）是一对士兵通过桥梁的例子。传统上，他们要停止齐步前进而要散步通过。这个理由是因为他们可能以等于桥的特征值之一的频率齐步行进，从而将发生共振。就像孩子的秋千那样，你一旦注意到一个秋千的频率，和此频率相配，你就使频率荡得更高。一个工程师总是试图使他的桥梁或他的火箭的自然频率远离风的频率或液体燃料的频率；而在另一种极端情况，一个证券经纪人则尽毕生精力与努力到达市场的自然频率线。特征值是几乎任何一个动力系统的最重要的特征。

其实，这个矩阵之所以能形成"频率的谱"，就是因为矩阵在特征向量所指的方向上具有对向量产生恒定的变换作用：周期性地增强（或减弱）特征向量的作用。进一步地，如果矩阵持续地叠代作用于向量，那么特征向量就会凸现出来。

比如，一个线性物理系统，其特性可以被一个矩阵所描述，那么这个系统的物理特性就可以被这个矩阵的特征值所决定，各种不同的信号（向量）进入这个系统中后，系统输出的信号（向量）就会发生相位超前或滞后、幅度放大或缩小等各种纷乱的变化。但只有特征信号（特征向量）被稳定地发生放大（或缩小）的变化。如果把系统的输出端口接入输入端口，那么只有特征信号（特征向量）第二次被放大（或缩小）了，其他信号继续其纷乱的不稳定变化幅度得不到增强……经过 n 次循环后，显然，乱七八糟的大量的向量群众们终不能成气候而渐渐消亡，只有特征向量们，心往一处想，劲往一线使，要么成功出人头地，要么失败杀身成仁。因

此我们可以在时间域上观察输出，就会得到一个或几个超级明显的特征信号（特征向量）出来。

弄过电路的哥们早看出了俺的含沙射影：切！绕什么绕，你说的不就是振荡器的原理嘛，前面讲的池塘是振荡器、桥梁是振荡器、电路中也有振荡器：振荡信号（电压、电流）构成了特征向量，特征值是决定了振荡信号的幅度和频率……

好家伙，被你看出来了，失败。我也不装神弄鬼了，下面干脆就直接弄个振荡器的电路来比对比对(不好意思，俺是学电子的，只能擅长弄点电路的例子来说事)。

3. 电子振荡器的特征向量

我们找个经典的振荡器来用矩阵分析分析，看看振荡器矩阵的特征值和特征向量到底是什么，振荡器的振荡频率与它们又是什么关系。

看图 5-38 所示的振荡器电路，这是一个著名的带有运放的振荡器电路，称为文氏（Wien）桥振荡器，它由一个运放以及一个串联和并联的 RC 电路构成反馈环路组成，两个附加的 R_1 和 $2R_1$ 电阻使运放的电压增益为 3。

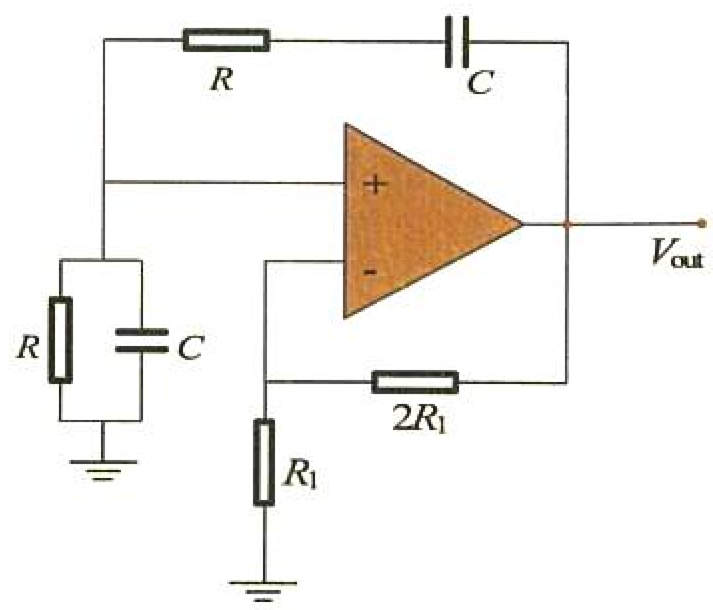


图 5-38 文氏桥振荡器电原理图

为了建立振荡器的传输参数矩阵，我们把反馈网络和运放分成两个二端口网络 $\boldsymbol{A}$ 和 $\boldsymbol{B}$（见图 5-39），分别求出其传输矩阵，然后进行级联得到整个电路的[$\boldsymbol{ABCD}$]矩阵。在这里，端口处的电压 U 和电流 I 构成了二维向量的两个元素。简略的计算如下：

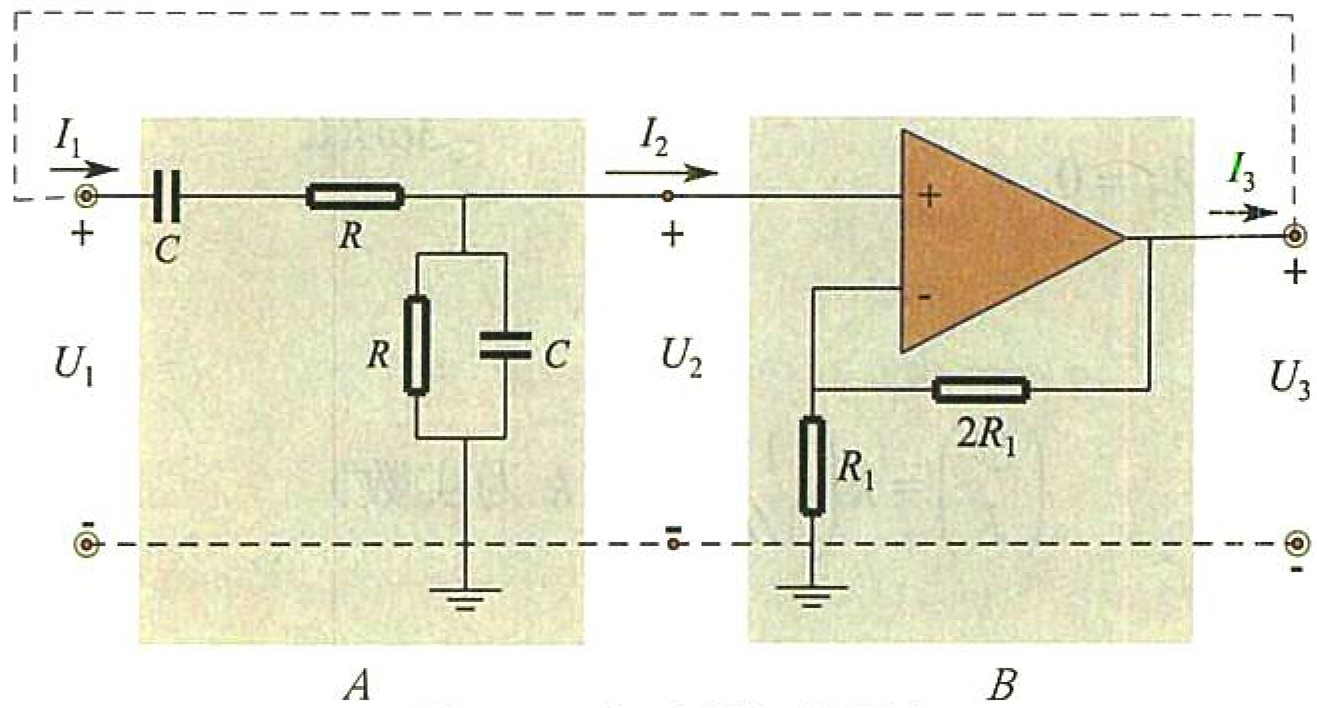


图 5-39 文氏桥振荡器电原理图

求得 $\boldsymbol{A}$ 网络（无源 RC 选频网络）的方程式如下：

$$\begin{pmatrix} U_2 \\ I_2 \end{pmatrix} = \begin{bmatrix} \dfrac{\mathrm{j}\omega RC}{1-(\omega RC)^2+\mathrm{j}3\omega RC} & 0 \\ \dfrac{\mathrm{j}\omega C\cdot(1+\mathrm{j}\omega RC)}{1-(\omega RC)^2+\mathrm{j}3\omega RC} & 1 \end{bmatrix} \begin{pmatrix} U_1 \\ I_1 \end{pmatrix}$$

传输矩阵 $\boldsymbol{A}$ 定义为

$$A=\begin{bmatrix}\dfrac{j\omega RC}{1-(\omega RC)^2+j3\omega RC} & 0\\ \dfrac{j\omega C\cdot(1+j\omega RC)}{1-(\omega RC)^2+j3\omega RC} & 1\end{bmatrix}$$

求得 ***B*** 网络（运放放大网络，运放被视为理想运放，+输入端电流为零，输出电流可以无穷大）的方程式如下：

$$\begin{pmatrix}U_3\\ I_3\end{pmatrix}=\begin{bmatrix}3 & 0\\ 1/r & 0\end{bmatrix}\begin{pmatrix}U_2\\ I_2\end{pmatrix}$$

r 为运放输出端的负载阻抗。

传输矩阵 ***B*** 定义为

$$\boldsymbol{B}=\begin{bmatrix}3 & 0\\ 1/r & 0\end{bmatrix}$$

那么整个网络的传输矩阵（先 ***A*** 后 ***B*** 故 ***BA***，参见 5.6.1 节）为

$$\boldsymbol{T}=\boldsymbol{BA}=\begin{bmatrix}3 & 0\\ 1/r & 0\end{bmatrix}\begin{bmatrix}\dfrac{j\omega RC}{1-(\omega RC)^2+j3\omega RC} & 0\\ \dfrac{j\omega C\cdot(1+j\omega RC)}{1-(\omega RC)^2+j3\omega RC} & 1\end{bmatrix}=\begin{bmatrix}\dfrac{j3\omega RC}{1-(\omega RC)^2+j3\omega RC} & 0\\ \dfrac{j(\frac{1}{r}\omega RC)}{1-(\omega RC)^2+j3\omega RC} & 0\end{bmatrix}$$

整个网络的向量方程式可重新表述为

$$\begin{pmatrix}U_3\\ I_3\end{pmatrix}=\boldsymbol{T}\begin{pmatrix}U_1\\ I_1\end{pmatrix}$$

至于传输矩阵 ***T*** 中的变量，R、C 和 r 对于一个具体确定的振荡电路来说是个固定的数值，j 也是固定的虚数符号；只有角频率 ω 是个不确定的值，范围可以从 0 到无穷大，甚至可以为包含负数在内的整个实数域。

至此，我们得到了描述振荡器网络的矩阵 ***T***。它是一个下三角矩阵，因此其特征值就是对角线上的元素，即

$$\begin{cases}\lambda_1=\dfrac{j3\omega RC}{1-(\omega RC)^2+j3\omega RC}=\dfrac{1}{1+\dfrac{(\omega RC)^2-1}{3\omega RC}j}\\ \lambda_2=0\end{cases}\tag{5-12}$$

进而求得对应的特征向量为

λ_1 对应的特征向量：

$$\begin{pmatrix}U\\ I\end{pmatrix}=k\begin{pmatrix}1\\ 1/r\end{pmatrix}\quad（k\text{ 为实数}）$$

λ_2 对应的特征向量：

$$\begin{pmatrix}U\\ I\end{pmatrix}=k\begin{pmatrix}0\\ 1\end{pmatrix}\quad（k\text{ 为实数}）$$

计算到这里，明显的怪异是，虽然传输矩阵的两个特征值分别是不确定的复数（和角频率 ω 有关）和零，但其对应的特征向量却和角频率 ω 无关，各是一条直线。

怪异是怪异，但传输矩阵的特征值和特征向量的实际意义还是不明确，和振荡电路到底什么关系嘛？你前面不是说**特征值是振动的谱，特征向量是线性不变量**吗？

哎……等等，你的问题点醒俺了。看官请仔细想想：

特征值的式子(5-12)不正是角频率 ω 的复变函数吗？其幅度正是振荡器的振动的频谱；λ_1 对应的特征向量是 $\begin{pmatrix} U \\ I \end{pmatrix} = k\begin{pmatrix} 1 \\ 1/r \end{pmatrix}$，这里可看出，特征向量里面的电压、电流元素满足 $I = U/r$；反过来说，只有满足欧姆定律的电压、电流的量才属于特征向量。简单点说，如果负载 r 是电阻，那么同相的电压、电流构成特征向量。

我们说，特征向量是个不变的向量（k 为某个定值时），特征向量的交变变化是由特征值所控制的，特征值是个交变的复变量，因此振荡信号的主要信息在特征值上。

老大！大家都知道振荡器始终会振荡在某一个频率点，其频谱是个线谱。从特征值的表达式上我怎么没看出来？

看官，实际上咱上述的分析是针对开环网络的，不是真正的振荡器的，真正的振荡器网络是闭合成环路的，***AB*** 网络的输出端接到其输入端。

从电路的理论来说，对于一个振荡器，白噪声中某一频率（$\omega = 1/(RC)$）的信号即微弱的电压（和电流）经过 n 次循环往复，信号电压（和电流）幅度越来越大，终于稳定的输出（如图 5-40）。而白噪声中的绝大多数频率 $\omega \neq 1/(RC)$ 的信号都没有成气候，渐渐湮灭了。

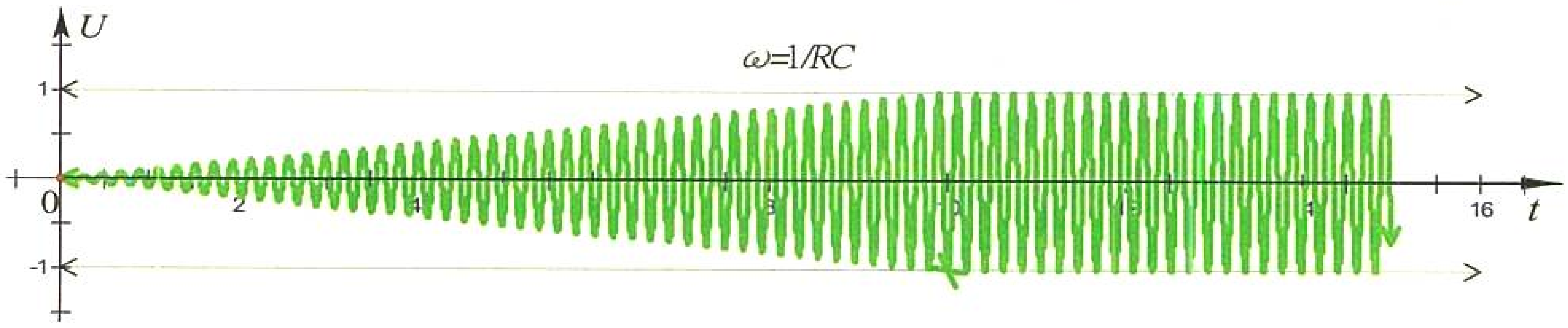


图 5-40 振荡器信号的起振过程

从数学的角度看，这样的一种情况正是线性代数里面的矩阵之迭代作用，或矩阵幂的结果。

好，下面我们对这个网络的传输矩阵 $\boldsymbol{T}$ 进行 n 次迭代运算 $\boldsymbol{T}^n$，看看矩阵幂和振荡器是如何联系在一块的？

由公式 $\begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix}^n = \begin{bmatrix} a^n & 0 \\ ba^{n-1} & 0 \end{bmatrix}$ 可知

$$\boldsymbol{T}^n = \begin{bmatrix} \dfrac{\mathrm{j}3\omega RC}{1-(\omega RC)^2+\mathrm{j}3\omega RC} & 0 \\ \dfrac{\mathrm{j}(\frac{1}{r}\omega RC)}{1-(\omega RC)^2+\mathrm{j}3\omega RC} & 0 \end{bmatrix}^n = \begin{bmatrix} \left[1+\dfrac{(\omega RC)^2-1}{3\omega RC}\mathrm{j}\right]^{-1} & 0 \\ \left\{r\left[3+\dfrac{(\omega RC)^2-1}{\omega RC}\mathrm{j}\right]\right\}^{-1} & 0 \end{bmatrix}^n$$

$$= \begin{bmatrix} \left[1+\dfrac{(\omega RC)^2-1}{3\omega RC}\mathrm{j}\right]^{-n} & 0 \\ \left\{r\left[3+\dfrac{(\omega RC)^2-1}{\omega RC}\mathrm{j}\right]\right\}^{-1}\left[1+\dfrac{(\omega RC)^2-1}{3\omega RC}\mathrm{j}\right]^{1-n} & 0 \end{bmatrix}$$

哇噻，是不是有点复杂呀，再往下怎么算呀，好像离通俗易懂的方向越来越远啦？

老实说，哥们我也在这里纠结了好久，终于有天晚上“灵感”来了。当时躺在床上正和 LP 聊天，忽然灵机一动：求解复数的 n 次方幂，应该用三角形式或复指数形式啊。唉，愚昧，怎么忘了棣莫弗定理了，赶紧把中学的复数一节复习了一遍，心里有了底。

好，把矩阵里面的复数先改成复三角形式，继续上面的计算：

$$T^n=\begin{bmatrix}\left[\sqrt{1+\left[\dfrac{(\omega RC)^2-1}{3\omega RC}\right]^2}(\cos\theta+\mathrm{j}\sin\theta)\right]^{-n} & 0\\ \left[r\sqrt{3^2+\left[\dfrac{(\omega RC)^2-1}{\omega RC}\right]^2}(\cos\theta+\mathrm{j}\sin\theta)\right]^{-1}\left[\sqrt{1+\left[\dfrac{(\omega RC)^2-1}{3\omega RC}\right]^2}(\cos\theta+\mathrm{j}\sin\theta)\right]^{1-n} & 0\end{bmatrix}$$

$$=\begin{bmatrix}\left\{1+\left[\dfrac{(\omega RC)^2-1}{3\omega RC}\right]^2\right\}^{-n/2}(\cos n\theta-\mathrm{j}\sin n\theta) & 0\\ r^{-1}\left\{3^2+\left[\dfrac{(\omega RC)^2-1}{\omega RC}\right]^2\right\}^{-1/2}\left\{1+\left[\dfrac{(\omega RC)^2-1}{3\omega RC}\right]^2\right\}^{(1-n)/2}\{\cos[(n-1)\theta+\varphi]-\mathrm{j}\sin[(n-1)\theta+\varphi]\} & 0\end{bmatrix}$$

唉，看起来越来越复杂了。

其实你只要从复数的几何意义着眼，立马上面的矩阵变得简单而亲切。我们把三角矩阵中的第一个元素（特征值）拿出来把玩把玩：

$$\left\{1+\left[\frac{(\omega RC)^2-1}{3\omega RC}\right]^2\right\}^{-n/2}(\cos n\theta-\mathrm{j}\sin n\theta)$$

这个元素同样是矩阵 $\boldsymbol{T}^n$ 的一个特征值（另外一个特征值仍是 0，因为 $\boldsymbol{T}^n$ 仍是一个三角矩阵，对角线上的元素就是特征值），我们分别看看此非零特征值的两部分因子所包含的几何和物理意义。

特征值后面部分的因子 $\cos n\theta-\mathrm{j}\sin n\theta$ 的物理意义是表示一个复振荡信号，实轴上的余弦振荡（类比于水珠的上下振动）和虚轴上的正弦振荡（类比于水珠的左右振动）随着 n 不间断地增大而持续进行着。这两个振荡可以合成一个复信号，这个复信号也可以看做二维单位向量（注意哈：这向量不是咱矩阵方程里研究的对象——电压电流的向量 $\begin{pmatrix}U\\I\end{pmatrix}$）。

这个单位向量没有闲着，它在作圆周运动（类比于小水珠在做圆周运动——上下和左右的合运动）。所以说因子 $\cos n\theta-\mathrm{j}\sin n\theta$ 的几何意义就是单位向量在复平面上绕着原点进行顺时针旋转，旋转随着 n 不间断地增大而持续进行着。

特征值前面部分的因子 $\left\{1+\left[\dfrac{(\omega RC)^2-1}{3\omega RC}\right]^2\right\}^{-n/2}$ 是在做圆周运动的单位向量的模，换句话说，表示振荡信号的幅度或者说旋转着的向量的长度。显然，这个振荡信号的幅度是频率 ω 的函数，也就是说，不同的频率的振荡信号幅度也不同。

好了，是不是有些清楚了。哦，还不清楚？那我假设这个具体的振荡器中电阻电容值的乘积 RC=0.5，令 n=2000，其振荡信号幅度因子的几何图形如图 5-41 所示。显然（图中横坐标为 ω），几何图形就是振荡器的频谱，这个频谱和频谱仪测量出来的相同（画出 n=5000 和 n=10000 的频谱图对比。可以看出振荡器的频谱随着 n 的增大，高斯波型谱逐渐变化为线型谱了）。

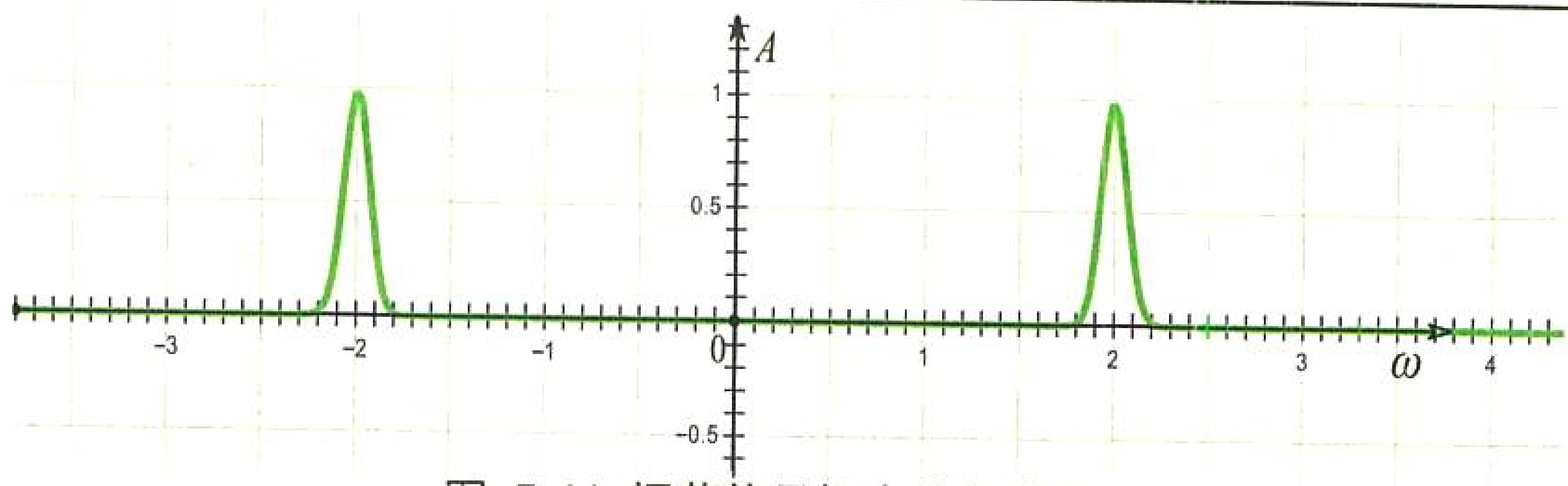


图 5-41 振荡信号幅度的频谱图

总结一下，振荡器里的所有特征向量中，只有振荡周期是 RC 的特征向量们才被反复地放大并被稳定输出（无相位差的余弦的电压和电流信号）。

矩阵的其他元素俺就不分析了，留给诸位研究吧。

将交流的电压和电流信号各看作一个向量，其实不难理解。在电路分析课程里，同频的电压和电流可以用相量或复数进行分析，这其实就是向量分析。把电压看作一个一维向量，电流看作另外一个一维向量，电信号如果通过一个电阻器件，则电压和电流向量夹角（相位差）不变；电信号如果通过一个电感器件（阻抗为 $j\omega L$），则电压和电流向量夹角（相位差）变为 90°，电流滞后；电信号如果通过一个电容器件（阻抗为 $1/j\omega C$），则电压和电流向量夹角（相位差）也变为 90°，电压滞后。

电压和电流信号的幅度和夹角变化刚好也可以用复数描述，因此我们就可以把一维的电压向量和一维的电流向量各看作一个复数，把两个复数组合就成为一个二维向量，这就是为什么振荡器分析要用二维的电压电流复向量而传输矩阵也是复矩阵的原因。

4. 激光器的特征向量

电子振荡器的交流电压和交流电流信号构成了一个二维向量。如果电压和电流的相位差保持不变，那么这个向量的方向就保持不变；如果电压值和电流值的平方和保持不变，那么这个向量的长度就保持不变。一些向量的频率如果刚好等于振荡器谐振频率，那么它们就是特征向量，特征向量进出振荡器不会改变向量方向，但其幅度在正负方向上按照余弦规律在变化。

如果要想象出交流的电压和交流电流信号所构成了一个二维向量如何在一条直线上正负交替振荡着反复通过电路并同时输出的图景还是有点挑战性的。别头疼了哈，其实还有更直观的特征向量模拟器呢。

随着电子振荡器的频率越来越高，信号变得难以约束在电路中，部分信号以电磁波向量场的形式向外部空间辐射，因此电压和电流量（比如位移电流）难以对电磁场信号进行描述或测量。咱们也应抛弃电压和电流概念而去采用电场和磁场组合为信号向量来分析。电磁波的传播是有方向的，更高频率的光波（光也是电磁波）的方向性更明显，初中物理就反复告诉同学们光是沿直线传播的。光波的组成是光子，又称为光粒子，传播方向是直线。呵呵，是不是光子特像向量了。

到这里，俺情不自禁地联想到了激光器的原理。是啊，就是 FP 激光器（法布里-伯罗激光器），它简直就是矩阵和特征向量模拟器。太形象了，下面我们就把物理中的激光器的原理搬到这里，供大家玩味一番。

激光器（LASER，Light Amplification by Stimulated Emission of Radiation）是利用受激辐射原理使光在某些受激发的物质中放大或振荡发射激光的器件。

激光器的大概原理是:

用光、电等办法对物质进行激励，使得其中一部分粒子激发到能量较高的状态，当这种高能量状态的粒子数大于能量较低状态的粒子数（所谓的粒子数反转）时，如果受到光子的激发，物质就能对某一波长的光产生放大作用，也就是这种波长的光辐射通过物质时，会发射强度放大并与入射光波位、频率和方向一致的光辐射。如图 5-42 所示，若把激发的物质放置于谐振腔内，光辐射在谐振腔内沿轴线方向往复反射传播，多次通过物质，光辐射被放大许多倍，形成一束强度大、方向集中的光束“激光”，这就是激光振荡器。

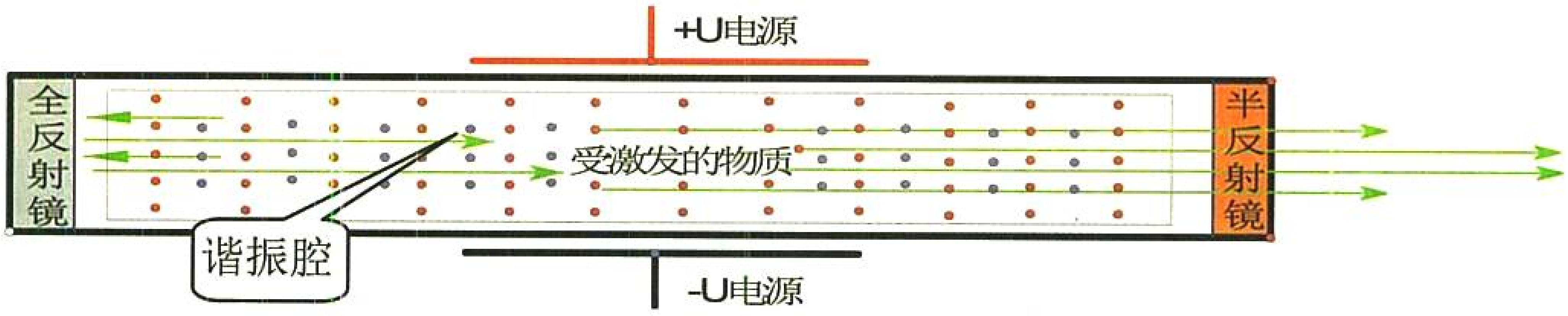


图 5-42 半导体激光振荡器的结构原理图

通常的激光器都是由工作物质、泵浦源和光学谐振腔这三个部分组成。稍微细致地讲，激光器是由一组平行平面反射镜把激光工作媒质夹在中间组成的光学谐振腔和从外界向激光器工作媒质供应能量的泵浦源组成（见图 5-42）的。

对半导体激光器来讲，泵浦源是由加载的电压形成的，电流注入作为其供应能量的方式。注入的能量将足够数量的粒子从低能级“抽运”到高能级，这很像水泵将水从低处抽到高处，因而将上述“抽运”过程称为“泵浦”。提供泵浦的设备称为**泵浦源**。泵浦源不断地向发光物质输入能量，把处在低能级的发光粒子激发到高能级上去，从而实现粒子数反转。

光学谐振腔是由放置在激光物质两边的两个反射镜组成的，其中半反射镜作为输出境使用。光学谐振腔有多种，最基本的一种称为法布里–伯罗共振器（即 FP 激光器），它由一对有少许透射率的平行平面镜组成。光在谐振腔内来回反射形成振荡，抑制自发发射，增强受激发射，不断获得光放大。这时从谐振腔轴线方向的半反射镜透射输出的光就是激光。

处在粒子数反转状态的**工作物质**，其大量的原子处在不稳定的激发态，它们很快会自发地由高能级向低能级跃迁，并辐射光子。原子的这种自发辐射是完全独立的，所以不同的原子发射的光子的方向全然不同。顷刻间，工作物质中出现了向四面八方传播的光子，如图 5-43（b）所示。

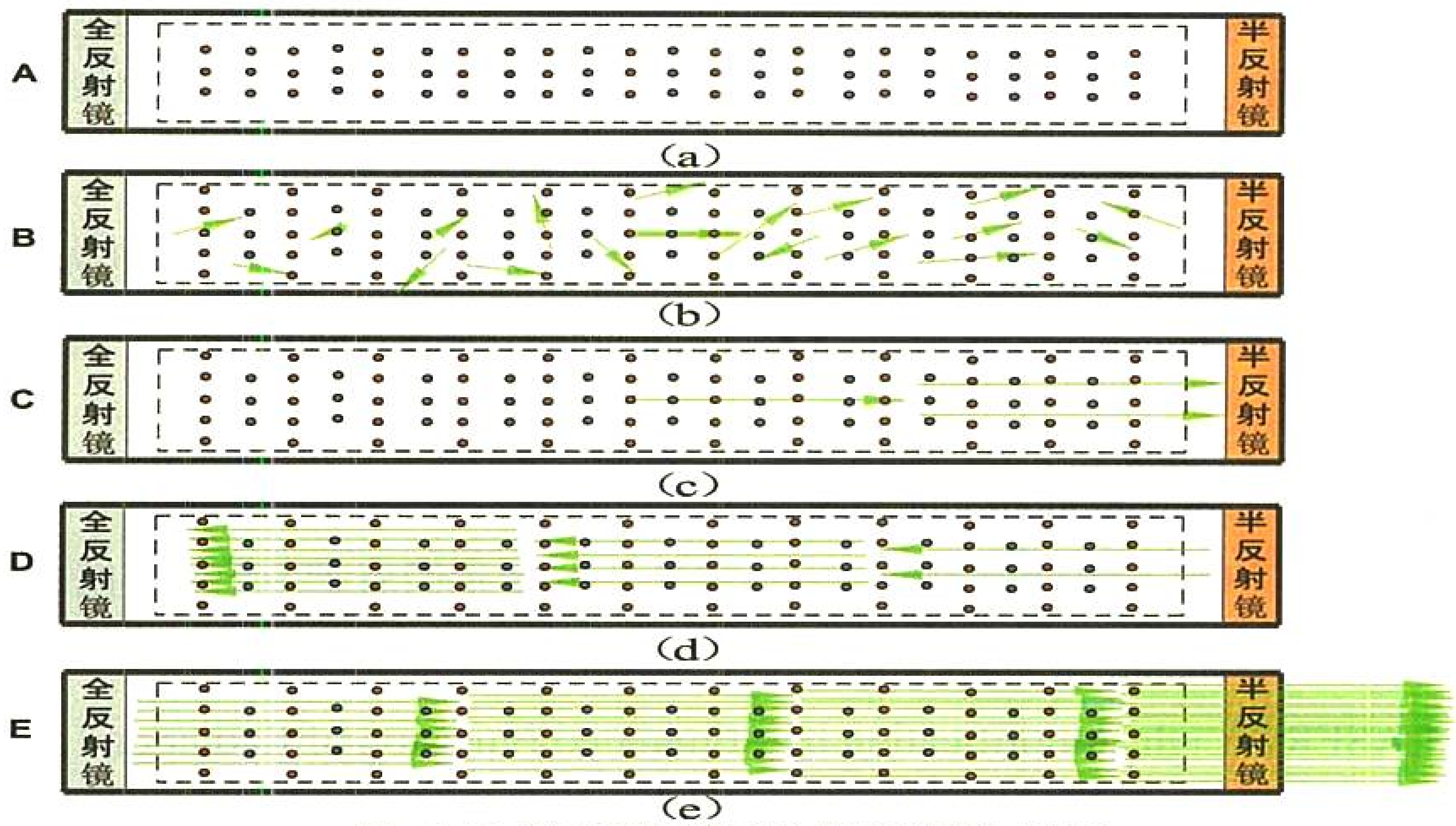


图 5-43 激光器受激产生级联发射示意图

图 5-43 的解释如下：
图（a）是级联发射之前，激光器晶体中的原子（黑点）处在基态；
图（b）中抽运光（箭头）被原子吸收，并且把大多数原子提升到激发态；
图（c） 虽然大多数光子跑到晶体外面去了，但当一个被激发了的原子自发发射出一个光子（与晶体轴平行的箭头）时，级联过程就开始了，这个光子激励另一个原子，使它贡献出第二个光子；
图（d）当这些光子在晶体的两个端面来回反射时，这一过程继续不断扩大；
图（e）晶体的右侧端面只是部分反射的，于是当放大到足够强的时候，穿出部分反射端面的光束就会很强。

假定工作物质具有圆柱形状，这些自发辐射光子必有一小部分沿着其中心轴的方向“同心同轴”地传播，而多数光子是与中心轴有一定夹角的，这些“离心离轴”的光子很快从工作物质的侧面逃逸出去，对激光的产生没有帮助。而那些“同心同轴”的光子在沿着工作物质中心轴方向运动时，将引起路径上处于高能级原子的受激辐射，产生与其具有相同频率、相同相位、沿相同方向传播的光子（这就是爱因斯坦提出的光子受激辐射理论，**在受激辐射跃迁的过程中，一个诱发光子可以使处在高能级上的粒子产生一个与诱发光子状态完全一样的光子**）。这些光子与诱发它们的光子一起，又向前传播，并激励其他的原子辐射出与它们完全相同的光子。如此下去，使光子数以神奇的速度按指数规律增长，产生了爆炸式的连锁反应，这就是**受激辐射光放大。**

神奇的是，由于所有的光子都是逐次受激辐射产生的，这使得它们具有相同的频率、相同的初相位、相同的偏振态和相同的传播方向！

理论上讲，只要工作物质足够长，则不管初始的自发辐射有多弱，最终总可以被放大到一定强度。但在实际的激光器中，不可能有所谓足够长度的工作物质。因此通常是在工作物质两端各放一块反射镜，使光能够来回反射多次通过工作物质并被不断放大，如图 5-43（c）、（d）所示。为充分利用光能，介质往往被置于一聚光腔体中，后者与端面反射共同构成光学谐振腔。可见，光学谐振腔有产生与维持激光两方面作用。由于有光学谐振腔的存在，一方面在光学正反馈作用下，腔内光子数因不断往返通过工作物质而被放大，另一方面由于谐振腔存在各种损耗，腔内光子数又不断减少。当放大与衰减互相抵消时就可以形成稳定的光振荡了，如图 5-43（e）所示。

从图 5-43（b）、（c）可以看出，光学谐振腔没有腔壁，许多依赖于内部反射的振荡模式就会自行消失。任何以一个斜角击中端面镜面的模式，最终都会自行从侧面跃出空腔而消失掉，因而不会把能量汇集起来。只有那些准确地按照轴向方向在两个镜面之间来回反射的模式，才能够留存下来进行振荡。

明白没，可以这样讲：谐振腔就是矩阵的 n 次幂，**激光器输出的光子全部是特征向量**。

根据上面的说法，光子是可以用向量表示的，激光器也可以用矩阵来描述。但好像没有人这样做。呃，差点忘了，上世纪 20 年代，德国的海森伯弄了个量子矩阵力学——量子力学用矩阵来解释。唉，科学先知就是科学先知，膜拜膜拜先。

频率有频谱，矩阵有矩阵的谱(特征值)。如果特征一样，那么我们认为两个矩阵相似。但是实际当中不存在特征值完全相同的矩阵，所以矩阵的相似程度就成了比对特征值本身。这种

比对通常是在欧氏空间中展开的。

矩阵的谱定理的最重要的应用是在经典力学和量子力学中。任何振动都是调和振动的线性组合。这个原理是物理学中分析声和光的基础。在量子力学中，谱定理具有更加重要的位置。

向量可以张成向量空间，特征向量当然可以张成特征向量的空间。激光器的特征向量空间是一条直线，就是激光射出的中轴线。城市的夜空中，高楼顶部的蓝色激光束直刺茫茫深空，向人们展示了特征向量空间让人惊秫的魅力。

5.8.3 特征向量空间的几何图景

特征向量空间就是矩阵的特征向量们所张成的线性子空间。我们知道，根据特征方程（5-7）的定义式，我们可以求解矩阵的特征值及特征向量（或特征向量空间）。将式（5-7）改写为

$$(\lambda \boldsymbol{E}-\boldsymbol{A})\boldsymbol{a}=\boldsymbol{0} \tag{5-13}$$

如果这个方程有非零解 $\boldsymbol{a}$，则说明矩阵存在特征向量。方程组有非零解的充要条件是系数行列式等于零，也就是

$$|\lambda \boldsymbol{E}-\boldsymbol{A}|=0$$

注意，这是一个关于特征值 λ 为变量的多项式方程。求解了这个方程，就可以得到矩阵所有的特征值 $\lambda_1,\lambda_2,\cdots$；把每个特征值分别代入式(5-13)，就可以得到对应每个特征值的特征向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots$。

下面通过例题给出特征空间的图形。

1. 特征子空间的图例

由等式 $\boldsymbol{Aa}=\lambda\boldsymbol{a}$ 知道，如果 $\boldsymbol{a}$ 是特征向量，那么 $k\boldsymbol{a}$ 也是特征向量(代到等式验算即可知道)。$k\boldsymbol{a}$ 是一条直线，直线上的所有向量都是对应特征值 λ 的特征向量，直线上的特征向量构成特征向量子空间，称之为特征子空间。

特征子空间不都是直线。如果一个特征值可以对应或求得两个线性无关的特征向量如 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$，那么这两个特征向量可以张成一个平面的特征子空间。这个特征平面里的所有向量都是特征向量，因为如果 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 满足 $\boldsymbol{Aa}=\lambda\boldsymbol{a}$，那么平面上任一向量 $k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2$ 也满足此式。

类似地，如果一个特征值对应 S 个线性无关的特征向量就是对应一个 S 维的特征子空间。

下面的例子给出了矩阵特征向量的求法及特征向量子空间的几何图形。

例 5.3 求矩阵

$$\boldsymbol{A}=\begin{bmatrix}4 & 6 & 0\\ -3 & -5 & 0\\ -3 & -6 & 1\end{bmatrix}$$

的特征值及特征向量，并说明其几何意义。

解 由矩阵 $\boldsymbol{A}$ 的特征方程：

$$|\boldsymbol{A}-\lambda\boldsymbol{E}|=\begin{bmatrix}4-\lambda & 6 & 0\\ -3 & -5-\lambda & 0\\ -3 & -6 & 1-\lambda\end{bmatrix}=(\lambda+2)(\lambda-1)^2=0$$

得到特征值$\lambda_1=-2,\ \lambda_2=\lambda_3=1$。把$\lambda_1=-2$代入式$(\boldsymbol{A}-\lambda\boldsymbol{E})\boldsymbol{x}=\boldsymbol{0}$得到齐次线性方程组：

$$\begin{cases}6x_1+6x_2=0\\3x_1+3x_2=0\\3x_1+6x_2-3x_3=0\end{cases}$$

它的基础解系为$\xi=\begin{pmatrix}-1\\1\\1\end{pmatrix}$，所以对应于$\lambda_1=-2$，$\boldsymbol{A}$的全部特征向量为

$$\boldsymbol{x}=c\xi=c\begin{pmatrix}-1\\1\\1\end{pmatrix}\qquad（c\neq 0，实数）$$

如图 5-44 所示，过向量ξ作直线L，则以原点o为起点，以L上除o点以外的任意点为终点的向量$c\xi$都是矩阵$\boldsymbol{A}$的关于特征值-2 的特征向量。它们全体构成$\boldsymbol{A}$的关于特征值-2 的特征向量子空间。此向量空间中的任意向量$\boldsymbol{x}$受矩阵$\boldsymbol{A}$作用后成为向量$-2\boldsymbol{x}$，它仍然位于直线L上，只是方向与$\boldsymbol{x}$相反，大小为$\boldsymbol{x}$的 2 倍。

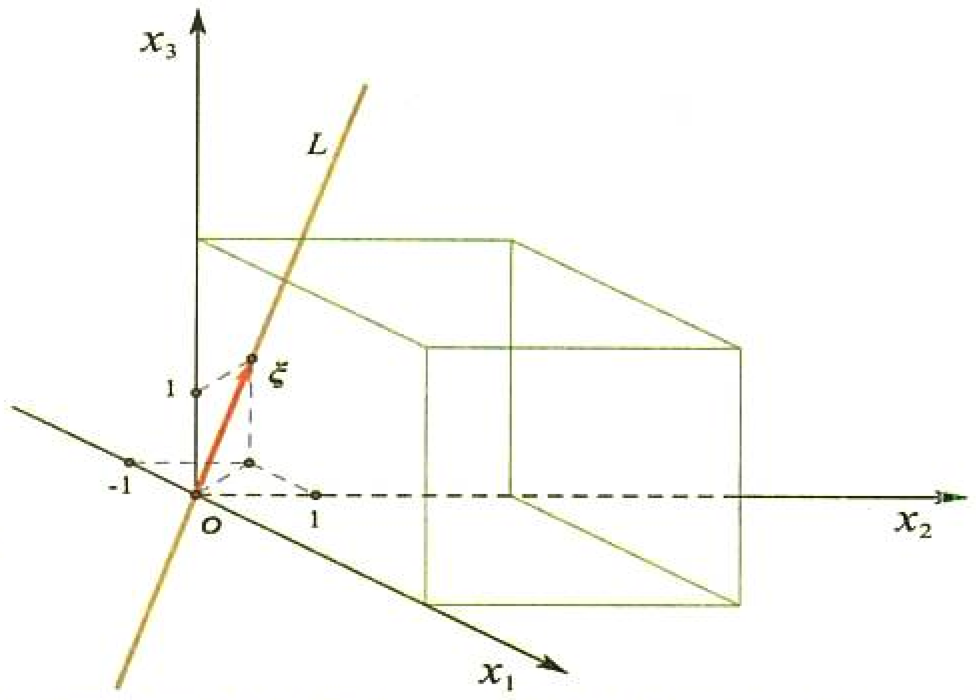


图 5-44 矩阵A的关于$\lambda_1=-2$的特征子空间——直线L

把$\lambda_2=\lambda_3=1$代入式$(\boldsymbol{A}-\lambda\boldsymbol{E})\boldsymbol{x}=\boldsymbol{0}$得到齐次线性方程组：

$$\begin{cases}3x_1+6x_2=0\\3x_1+6x_2=0\\3x_1+6x_2=0\end{cases}$$

它的基础解系为$\xi_1=\begin{pmatrix}-2\\1\\0\end{pmatrix}$，$\xi_2=\begin{pmatrix}0\\0\\1\end{pmatrix}$，所以对应于特征值$\lambda_2=\lambda_3=1$，$\boldsymbol{A}$的全部特征向量为

$$\boldsymbol{x}=c_1\xi_1+c_2\xi_2=c_1\begin{pmatrix}-2\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\0\\1\end{pmatrix}\qquad（c_1、c_2为不同时等于 0 的实数）$$

特征子空间为一平面，如图 5-45 所示的平面Π。

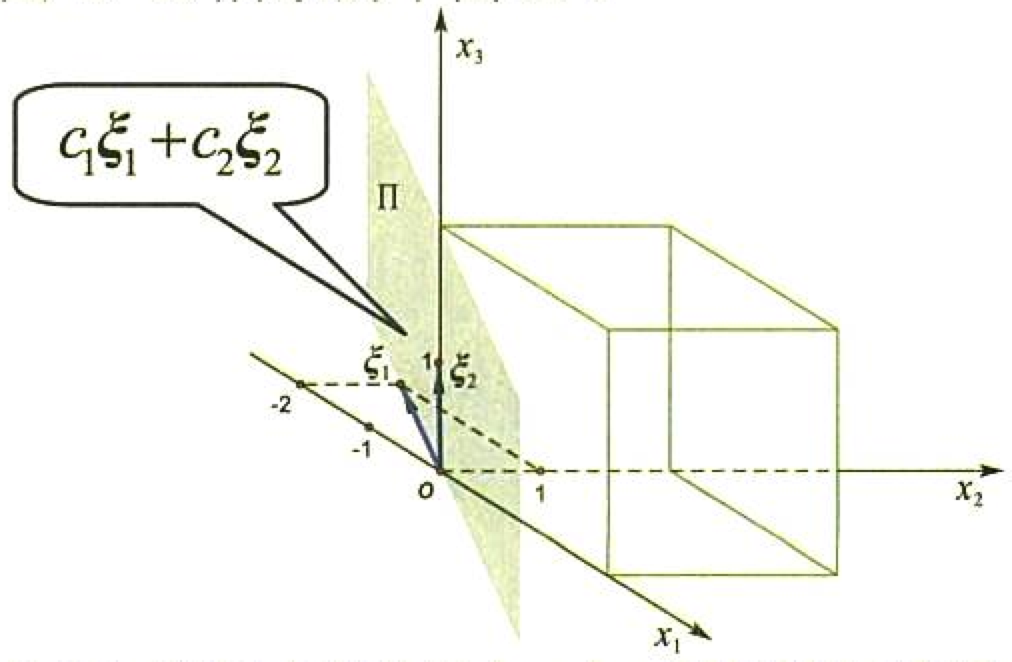


图 5-45　矩阵$\boldsymbol{A}$的关于$\lambda_2=\lambda_3=1$的特征子空间——平面Π

如图 5-45 所示，过向量ξ_1、ξ_2作平面Π，则以原点o为起点，以Π上除o点以外的任意点为终点的向量$c_1\xi_1+c_2\xi_2$都是矩阵$\boldsymbol{A}$的关于特征值为 1 的特征向量，它们的全体构成$\boldsymbol{A}$的关于 1 的特征向量子空间。此向量平面空间中的任意一个向量$\boldsymbol{x}$受矩阵$\boldsymbol{A}$作用后没有任何变换，或者说它仍然位于平面Π上，方向、大小均不改变。

从上例我们已清楚地看到，$\boldsymbol{A}$的特征向量空间中位于某直线上的特征向量，受$\boldsymbol{A}$作用后得到的向量有着仍然位于该直线上的几何特征。

以上举的例子都局限于三维线性空间，同样可推广到n维线性空间。

因此，对于某一个特征值，其特征空间可以是一条直线，也可以是一个平面，亦或是更高维的空间。空间里的向量们受到幅度相同的作用后并且保持着原方向。对应于激光振荡器，激光输出是一个直线束对应着的特征空间是一个直线，输出激光面则对应着特征平面，输出激光柱则对应着特征三维空间。

2. 特征值的代数及几何重数的意义

上节的例子里，在特征多项式方程$|\boldsymbol{A}-\lambda\boldsymbol{E}|=(\lambda+2)(\lambda-1)^2=0$中，特征值为-2 的求解因子式$(\lambda+2)$是一次的，对应的特征子空间——直线的维数是一维的；特征值为 1 的求解因子式$(\lambda-1)^2$是二次的，对应的特征子空间——平面的维数是二维的。

这里有个术语，一个特征值的求解因式的次数被称之为代数重数，特征值的特征子空间的维数被称之为几何重数。

所以上节矩阵例子里，特征值-2 的代数重数和几何重数都是 1；特征值 1 的代数重数和几何重数都是 2。

这个结论对一般的矩阵成立吗？不。

正确的结论是，特征值的代数重数大于或等于几何重数。

因为有时候代数重数≥2 的特征值，它的特征向量子空间会亏损——子空间会重合，子空间重合不是子空间直和，所以特征子空间的维数会变小，也就是几何重数可能≤2。

举个例子。对于三角矩阵：

$$\boldsymbol{A}=\begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix}$$

我们知道，a和b是矩阵的两个特征值（三角矩阵的对角线元素就是矩阵的特征值）。容易求得：a的特征向量子空间是$k\begin{pmatrix}1\\0\end{pmatrix}$，这是一条直线——$x_1$坐标轴；$b$的特征向量子空间是$k\begin{pmatrix}1\\b-a\end{pmatrix}$，这也是一条直线。$a$和$b$的代数重数和几何重数都是 1。

然而，当b的值逐渐接近于a的值，即$b\to a$时，b的特征向量子空间逐渐接近于a的特征向量子空间，即$k\begin{pmatrix}1\\b-a\end{pmatrix}\to k\begin{pmatrix}1\\0\end{pmatrix}$。

当b等于a时，特征值合二为一，同时特征子空间也合二为一——两根线重合为一根线，特征子空间亏损了。这时，矩阵特征值的重数为 2，但特征子空间的维数仍然是 1，即几何重数是 1。

一个极端的例子是n阶三角矩阵：

$$\begin{bmatrix} a & 1 & & & \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & a & 1 \\ & & & & a \end{bmatrix}$$

这个矩阵只有一个特征值 a，代数重数为 n，其特征子空间是一根直线（即 x_1 轴），几何重数是 1。

5.8.4 实对称矩阵的特征值和特征向量

对称（Symmetric）矩阵是满足 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$ 的矩阵，其主对角线元素是任意数，而其他元素则沿主对角线两侧对称相等。因为行元素和列元素对称，显然行数和列数相等，这样的矩阵必是一个方阵。实对称矩阵的元素都是实数。

我们看看一个矩阵由非对称的元素更换为对称的元素后，矩阵的变换图形会有何变化？其特征值和特征向量又会作何变化？

设 $\boldsymbol{A}=\begin{bmatrix}1 & 1.5\\0 & 1\end{bmatrix}$（见 5.8.1 节的例子），我们已经知道，矩阵的特征向量的图形是图 5-46(a)。其特征向量在 x_1 轴上，或者说特征子空间是 x_1 轴。x_1 轴和椭圆图形显然没有对称关系。

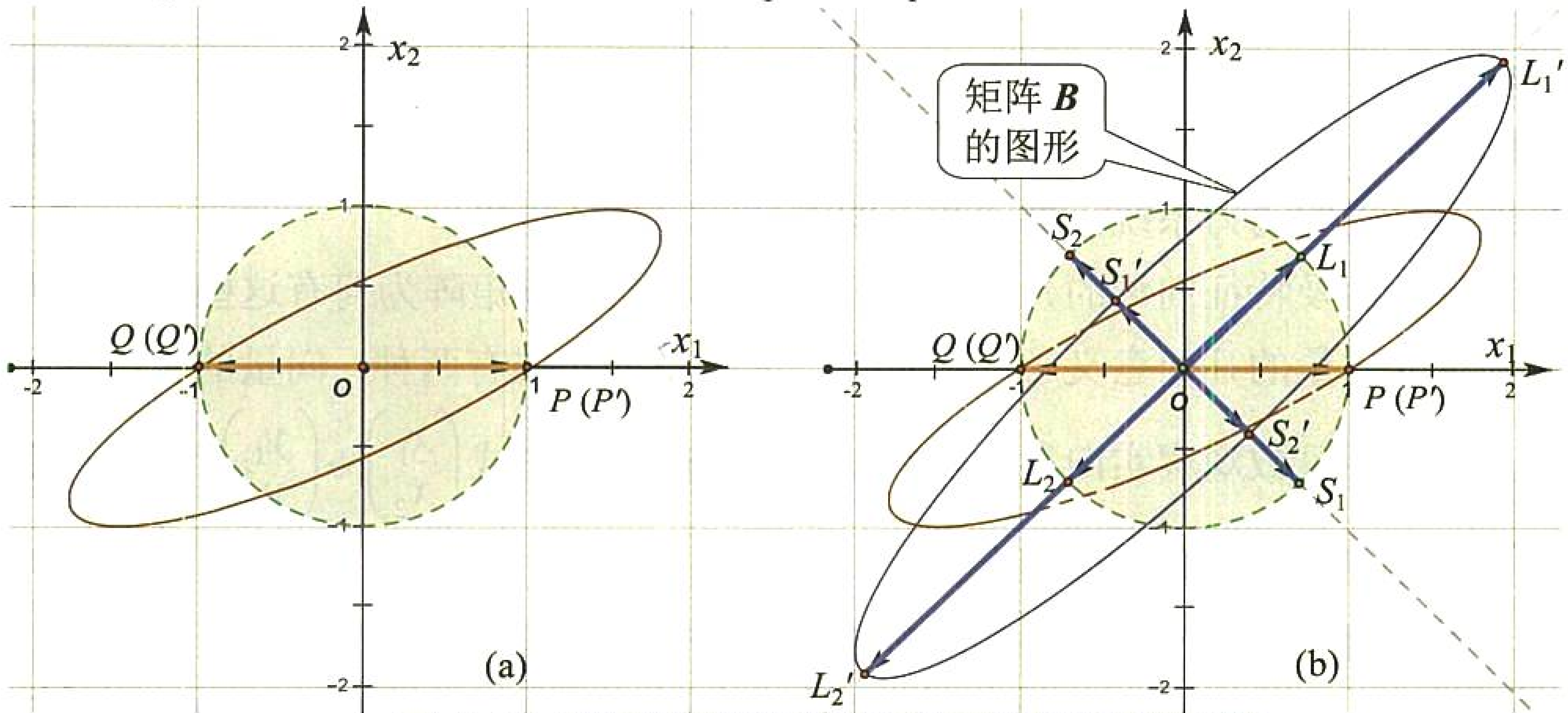


图 5-46 非对称矩阵和对称矩阵的特征向量的比较

当我们把矩阵 $\boldsymbol{A}$ 改造成类似的对称矩阵 $\boldsymbol{B}=\begin{bmatrix}1 & 1.5\\1.5 & 1\end{bmatrix}$ 时，其图形见图 5-46(b)。

矩阵 $\boldsymbol{B}$ 的特征值分别是-0.5 和 2.5。对应特征值-0.5 的特征向量中我们取两个单位向量 $\overrightarrow{oS_1}=\left(\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)$ 和 $\overrightarrow{oS_2}=\left(-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$，向量 $\overrightarrow{oS_1}$ 和 $\overrightarrow{oS_2}$ 分别乘以特征值-0.5 得到 $\overrightarrow{oS_1'}$ 和 $\overrightarrow{oS_2'}$，这两个特征向量构成了 $\boldsymbol{B}$ 矩阵椭圆形的两个短轴 $\overrightarrow{oS_1'}=\left(-\frac{\sqrt{2}}{4},\frac{\sqrt{2}}{4}\right)$ 和 $\overrightarrow{oS_2'}=\left(\frac{\sqrt{2}}{4},-\frac{\sqrt{2}}{4}\right)$，这两个向量分别是单位向量变换来的，长度缩小了 $\frac{1}{2}$，方向反转。

对应特征值 2.5 的特征向量中我们也取两个单位向量 $\overrightarrow{oL_1}=\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$ 和 $\overrightarrow{oL_2}=\left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)$，向量 $\overrightarrow{oL_1}$ 和 $\overrightarrow{oL_2}$ 分别乘以特征值 2.5 得到 $\boldsymbol{B}$ 矩阵椭圆形的两个长轴 $\overrightarrow{oL_1'}=$

$(\frac{5\sqrt{2}}{4},\frac{5\sqrt{2}}{4})$ 和 $\overrightarrow{oL_2}'=(-\frac{5\sqrt{2}}{4},-\frac{5\sqrt{2}}{4})$，这两个向量分别是单位向量变换来的，长度放大了 2.5 倍，方向相同。

我们发现，变换后的所有的 $\boldsymbol{Bx}$ 向量末端形成倾斜的椭圆图形，在这个椭圆的两个主轴上，向量 $\boldsymbol{x}$ 和变换后的向量 $\boldsymbol{Bx}$ 之间的关系满足 $\boldsymbol{Bx}=\lambda\boldsymbol{x}$。当 λ 为正时，向量 $\boldsymbol{x}$ 和向量 $\boldsymbol{Bx}$ 的方向相同，长度相差 λ 倍；当 λ 为负时，向量 $\boldsymbol{x}$ 和向量 $\boldsymbol{Bx}$ 的方向相反，长度也相差 λ 倍。也就是说，**对称矩阵 B 的特征向量刚好是在其图形的主轴方向上。**

我们还观察到，矩阵 $\boldsymbol{B}$ 的特征向量看起来是垂直的，实际上有以下的定理：

对于实对称矩阵，对应于不同特征值的特征向量是正交的；换句话说，不同特征空间的任意两个特征向量是正交的。

5.8.5 复数特征值及特征向量的几何意义

最后我们讨论当特征值和特征向量是复数的时候，如何理解其几何意义。

我们知道，n 阶方阵的特征方程是 n 次多项式，在复数域下恰有 n 个根，因此不只复矩阵有复特征值和特征向量，实矩阵也有。

复数的东西感觉很虚幻，其实很有实际应用价值，因为咱觉得复数的运算规则实际是符合自然界物质或能量在高维度空间里的变化规律的。复数特征值的研究可帮助我们发现或揭示实矩阵中隐含的信息，理解那些难以理解的结论。含有复数特征值和特征向量的矩阵常常出现在很多蕴含周期运动的实动力系统、振动和空间的某些旋转分析中。

讨论复数特征值及特征向量的几何意义可以帮助我们理解矩阵为何有这些旋转或振动。

前面讨论过复向量的几何意义（见 2.9 节）。复向量的构成有两种：向量的复数形式和复数的向量形式，为了用复数观察实向量，我们使用向量的复数形式 $\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}y_1\\y_2\end{pmatrix}\mathrm{i}$，这样把复向量分成实向量和虚向量，讨论起来比较方便。

下面将看到，对于特征值的看法将随着复数特征值的出现而改变。

前面我们讲过，一个实数特征值乘以对应的特征向量，除了使特征向量的长度有变化外，负的实数特征值还使向量的方向发生反方向的改变。其实我们可以认为特征向量在负数特征值的作用下发生了转动——弧度 π 的旋转，这将和虚数特征值使特征向量发生 π/2 或负 π/2 的旋转的逻辑相一致。进而一个实数和一个虚数合成的复数特征值将允许使特征向量发生任意可能角度的旋转。

这样仍然不会使特征向量失去“特征”的特定意义，因为复特征向量在实和虚平面的分向量进行了同样的变化（同样的幅度和转角的变化）。

我们可以从实特征向量开始对比。

二维实特征向量在两个实轴 x_1、x_2 上的分向量在特征值的作用（乘积）下，分向量作等比例的伸缩变化。这是显然的，看公式就知道了：

$$\boldsymbol{A}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\lambda\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}\lambda x_1\\\lambda x_2\end{pmatrix} \tag{5-14}$$

对于复向量，它的分向量分布在实平面和虚平面上，如果λ为实数，那么有

$$A\begin{pmatrix}x\\y\end{pmatrix}=\lambda\left(\begin{pmatrix}x_1\\y_1\end{pmatrix}+\begin{pmatrix}x_2\\y_2\end{pmatrix}\mathrm{i}\right)=\lambda\begin{pmatrix}x_1\\y_1\end{pmatrix}+\lambda\begin{pmatrix}x_2\\y_2\end{pmatrix}\mathrm{i} \tag{5-15}$$

就是说，二维复特征向量在虚、实两个平面上的分向量在特征值的作用（乘积）下，分向量也作等比例的伸缩变化。

如果特征值λ为虚数$k\mathrm{i}$，那么有

$$A\begin{pmatrix}x\\y\end{pmatrix}=k\mathrm{i}\left(\begin{pmatrix}x_1\\y_1\end{pmatrix}+\begin{pmatrix}x_2\\y_2\end{pmatrix}\mathrm{i}\right)=k\begin{pmatrix}x_1\\y_1\end{pmatrix}\mathrm{i}-k\begin{pmatrix}x_2\\y_2\end{pmatrix} \tag{5-16}$$

就是说，二维复特征向量在虚、实两个平面上的分向量在虚特征值的作用（乘积）下，分向量除了等比例的伸缩变化外，还同时进行了同角度（π/2 弧度）的旋转：实向量从实平面旋转到虚平面变成了虚向量，虚向量从虚平面旋转到实平面变成了实向量。

如果λ为有实有虚的复数，结果就是上述的综合（向量合成），仍然还是伸缩加旋转。在实平面分向量伸缩的影响下，整个复向量除了长度伸缩外，还在四维空间里进行了一个特定角度的旋转。

这个旋转/伸缩的过程发生在两个复平面上，结果体现在实和虚的两个平面上。

你说什么来着？思绪有点凌乱。那就来个例子吧。

例 5.4 假设$A=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$，那么$\mathbf{R}^2$上的线性变换$\boldsymbol{x}\mapsto A\boldsymbol{x}$将平面的向量逆时针旋转$\pi/2$。矩阵的作用是周期性的。因为在旋转 4 次后，向量又回到了它的初始位置。显然没有非零向量被映射成自身的数倍，因此A在$\mathbf{R}^2$上没有特征向量，也没有实特征值。

实际上，A的特征方程是

$$\lambda^2+1=0$$

求解，只有复数$\lambda_1=\mathrm{i},\lambda_2=-\mathrm{i}$。但如果我们让矩阵$A$作用在复数域$\mathbf{C}^2$上，那么有

$$\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}=\begin{pmatrix}\mathrm{i}\\1\end{pmatrix}=\mathrm{i}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}$$

$$\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{pmatrix}1\\\mathrm{i}\end{pmatrix}=\begin{pmatrix}-\mathrm{i}\\1\end{pmatrix}=-\mathrm{i}\begin{pmatrix}1\\\mathrm{i}\end{pmatrix}$$

因此i、$-\mathrm{i}$是特征值，$\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}$、$\begin{pmatrix}1\\\mathrm{i}\end{pmatrix}$是对应的特征向量。

图 5-47 给出了特征值 i 及其对应的特征向量$\boldsymbol{a}=\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}$的变换图形。也就是旋转矩阵$A$对特征向量$\boldsymbol{a}$的变换过程：

$$\begin{aligned}A\boldsymbol{a}&=\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}\\&=\begin{bmatrix}0&-1\\1&0\end{bmatrix}\left(\begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\-\mathrm{i}\end{pmatrix}\right)=\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}\mathrm{i}\\0\end{pmatrix}=\begin{pmatrix}\mathrm{i}\\1\end{pmatrix}\\&=\mathrm{i}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}=\boldsymbol{b}\end{aligned}$$

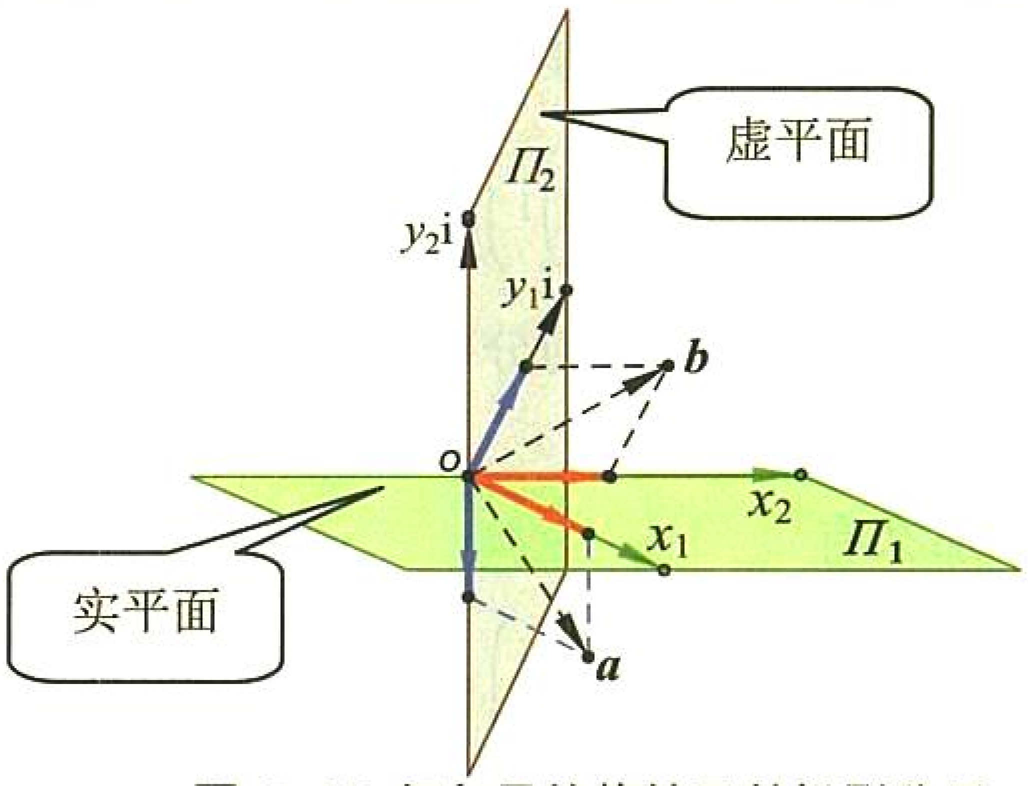


图 5-47 复向量的旋转及其投影分量

图中特征向量 $\boldsymbol{a}$ 被旋转矩阵 $\boldsymbol{A}$ 变换为 $\boldsymbol{b}=\begin{pmatrix}\mathrm{i}\\1\end{pmatrix}$，$\boldsymbol{a}$、$\boldsymbol{b}$ 向量之间为比例 i 的关系。但 $\boldsymbol{a}$ 向量的实、虚分量各自在实平面 Π_1 和虚平面 Π_2 上同时旋转了 90°，具体就是 $\boldsymbol{a}$ 的实平面上分向量 $\begin{pmatrix}1\\0\end{pmatrix}\to\begin{pmatrix}0\\1\end{pmatrix}$，虚平面上分向量 $\begin{pmatrix}0\\-\mathrm{i}\end{pmatrix}\to\begin{pmatrix}\mathrm{i}\\0\end{pmatrix}$。

在实数域中我们看不到虚平面的虚数及其变化，管中窥豹，只看到了实数空间里（实平面上）的 90° 旋转变换，这正是实矩阵 $\boldsymbol{A}$ 所对应变换的几何意义。

5.9 矩阵相似的几何意义

相似矩阵的定义是：

如果有可逆方阵 $\boldsymbol{P}$，使得方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足 $\boldsymbol{A}=\boldsymbol{PBP}^{-1}$，那么矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 被互称为相似矩阵。

多么简洁、深刻的定义啊。深刻得让俺看了 n 遍都不明白怎么俩矩阵就相似了？哪里相似了。后来听说了线性变换，又听说相似矩阵都表示同一个线性变换的话，俺的心里才有了点底。直到网上有个叫孟岩的牛人说各个相似矩阵就是给一头小猪拍的不同角度的照片！俺才总算有了感性认识。

5.9.1 什么是相似矩阵

前面讲过，线性变换用矩阵表示是与空间的一组基相联系的。一般情况下，一个线性变换就是一个描述，比如平面旋转 π/4 弧度的变换，比如四维空间对于一个平面镜像的变换等。那么要把这些线性变换转化为矩阵，就要根据情况选择某一个坐标系及其单位等。选择坐标系及其单位就是确定某一个基。所以一个线性变换在不同的基下的表示矩阵是不相同的，下面的定义及定理揭示了同一个线性变换在不同基下的矩阵之间的相互关系。

相似方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足 $\boldsymbol{A}=\boldsymbol{PBP}^{-1}$，那么矩阵 $\boldsymbol{A}$ 变换到矩阵 $\boldsymbol{B}$ 的过程，被称为矩阵的相似变换。

实际上，相似矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是同一个线性变换（在同一线性空间中）在两个不同基下的表示矩阵，而可逆矩阵 $\boldsymbol{P}$ 就是基变换矩阵。

同一个线性变换在不同基下的矩阵是相似矩阵；反之，两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 如果是相似矩阵，

并且 $\boldsymbol{A}$ 是一个线性变换在一个基下的矩阵，那么矩阵 $\boldsymbol{B}$ 必然是这个线性变换在另一个基下的矩阵。

呵呵，前面的定义、定理绕来绕去的，其实核心一句话就是：

相似矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是同一个线性变换在两个不同基下的表示矩阵。

这也是相似矩阵的几何意义啊。

打个比方说，就像两个观众看一场演出，台上演员的某一演出动作就是一个变换，是实实在在的、唯物主义的不以谁看为转移的一个变换。但是两个观众张三和李四的位置不一样，从不同角度观看，这就是取的坐标不同，基不同了。显然，基不同，看到的演员的动作也不同了。扮演猴子的演员在舞台中间从左往右翻跟头，假设演员功夫好，翻的是标准的圆周运动。左前方的张三看起来猴子的“跟头”变换是顺时针椭圆周运动，此运动表示为矩阵 $\boldsymbol{A}$；在后方一角落的李四（李四是剧团工作人员，在帷幕后面闲看）看起来猴子的“跟头”变换是逆时针椭圆周运动，此运动表示为矩阵 $\boldsymbol{B}$。两个人看到的运动应是差不多的，很相似，因此 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 称为相似矩阵（注意：不止这两人，所有的观众看到的运动都是相似矩阵）。

还有第三个人王二麻子很明智，知道两个人看的运动有些走样了。就到观众席的正中央正襟危坐观看，呵！标准的圆周运动。王二麻子告诉我们：在一大堆相似矩阵中，正面的矩阵看起来不走样，最爽（矩阵有用啊，证明了为何前排中间的位置票价最贵啊）。

关于相似矩阵的比喻，还有网友孟岩的“猪照”论，奇文共赏，节录如下：

……

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，那就干脆来个很通俗的类比。

比如有一头宠物小萌猪，你打算给它拍照片，只要你给照相机选定一个镜头位置，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。

同样地，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？同样地，你给我两个矩阵，我怎么知道这两个矩阵描述的是同一个线性变换呢？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 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 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 $\boldsymbol{P}$，使得 $\boldsymbol{A}$、$\boldsymbol{B}$ 之间满足这样的关系：

$$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$$

对线性代数稍微熟悉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 $\boldsymbol{P}$，其实就是矩阵 $\boldsymbol{A}$ 所基于的基与矩阵 $\boldsymbol{B}$ 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，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矩阵与先前的那个矩阵式相似，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。

……

（本文来自 CSDN 博客，出处：http://blog.csdn.net/go2newlife/archive/2007/12/02/1912439.aspx）。

5.9.2 矩阵相似的几何意义

进一步细究啊，因为线性变换是一个运动，描述的是一个投射的瞬态过程。因此这头猪应该是一头身手矫健的猪猪侠，在打一个降龙十巴掌的招数。矩阵呢应该是一个高速摄像机录下的一小段视频。不同位置录下的视频是相似的，最能直接表现猪猪侠的武功招数的那段视频应该是对角矩阵。

好歹前面的讲述算是相似矩阵的物理意义吧。下面咱们回归到两个基下的线性变换的讨论上来。

如图 5-48 所示，矩阵 $\boldsymbol{A}$ 表示一个线性变换，把一个向量 $\boldsymbol{x}$ 变换成另一个向量 $\boldsymbol{Ax}$:$\boldsymbol{x}\to\boldsymbol{Ax}$（这是张三看到的演出动作）；在另外一个基下（等于换了一个坐标系，改变了观察角度），同样的一个变换动作表示成了 $\boldsymbol{x}'\to\boldsymbol{Bx}$（这是李四看到的演出动作）。

从张三的坐标系变换到李四的坐标系，就要乘以一个 $\boldsymbol{P}$（图 5-48中给出的是向量 $\boldsymbol{x}$ 的变换，因此乘以 $\boldsymbol{P}^{-1}$）；现实地操作就是张三跑到李四的座位上去就可以了。小学肄业的张三不知道，他竟然使用了高等代数的矩阵乘法。

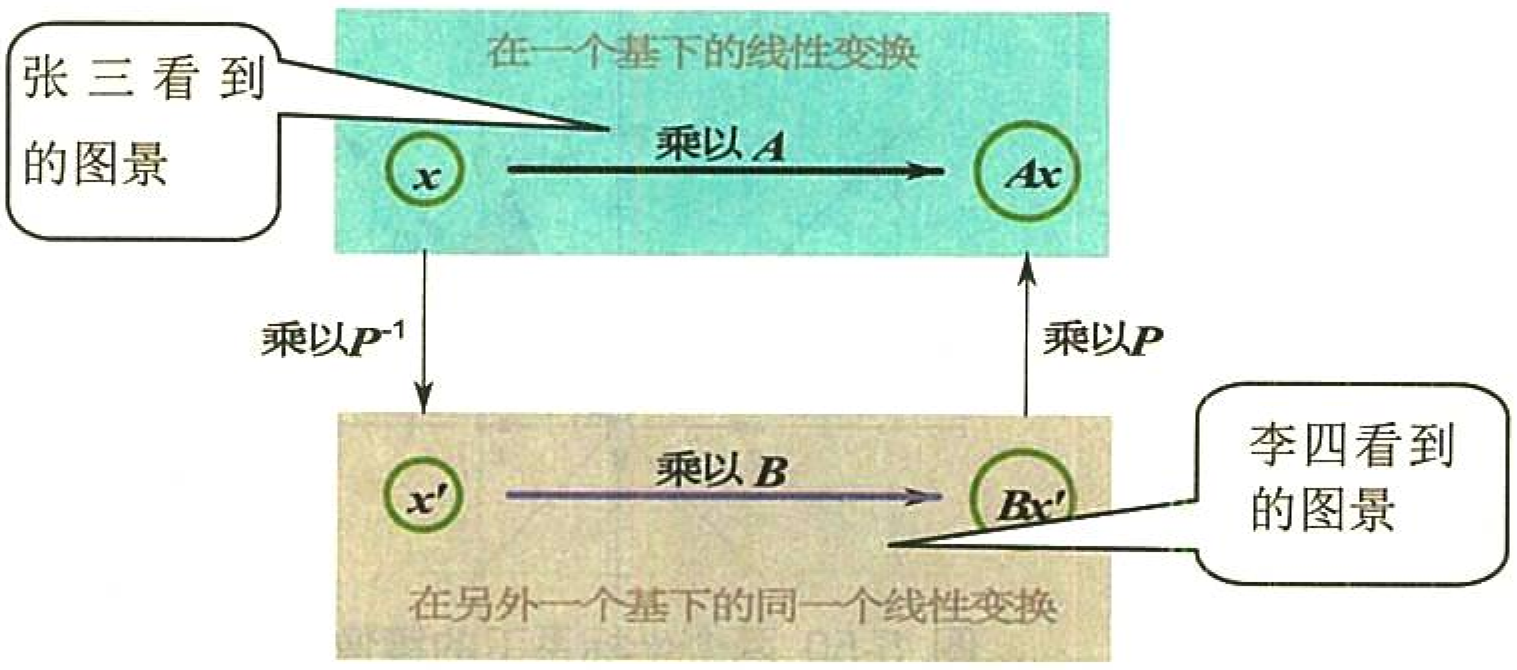


图 5-48 相似矩阵是两个基下的一个线性变换

空讲千言不如小题一例，下面我们给出一个例子来具体理解它。

例5.5 设有一线性变换：如图 5-49所示，它将任意向量$\begin{pmatrix}x\\y\end{pmatrix}$映射为关于$45^\circ$直线的镜像$\begin{pmatrix}y\\x\end{pmatrix}$。

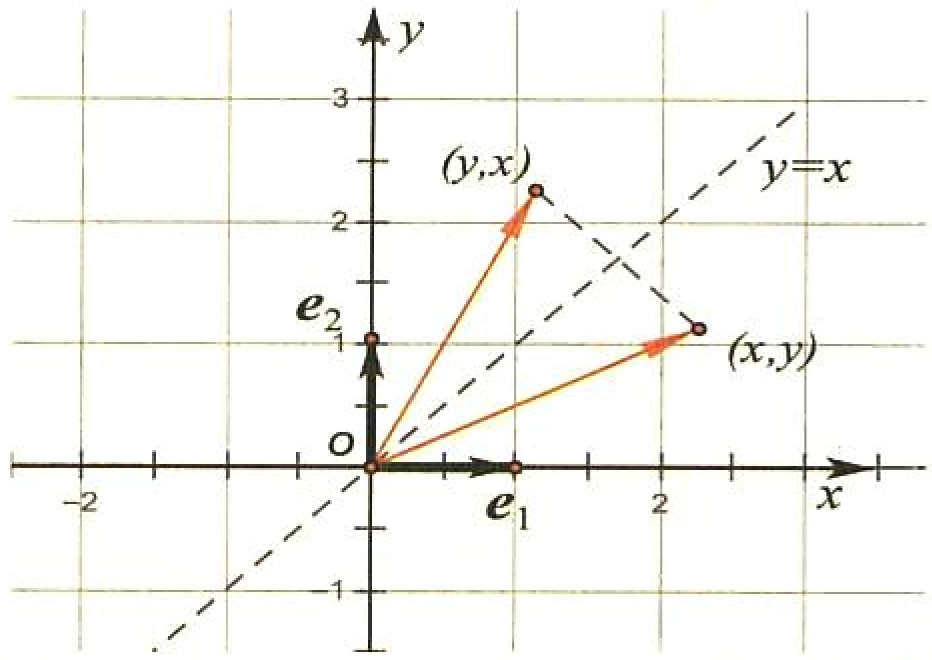


图 5-49 平面上关于45°直线：y=x 的镜像变换

取直角坐标系，其标准正交基$\boldsymbol{e}_1=\begin{pmatrix}1\\0\end{pmatrix}$和$\boldsymbol{e}_2=\begin{pmatrix}0\\1\end{pmatrix}$，根据线性变换的矩阵定理，则相应的线性变换矩阵$\boldsymbol{A}$容易求出。因为$\boldsymbol{A}$将$\boldsymbol{e}_1$映射为$\boldsymbol{e}_2$，将$\boldsymbol{e}_2$映射为$\boldsymbol{e}_1$，所以这个镜像映射在基$\boldsymbol{e}_1$和$\boldsymbol{e}_2$下的坐标表达式为

$$\begin{pmatrix}y\\x\end{pmatrix}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

其中把变换矩阵记为$[\boldsymbol{A}]_e=\begin{bmatrix}0&1\\1&0\end{bmatrix}$，表示矩阵$\boldsymbol{A}$是以$\boldsymbol{e}_i$为基的。

下面我们再找一个新的基底（见图 5-50），使得新的基向量之一$\boldsymbol{e}_1'$沿着45°直线，即$\boldsymbol{e}_1'=\begin{pmatrix}1\\1\end{pmatrix}$，而另一个基向量与之垂直，即$\boldsymbol{e}_2'=\begin{pmatrix}-1\\1\end{pmatrix}$。则新基与旧基的转换关系为

$$\begin{cases}\boldsymbol{e}_1'=\ \ \boldsymbol{e}_1+\boldsymbol{e}_2\\ \boldsymbol{e}_2'=-\boldsymbol{e}_1+\boldsymbol{e}_2\end{cases}$$

将其改写为

$$(\boldsymbol{e}_1',\boldsymbol{e}_2')=(\boldsymbol{e}_1,\boldsymbol{e}_2)\begin{bmatrix}1&-1\\1&1\end{bmatrix}$$

其中把基变换矩阵记为$\boldsymbol{P}=\begin{bmatrix}1&-1\\1&1\end{bmatrix}$。

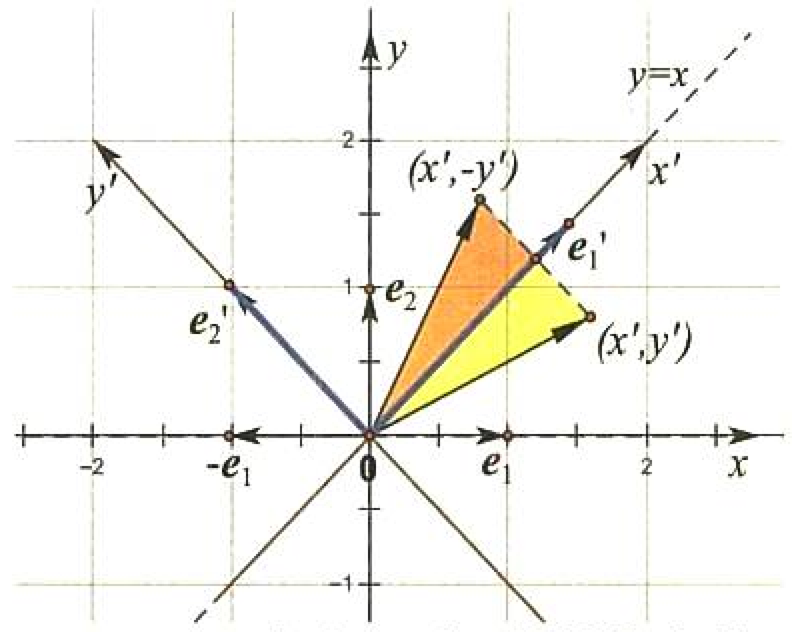


图 5-50 两个坐标系下的镜像变换

在这组新基上，这个镜像运动的线性变换A事实上被简化了。因为新基向量e_1'在45°直线上，它是它本身的镜像，即$Ae_1'=e_1'$。另一个新基向量e_2'正好被翻转过来，即$Ae_2'=-e_2'$。于是，原矩阵A所表示的线性变换在新基e_1'和e_2'下的坐标表达式为

$$\begin{pmatrix} x' \\ -y' \end{pmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}$$

其中把线性变换矩阵记为$[B]_{e'}=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$，表示矩阵$B$是以$e_i'$为基的。

这个矩阵B与单位矩阵很接近，确实比较简单。好了，下面把这个例子和前面的内容对对号：

(1) 矩阵A和B是一对相似矩阵，因为哥俩都是表示的同一变换“关于固定的一直线的镜像映射”。

(2) 矩阵P是基变换矩阵，其作用就是过河的小桥，把一个旧基换成另一个新基，乘以P；回去的小桥是乘以P^{-1}。

(3) 矩阵P的每个列都是由A的特征向量组成的，特征向量用旧基上的坐标表示。

(4) 矩阵B是简单的对角阵，对角上的元素就是特征值，从左上到右下排列的特征值分别对应着P矩阵的从左到右排列的特征向量。

(5) A和B俩矩阵既然表示同样的线性变换，因此特征值也是同样的，它和基没关系。

三角形相似的变换矩阵

咱对相似这个名词并不陌生，初中的时候就有相似三角形的说法。三角形的相似和矩阵的相似有关系吗？关系不大。俺只要把一个三角形与一个相似三角形所对应的变换矩阵给你看就清楚了。

这个变换矩阵就是被称为纯量矩阵的家伙：

$$\begin{bmatrix} k & & 0 \\ & k & \\ & & \ddots \\ 0 & & k \end{bmatrix}$$

这个矩阵就是主对角线上的元素都为数k，其他的元素全是 0。这个矩阵显然等于单位阵的k倍，因此纯量矩阵可以表示为kE。为何叫纯量矩阵？因为用纯量矩阵乘以一个矩阵A，好像用一个数k相乘的效果一样：$kE\cdot A=A\cdot kE=k\cdot A$。

好了，一个三角形ΔOAB是由三个向量$\overrightarrow{OA}=(a_1,a_2)$，$\overrightarrow{OB}=(b_1,b_2)$，$\overrightarrow{AB}=(b_1-a_1,b_2-a_2)$组成。若将每个变量都变为原来长度的$k$倍，则$\overrightarrow{OA}$、$\overrightarrow{OB}$、$\overrightarrow{AB}$分别都延长$k$倍而变为$\overrightarrow{OA'}$、$\overrightarrow{OB'}$、$\overrightarrow{AB'}$，用矩阵表示就是

$$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}\begin{bmatrix} a_1 & b_1 & b_1-a_1 \\ a_2 & b_2 & b_2-a_2 \end{bmatrix}=\begin{bmatrix} ka_1 & kb_1 & k(b_1-a_1) \\ ka_2 & kb_2 & k(b_2-a_2) \end{bmatrix}$$

5.9.3 矩阵相似对角化的几何解释

矩阵的对角化有不同的方法，但我们常常讲的矩阵的对角化，多数是所谓的相似对角化。

鉴于大家都有花大价钱买甲等票到王二麻子舞台正前方的位置看戏的习惯，数学家当然也喜欢把一个矩阵相似变换到另一个比较爽的矩阵上去进一步研究这个线性变换。这个最爽的位置就是用特征向量当坐标轴，最爽的矩阵就是对角阵，次一些的当算约当阵了。实际上，这是一个常见的工程处理方法，把一个对象从一个领域变换到另外一个领域去以便研究（当然要保证被研究对象的本质不能被变换掉）。

1. 相似对角化的好处

举个矩阵相似的事例，比如我们开车郊游（图 5-51），到乡野别院度假。从地点 $\boldsymbol{x}$ 驾驶到目的地 $\boldsymbol{Ax}$，走眼前的小道不太方便，那就绕个弯，先从 $\boldsymbol{x}$ 过桥到 $\boldsymbol{x}'$，从 $\boldsymbol{x}'$ 到 $\boldsymbol{Bx}'$ 的行车就很方便，处理简单。可到了 $\boldsymbol{Bx}'$ 还没完啊，度假目的地是 $\boldsymbol{Ax}$ 啊。好，再过桥把 $\boldsymbol{Bx}'$ 变到 $\boldsymbol{Ax}$。

确实，走笔直的马路当然比走九曲十八弯爽了。这里的马路也是对角阵。数学上的对角阵看起来爽，除了对角线上外其他的元素都是0啊，美的东西必然简洁，观赏性较强。作为数学工具，当然计算起来就比较好用了。

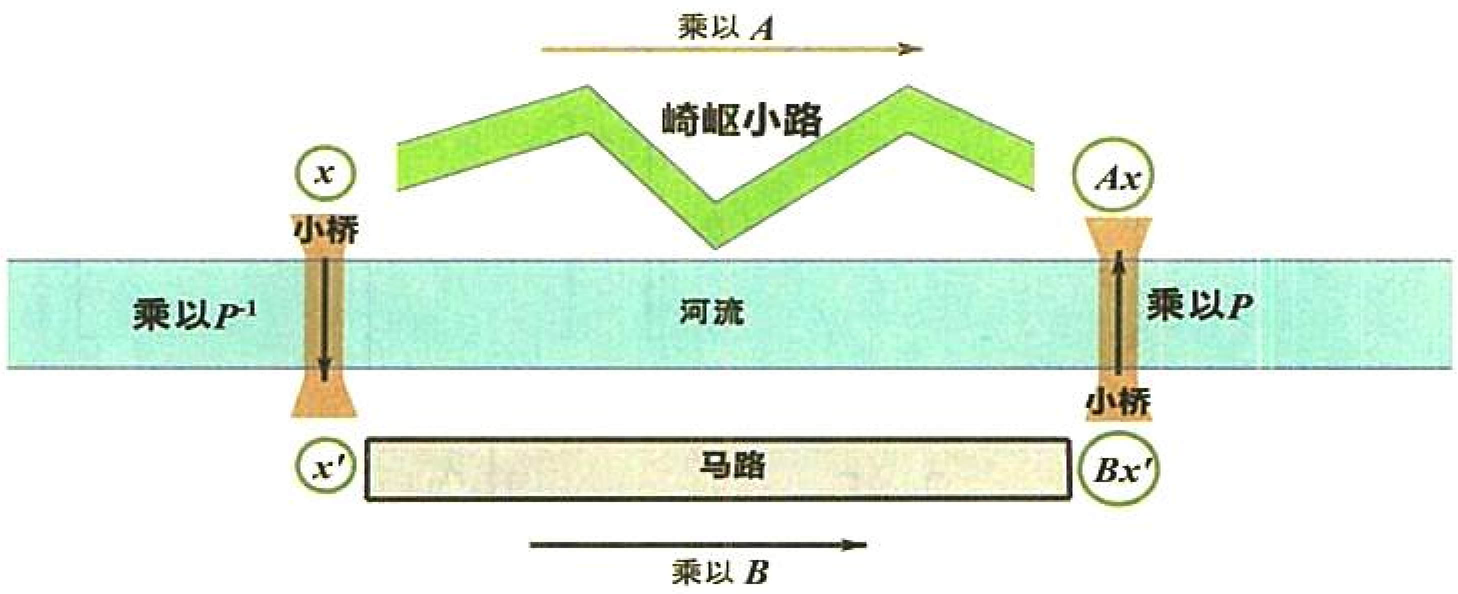


图 5-51 相似的矩阵就是走相似的路

Bx'
Bx'

假如一个矩阵对空间中一个向量作用，得到一个新向量：

$$\boldsymbol{Ad}=\begin{bmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}\begin{pmatrix}d_1\\ d_2\\ d_3\end{pmatrix}=\begin{pmatrix}a_1d_1+a_2d_2+a_3d_3\\ b_1d_1+b_2d_2+b_3d_3\\ c_1d_1+c_2d_2+c_3d_3\end{pmatrix}$$

而如果是一个对角阵对空间中一个向量作用的话，计算简便规律明了：

$$\boldsymbol{\Lambda d}=\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{bmatrix}\begin{pmatrix}d_1\\ d_2\\ d_3\end{pmatrix}=\begin{pmatrix}\lambda_1 d_1\\ \lambda_2 d_2\\ \lambda_3 d_3\end{pmatrix}$$

另外，对角阵 $\boldsymbol{\Lambda}$ 计算的简便性还体现在乘幂的时候，我们有以下的简便公式：

$$\boldsymbol{\Lambda}^n=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m\end{bmatrix}^n=\begin{bmatrix}\lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_m^n\end{bmatrix}$$

计算一般的矩阵的幂也可以利用上面的公式。实际上，如果 $\boldsymbol{A}=\boldsymbol{P\Lambda P}^{-1}$，那么有

$$\boldsymbol{A}^2=(\boldsymbol{P\Lambda P}^{-1})(\boldsymbol{P\Lambda P}^{-1})=\boldsymbol{P\Lambda}(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{\Lambda P}^{-1}=\boldsymbol{P\Lambda\Lambda P}^{-1}=\boldsymbol{P\Lambda}^2\boldsymbol{P}^{-1}$$

$$\boldsymbol{A}^3=(\boldsymbol{P\Lambda}^2\boldsymbol{P}^{-1})(\boldsymbol{P\Lambda P}^{-1})=\boldsymbol{P\Lambda}^3\boldsymbol{P}^{-1}$$

$$\vdots$$

$$A^n = P\Lambda^n P^{-1}$$

2. 相似对角化定理的推导

总结前面的例子，对角阵 Λ 的求解法当有下面的对角化定理：

$n\times n$ 矩阵 A 可对角化的充分必要条件是 A 有 n 个线性无关的特征向量。

具体讲，对于 $P^{-1}AP=\Lambda$，Λ 为对矩阵的充要条件是 P 的列向量是 A 的 n 个线性无关的特征向量。这时，Λ 的主对角线上的元素分别是 A 的对应于 P 中特征向量的特征值 λ_i，即

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

矩阵的对角化怎么就与矩阵的特征值、特征向量纠结在一起了呢？其实是因为我们可以利用特征向量的定义来得到对角阵。示意性的推导如下：

把 A 的特征向量 $P_1, P_2, \cdots, P_n$ 排成矩阵 P 的列，运用分块运算技术，把矩阵 A 作为一个整块，或作为一个常数；把矩阵 P 中每个列看做一个块（或元素），运算如下：

$$AP = A\begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix} = \begin{bmatrix} Ap_1 & Ap_2 & \dots & Ap_n \end{bmatrix}$$

好，这时候用上了特征向量的定义式 $AP=\lambda P$，上式继续有

$$AP = \begin{bmatrix} Ap_1 & Ap_2 & \dots & Ap_n \end{bmatrix} = \begin{bmatrix} \lambda_1 p_1 & \lambda_2 p_2 & \dots & \lambda_n p_n \end{bmatrix}$$

把最后的矩阵按完全不同的方式分解：

$$AP = \begin{bmatrix} P_1\lambda_1 & P_2\lambda_2 & \cdots & P_n\lambda_n \end{bmatrix} = \begin{bmatrix} P_1 & P_2 & \cdots & P_n \end{bmatrix}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix} = P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$$

显然，把最后的矩阵左乘一个 P 矩阵的逆，把 P 消简成单位阵，就得到了由特征值构成的对角阵 Λ，也即 $P^{-1}AP=\Lambda$ 成立。

为了直接理解 $P^{-1}AP=\Lambda$ 的变换过程，首先，咱回顾一下“线性变换的矩阵定理”。想起来啦，好，由此定理，我们理解一下下面似乎无聊的计算：

$$A = AE = A[e_1 \quad e_2 \quad \dots \quad e_n]=[Ae_1 \quad Ae_2 \quad \dots \quad Ae_n]$$

其中 $\{e_1, e_2, \dots e_n\}$ 为单位基向量。

啥意思？就这意思：作为一线性变换，矩阵 A 就等于自己逐个乘以坐标基向量 e_i 后得到的新矩阵。相当于对 E 内的每个列向量 e_i 进行了一次线性变换。E 就是此坐标系下的单位基向量矩阵。

这和矩阵的几何意义相符：一个矩阵就是把第一象限的单位立方体变换到其他象限的多面体，单位立方体由单位基向量张成，多面体由矩阵列向量张成；而列向量是由基向量变换得到的。

回过头来，P 是原坐标系下的基向量矩阵（但不一定是单位基向量矩阵）。一个 A 乘以 P 矩阵，相当于对 P 内的每个列向量 $\{p_1, p_2, \dots, p_n\}$ 进行了一次线性变换，也就是对基向量组进

行了一次线性变换，得到了一组新向量组$\{Ap_1, Ap_2, ..., Ap_n\}$。这组表达线性变换的向量组还在使用老的坐标系呢，好，再左乘个P^{-1}矩阵把它们统统变换到以$\{p_1, p_2, ..., p_n\}$为坐标系的空间里面去，得到向量组$\{P^{-1}Ap_1, P^{-1}Ap_2, ..., P^{-1}Ap_n\}$，把它们按序组成矩阵就是原线性变换在新基下的矩阵。

如果P矩阵的每个列向量$\{p_1, p_2, ..., p_n\}$都是特征向量，则矩阵$\{P^{-1}Ap_1, P^{-1}Ap_2, ..., P^{-1}Ap_n\}$可以简化为$[\lambda_1 e_1 \ \ \lambda_2 e_2 \ \ ... \ \ \lambda_n e_n]$，即以特征值为元素的对角阵。

3. 相似对角化的例子

例 5.6 用前面镜像变换的例子验证一下如何相似对角化。如图 5-52 所示，在原坐标系$\{xoy\}$下的镜像变换矩阵和基变换距阵分别为

$$A=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix},\quad P=\begin{bmatrix}1 & -1\\1 & 1\end{bmatrix}$$

那么相似对角化的过程$A \to AP \to P^{-1}AP = \Lambda$具体是：

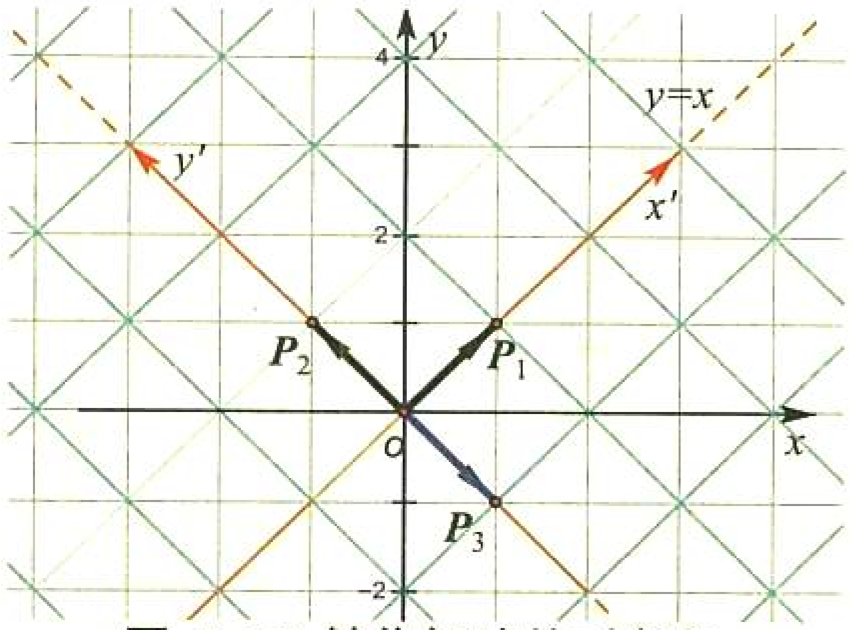


图 5-52 镜像矩阵的对角化

镜像变换A左乘矩阵P，就把特征向量组$\{p_1, p_2\}=\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$变换成了$\{p_1, p_3\}=\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\right\}$，这两个向量的元素都是原坐标系下的数值；再左乘逆矩阵$P^{-1}=\begin{bmatrix}0.5 & 0.5\\-0.5 & 0.5\end{bmatrix}$，就把$\{p_1, p_3\}$变成了新坐标系$\{x'oy'\}$下的向量组$\{p'_1, p'_3\}=\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\-1\end{pmatrix}\right\}$，把$P'_1$、$P'_3$组成矩阵，就得到了表示镜像变换的新矩阵$\Lambda=\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$。这是新坐标系$\{x'oy'\}$下的镜像变换矩阵。

从图上看到没？镜像变换很简单，就那么三个向量在那里放着没动，就我们自个儿把矩阵啊向量啊乘过来除过去的，折腾个没完——不为别的，就为了那个传说中的对角阵。

其实，我们很朴素地想一想特征向量的定义式$Ax=\lambda x$，这公式本身就是一种简化：矩阵A对特征向量x的变换结果Ax等于一个常数（特征值）λ对特征向量x的变换结果λx。这样看来，矩阵的对角简化与特征向量的联系就不突然了。实际上，想把A对角化，就要有足够的特征向量形成$\mathbf{R}^n$的一个基。

因此，线性变换的相似对角化实质是寻找一个适当的坐标系，使得该变换对这个新的坐标系上的单位向量(或基向量)只做伸缩变换、不做旋转变换。

一般实矩阵相似于对角阵的充要条件是它有n个线性无关的特征向量这一基本结论，正因为此，要知道并不是任何实矩阵都可以相似于对角阵，但实对称方阵一定可以与对角阵相似（可

对角化）。对这一点一定不要有所怀疑。

5.10 矩阵行列式的几何意义

与先前有别，本节的行列式的几何意义主要是结合矩阵的线性变换的几何意义进行的。

前面我们提到过，行列式的几何意义有两个解释：一个解释是行列式就是行列式中的行或列向量所构成的超平行多面体的**有向面积或有向体积**；另一个解释是矩阵 $\boldsymbol{A}$ 的行列式 $\det\boldsymbol{A}$ 就是线性变换 $\boldsymbol{A}$ 下的图形面积或体积的**伸缩因子**（估计还有人说，如果把矩阵看成二阶张量的话，它的行列式具有更丰富的意义——这些探究不弄了，咱这里不扯远了）。

在第 3 章中，我们主要使用行/列向量的概念详细地研究了行列式的几何意义，经过本章对矩阵的几何意义有了初步了解后，我们又如何从矩阵的角度考查它的行列式的意义呢？显然，我们更喜欢或习惯于把行列式看做矩阵作为线性变换下的图形面积或体积的**伸缩因子**。

这两个几何解释一个是静态的体积概念，一个是动态的变换比例概念，但具有相同的几何本质，因为矩阵 $\boldsymbol{A}$ 表示的（矩阵向量所张成的）几何图形相对于单位矩阵 $\boldsymbol{E}$ 所表示的单位面积或体积（即正方形或正方体或超立方体的容积等于 1，又称为体积元）的几何图形而言，伸缩因子本身就是矩阵 $\boldsymbol{A}$ 表示的几何图形的面积或体积，也就是矩阵 $\boldsymbol{A}$ 的行列式。

我们知道，对于二维平面，行列式的几何意义是行向量或列向量构成的平行四边形的有向面积。一个矩阵作用于某个图形（另一个矩阵）所引起的新图形的面积的变化，是和这个矩阵的行列式有直接关系的，行列式的值就是图形面积变化的比例值。比如，行列式的绝对值为 1，表示变换后的图形的面积不变。如果行列式的值大于 1，则表示变换后的图形面积会放大；反之就会缩小。

常常把矩阵的几何图形描绘成平行多面体是因为这个矩阵的行列式就是对应的平行多面体的有向面积值，所以矩阵所表达的几何变换也就和行列式建立了直接的联系。从行列式的值出发，我们就能知道许多关于这个矩阵的几何性质。

5.10.1 二阶矩阵行列式的几何意义

在这一节中，我们将会讨论，二阶及三阶矩阵的行列式是如何表示面积的伸缩率的。

首先讨论二阶的情况。我们前面说过，矩阵 $\boldsymbol{A}$ 将平行直线变换为平行直线，将平行四边形变换为平行四边形。假设 $\boldsymbol{A}$ 是一个列向量为 $\boldsymbol{a}$、$\boldsymbol{b}$ 的 2×2 矩阵 $\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$，那么，这里的线性变换 $\boldsymbol{A}$ 是指将 $\mathbf{R}^2$ 中的单位正方形变成 $\mathbf{R}^2$ 中以 $\boldsymbol{a}$、$\boldsymbol{b}$ 为邻边的平行四边形。

如图 5-53 所示，矩阵 $\boldsymbol{A}$ 是在标准坐标系下的某线性变换的表示。图 5-53（a），在标准坐标系下，单位标准基 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$ 张成了正方形元，面积为 1。图 5-53（b），此线性变换把正方形元变换成了大平行四边形，四边形的边就是矩阵 $\boldsymbol{A}$ 的列向量 $\boldsymbol{a}$、$\boldsymbol{b}$。在 $\boldsymbol{A}$ 的作用下，$\boldsymbol{e}_1 \to \boldsymbol{a}$，$\boldsymbol{e}_2 \to \boldsymbol{b}$。

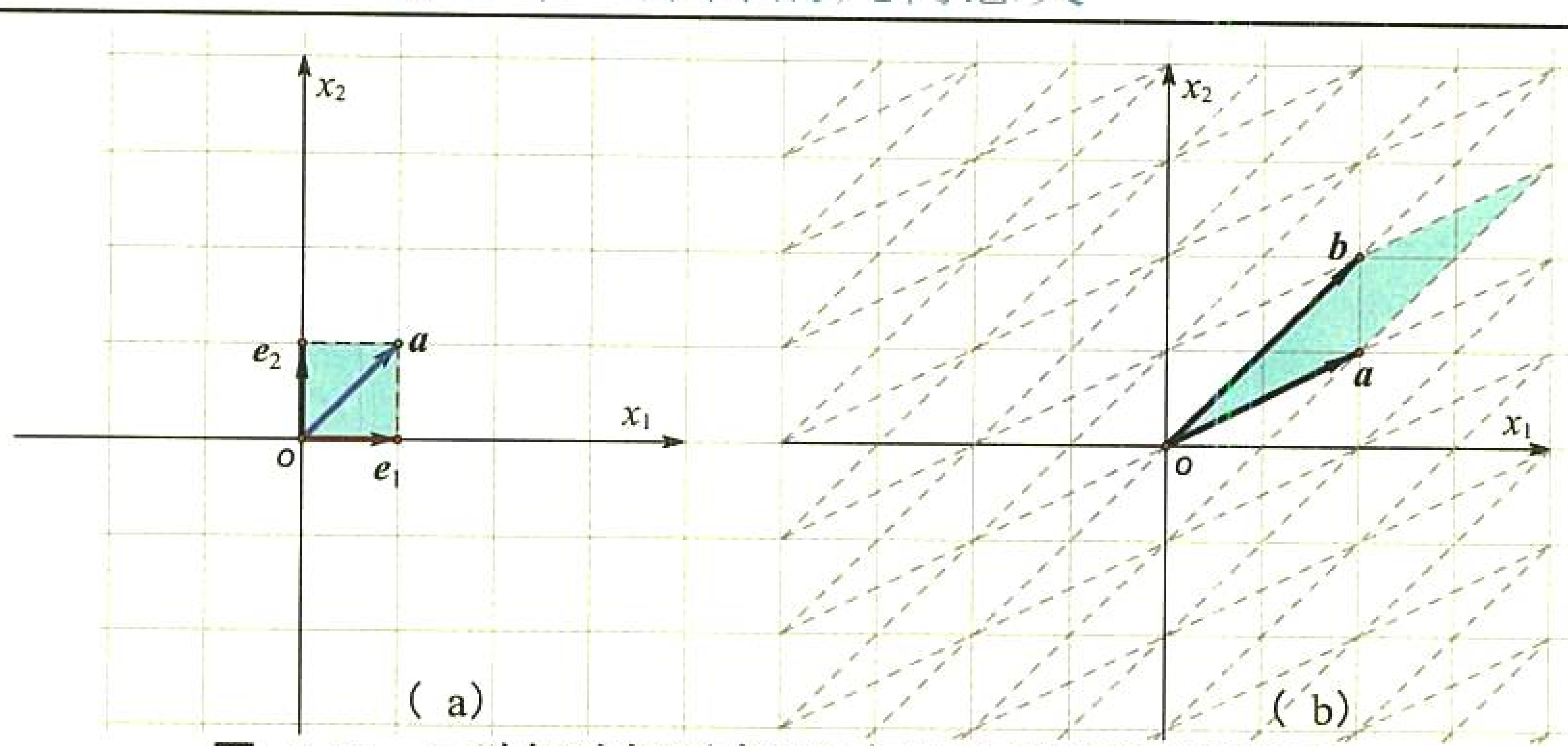


图 5-53　二阶矩阵行列式是平行四边形面积的变化率

我们已经证明平行四边形的面积等于矩阵的行列式。显然，原单位正方形的面积是 1，因此矩阵的行列式也等于由单位正方形变换为平行四边形的面积的伸缩率。

如果原图形为一个单位圆，则线性变换 $\boldsymbol{A}$ 将之变成一个椭圆。同样，椭圆的面积比上单位元的面积就是线性变换 $\boldsymbol{A}$ 的行列式。

为简单起见，我们设矩阵 $\boldsymbol{A}$ 为对角阵$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$，设向量$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$为单位圆上任一向量（见图 5-54），那么有单位圆方程$x_1^2 + x_2^2 = 1$成立。再设向量$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$为经过矩阵 $\boldsymbol{A}$ 变换后的图形上的一个向量，那么有

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

由此式得到

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} a^{-1} & 0 \\ 0 & b^{-1} \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a^{-1}y_1 \\ b^{-1}y_2 \end{pmatrix}$$

把上式带入单位圆方程得到

$$\frac{y_1^2}{a^2} + \frac{y_2^2}{b^2} = 1$$

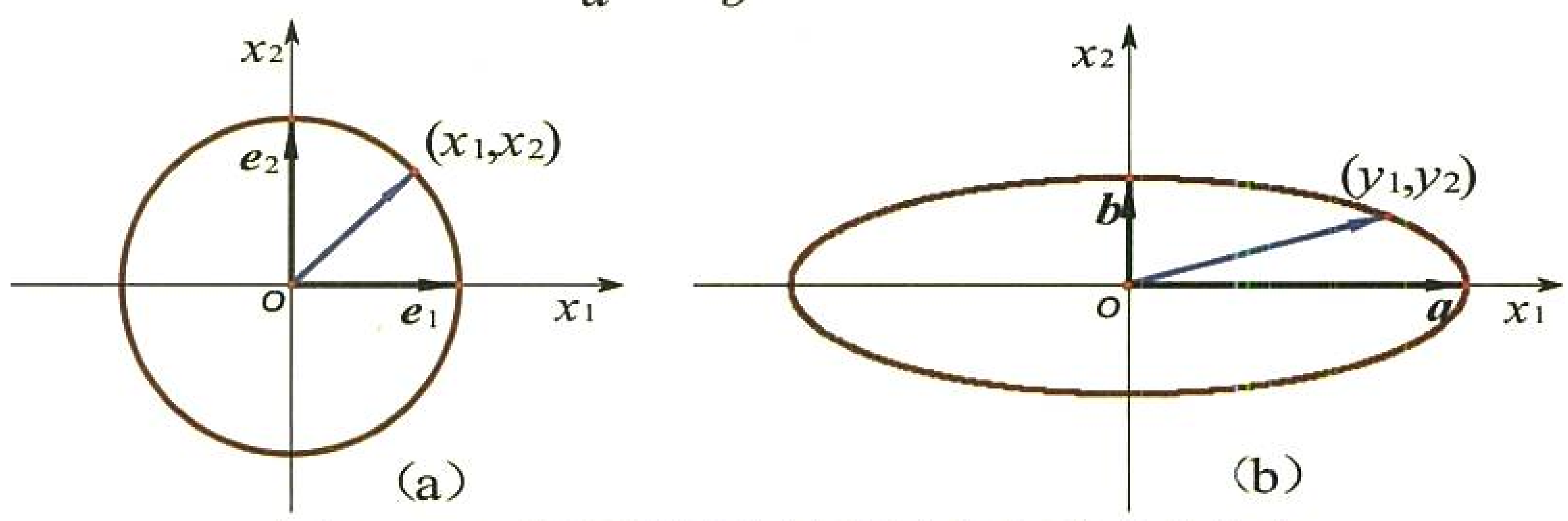


图 5-54 二阶矩阵行列式是圆/椭圆面积的变化率

这个方程正是椭圆方程的标准式。双轴长分别为a、b，其面积为 πab 。

而单位圆的面积是π，因此从单位圆变换到椭圆的面积的伸缩率是 ab，这正是行列式的值。实际上，对于平面空间上具有有限面积的任意区域，矩阵的变换所造成的面积伸缩比率同样也是行列式的值。

如图 5-55 所示，如果坐标网格的间距进一步缩小，就会越来越逼近不规则的图形，那么无穷小的方格就会变换为无穷小的平行四边形，进而得到同样的结论。

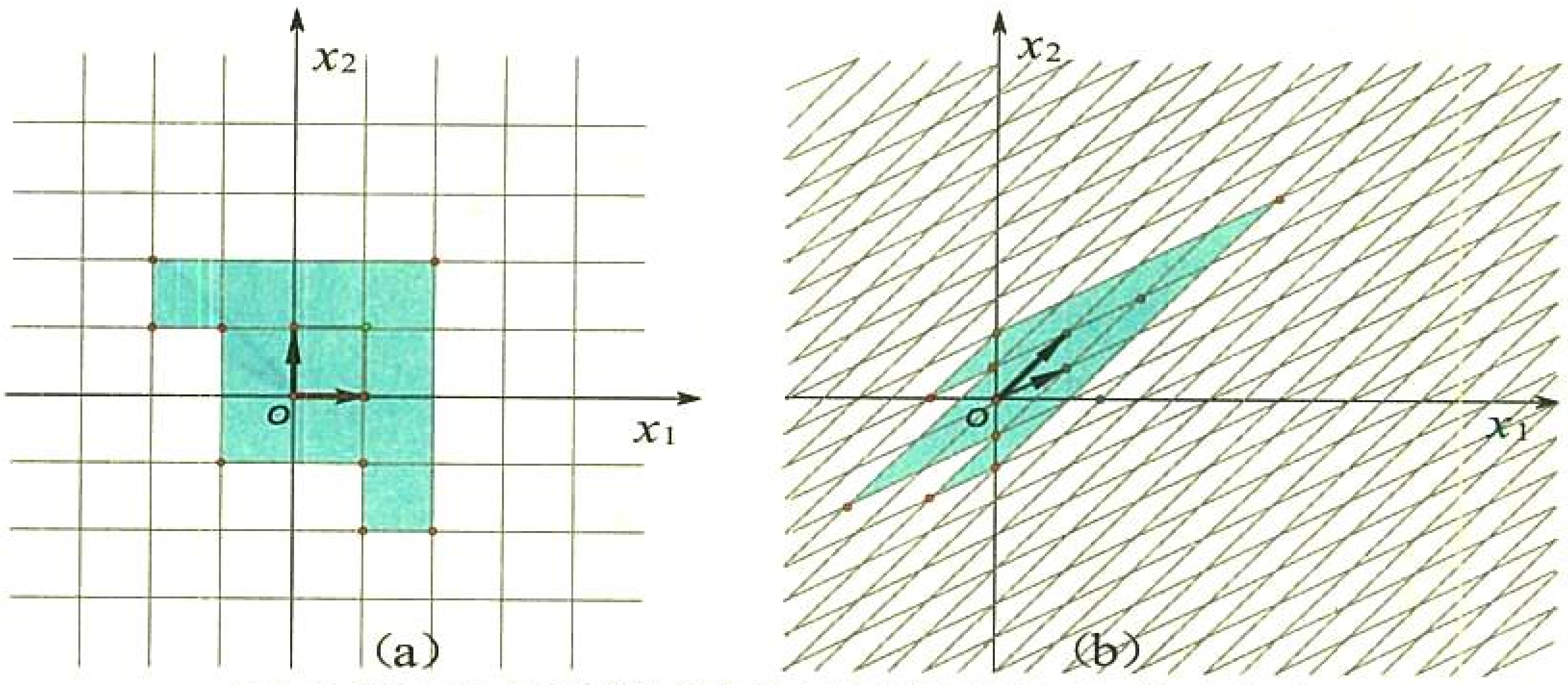


图 5-55二阶矩阵行列式的是任意图形面积的变化率

同样，在 3×3 的矩阵情形下，$\boldsymbol{A}$ 将 $\mathbf{R}^3$ 中的一个单位立方体映射成 $\mathbf{R}^3$ 中由 $\boldsymbol{A}$ 的列向量确定的平行六面体；如果原图形为一个球，则线性变换 $\boldsymbol{A}$ 将之变成一个椭球。

一般地，一个 $n\times n$ 矩阵 $\boldsymbol{A}$ 将 $\mathbf{R}^n$ 中的单位 n 立方体变成 $\mathbf{R}^n$ 中由 $\boldsymbol{A}$ 列向量确定的 n 维平行体。对非单位正方形(立方体或超立方体)以同样的方式变换，即伸缩因子为

$$\frac{\text{像域的容积}}{\text{原域的容积}}$$

而 $n\times n$ 矩阵 $\boldsymbol{A}$ 的行列式 $\det\boldsymbol{A}$ 就是这个伸缩因子。

为什么只有方阵才有行列式

当然啦，行列式定义中就要求是等行等列数的计算，只有方阵才满足啊。呵呵，有点耍赖。另外，从矩阵的线性变换的角度出发，矩阵的行列式是有向面积或体积的变换比例的意义，是同一个空间中的容积的变化率。但一个 2 行 3 列的矩阵会把一个三维向量变换成二维向量，三维空间的东东时空大挪移，整到二维空间里面去了。空间都不一样，咋比较？要比较也是比例总是 0，因为这个矩阵把一个立方体变成了一个平面。反之，一个 3 行 2 列的矩阵把二维空间中的一个平面既可以变成三维空间中的一个平面也可以变成一个平行六面体。如果平面变平面的话还可以考虑推广行列式的事；但如果平面变六面体的话，变换比率可是∞，变换比例因子一会儿 0 一会儿实数一会儿∞，这个变换的比例因子的事情就不好玩了。

5.10.2 矩阵运算的行列式的几何意义

1. 两个相似矩阵有相等的行列式

设 T 是 n 维线性空间 V 到自身的线性变换，T 在 V 的任意一组基下的变换矩阵的行列式都是相等的。设 $\boldsymbol{B}$ 是 V 的一组基，那么 T 的行列式就是 T 在 $\boldsymbol{B}$ 下的变换矩阵 $\boldsymbol{A}$ 的行列式 $|\boldsymbol{A}|$。

既然可以给线性变换定义行列式，而且线性变换的行列式就是变换矩阵的行列式，那就是说，行列式是不随基改变而改变的了。事实确实如此，我们还有下面进一步的结论：

相似矩阵是同一个线性变换在不同基下的变换矩阵，因此，两个相似矩阵有相等的行列式。

来个简单的例子吧。某个线性变换在标准坐标系下的矩阵为 $\boldsymbol{A}=\begin{bmatrix}5&3\\3&5\end{bmatrix}$，把它相似对角化得到 $\boldsymbol{\Lambda}=\begin{bmatrix}8&0\\0&2\end{bmatrix}$，基坐标系转换矩阵为 $\boldsymbol{P}=\begin{bmatrix}1&-1\\1&1\end{bmatrix}$。

上面的意思换句话就是说，矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{\Lambda}$ 是一对相似矩阵。新的坐标系的基为 $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$。

这些是我们在前面矩阵的相似对角化一节中学到的。我们先验证一下 $\boldsymbol{A}$ 和 $\boldsymbol{\Lambda}$ 的行列式是否相等：

$$|\boldsymbol{A}|=\begin{vmatrix}5&3\\3&5\end{vmatrix}=16,\quad |\boldsymbol{\Lambda}|=\begin{vmatrix}8&0\\0&2\end{vmatrix}=16$$

OK，行列式确实相等。为何相等？我们给出几何解释。如图 5-56 所示，矩阵 $\boldsymbol{A}$ 是在标准坐标系下的某线性变换的表示。图中，在标准坐标系 $\{x_1ox_2\}$ 下，单位标准基 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$ 张成了正方形元，面积为 1。此线性变换把正方形元变换成了大平行四边形，四边形的边就是矩阵 $\boldsymbol{A}$ 的列向量 $\boldsymbol{a}_1=(5,3)^{\mathrm{T}}$，$\boldsymbol{a}_2=(3,5)^{\mathrm{T}}$。四边形的面积是 16。面积的伸缩比率是 $16/1=16$，因此 $|\boldsymbol{A}|=16$。

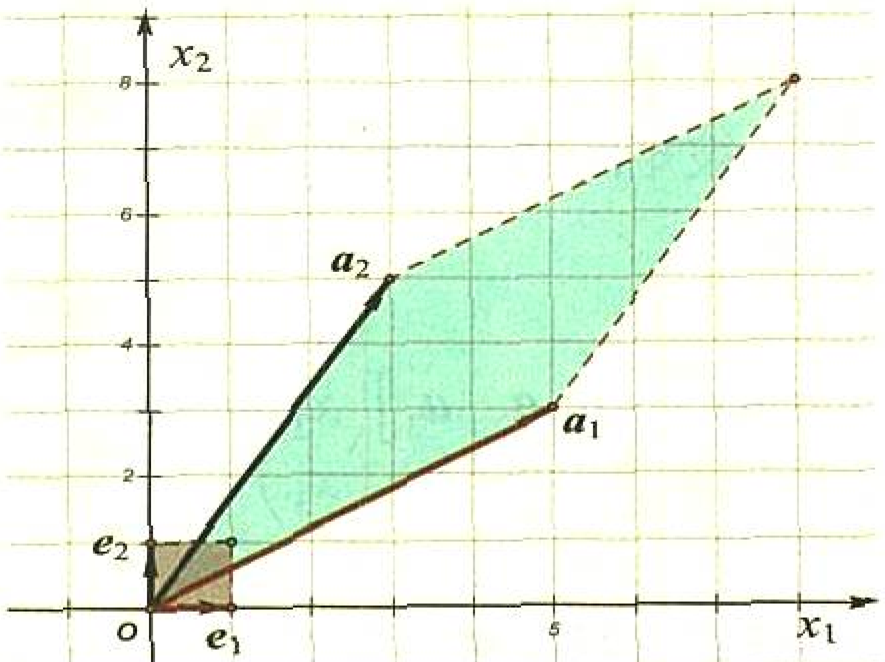


图 5-56 矩阵A的行列式的几何意义

在新的坐标系 $\{y_1oy_2\}$ 下，新基为 $\boldsymbol{P}_1=(1,1)^{\mathrm{T}}$，$\boldsymbol{P}_2=(-1,1)^{\mathrm{T}}$，同一线性变换的矩阵表达式为 $\boldsymbol{\Lambda}$。图 5-57 中，新坐标系的基 $\boldsymbol{p}_1$、$\boldsymbol{p}_2$ 也张成了正方形元，面积也为 1（注意，此面积元在老坐标系下的面积为 2）。线性变换也把正方形元变换成了大的正方形，大正方形的边就是矩阵 $\boldsymbol{\Lambda}$ 的列向量 $\boldsymbol{a}_1=(8,0)^{\mathrm{T}}$ 和 $\boldsymbol{a}_2=(0,2)^{\mathrm{T}}$。大正方形的面积也是 16。面积的伸缩比率是 $16/1=16$，因此 $|\boldsymbol{\Lambda}|=16$。

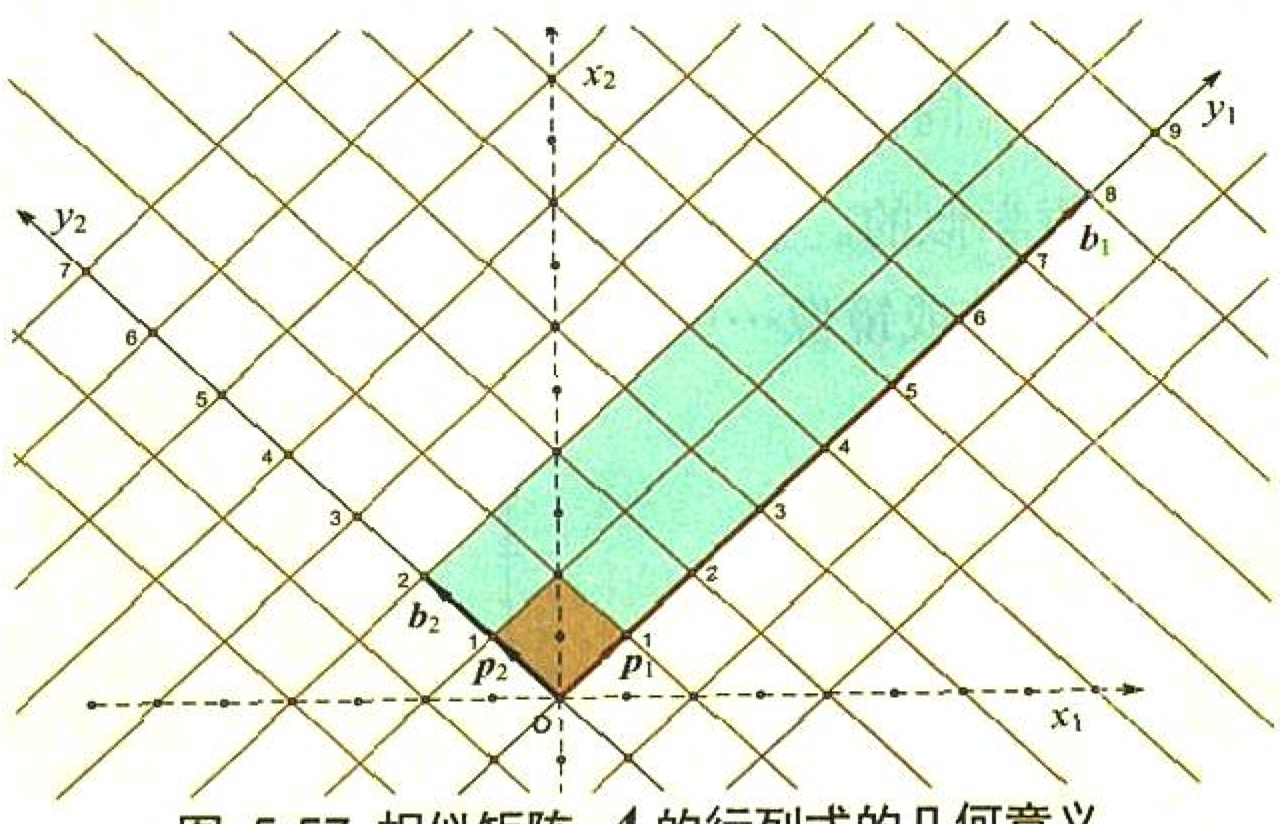


图 5-57 相似矩阵 $\boldsymbol{\Lambda}$ 的行列式的几何意义

俩相似矩阵的行列式都为 16，验证完毕。

同样，由对角阵，我们很自然地明白下面的结论：行列式是所有特征值之积；对角（或三角）阵的行列式等于对角线上元素的乘积。

2. 两个矩阵乘积的行列式等于行列式的乘积

实际上，矩阵的乘法除了把一个几何图形变换为一个新的图形外，也可以等价地看做空间的坐标系的变换。矩阵把原坐标系的基向量换成了自己的列向量，就是说矩阵的列向量成了新坐标系的基。变换后的新图形就是因为基向量的变化进而对原图形进行了连带的拉扯形成的。

这种几何意义的看法是较高级些的观点，可以有效地理解矩阵的乘法、矩阵的行列式及其他各类的变换。

好，我们现在先从矩阵乘法的代数意义出发，进而给出其几何解释，顺便给出行列式的新的几何意义出来。

先定义三维空间中标准坐标系下的两个矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$，各由三个列向量组成，即

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3\end{bmatrix}=\begin{bmatrix}a_{11} & a_{21} & a_{31}\\ a_{12} & a_{22} & a_{32}\\ a_{13} & a_{23} & a_{33}\end{bmatrix},\quad \boldsymbol{B}=\begin{bmatrix}\boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3\end{bmatrix}=\begin{bmatrix}b_{11} & b_{21} & b_{31}\\ b_{12} & b_{22} & b_{32}\\ b_{13} & b_{23} & b_{33}\end{bmatrix}$$

空间中任一向量 $\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$，被矩阵 $\boldsymbol{A}$ 变换为 $\boldsymbol{y}=\boldsymbol{Ax}$，写成具体的代数式为

$$\boldsymbol{y}=\boldsymbol{Ax}=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+x_3\boldsymbol{a}_3$$

这个式子的意义我们已说过，再重复一下啊：

如果把 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$、$\boldsymbol{a}_3$ 看做一个新坐标系的基，那么 (x_1,x_2,x_3) 就是向量 $\boldsymbol{y}$ 在新基下的坐标。

有点意思嗨，着眼点变了，结论也新鲜了：左乘一个矩阵，改变的是基，不变的是向量的坐标值。

好，记住上面的话，继续，如果向量 $\boldsymbol{y}$ 被矩阵 $\boldsymbol{B}$ 变换为 $\boldsymbol{z}=\boldsymbol{By}$，写成具体的代数式为

$$\boldsymbol{z}=\boldsymbol{By}=\boldsymbol{BAx}=\boldsymbol{B}\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{bmatrix}\boldsymbol{Ba}_1 & \boldsymbol{Ba}_2 & \boldsymbol{Ba}_3\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=x_1(\boldsymbol{Ba}_1)+x_2(\boldsymbol{Ba}_2)+x_3(\boldsymbol{Ba}_3)$$

接着前面的意思，这个式子的意义应该是，新坐标系的基被矩阵 $\boldsymbol{B}$ 又进行了变换，又变成了一组新基，相当于坐标系又改变了。同样地，基又改变了，向量的坐标还是不变，仍然是 (x_1,x_2,x_3)。

真是玄，数学咋整得像哲学似的。书的前面讲，矩阵把一个向量拉伸、旋转或镜像改变，再乘以一个矩阵，继续拉伸、旋转或镜像……是的，这是在默认基或坐标系不变的基础上讲的，符合思维惯性，见图 5-58。

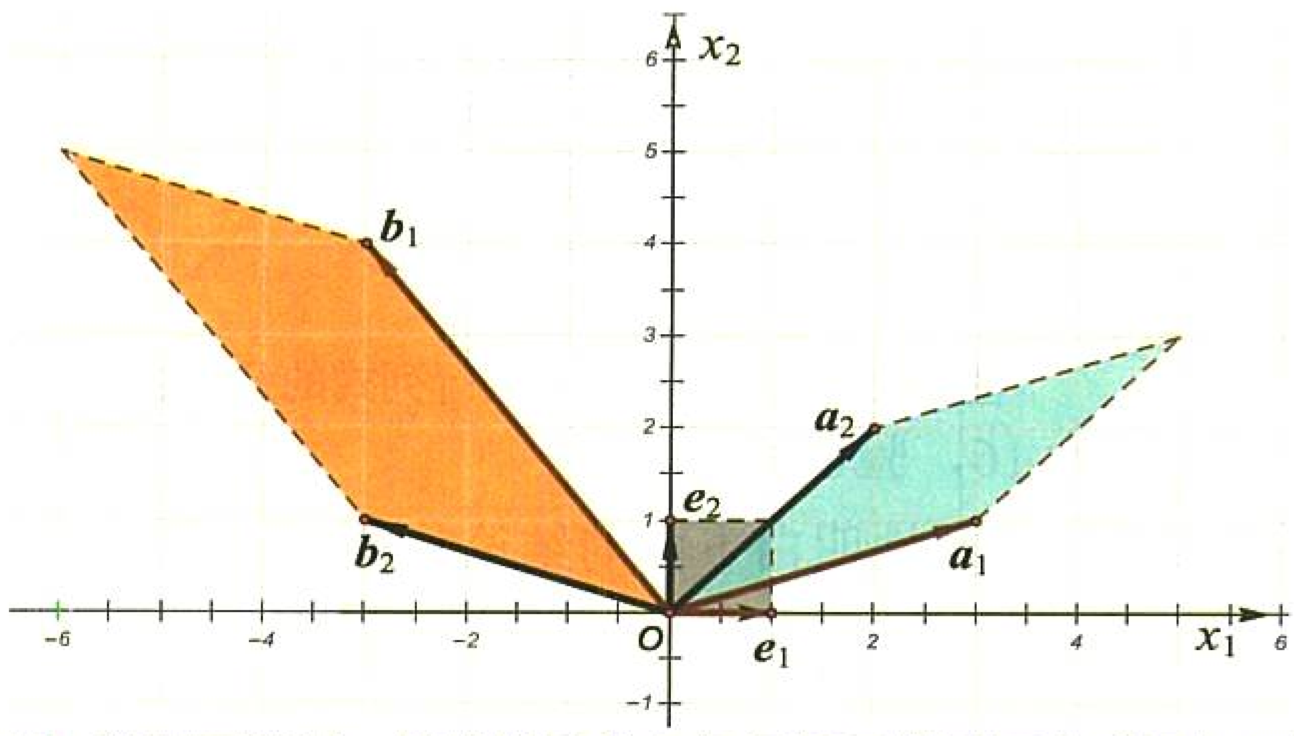


图 5-58 坐标不变时，矩阵连乘就是把平行四边形再变成平行四边形

现在想说的是，向量（注意是向量的坐标数字）总是不变，乘一次矩阵 $\boldsymbol{A}$，坐标系变一次：基向量拉伸、旋转或镜像变一次；再乘一次矩阵 $\boldsymbol{B}$，基向量继续拉伸、旋转或镜像……坐标系因此再变一次。

好，继续引申下去，标准坐标系下的单位元图形，比如平面下的边长为 1 的面积元、立体

空间下的棱长为 1 的体积元等，这些单位元都是由标准基向量张成的。这些基向量被两个以上矩阵左乘时可以看做没有改变坐标，因而可以认为单位元也没有变换，一次次改变的只是坐标系。

换句话说，单位元总是单位元；正方形、立方体是笛卡儿坐标系下的单位元，由此变换来的平行四边形、平行长方体是另一个坐标系下的单位元。

图 5-59 中，在三个坐标系 $\{x_1ox_2\}$、$\{y_1oy_2\}$、$\{z_1oz_2\}$ 下，正方形、小平行四边形、大平行四边形分别是它们的单位元。在各自的坐标单位下，其面积数值全是 1。

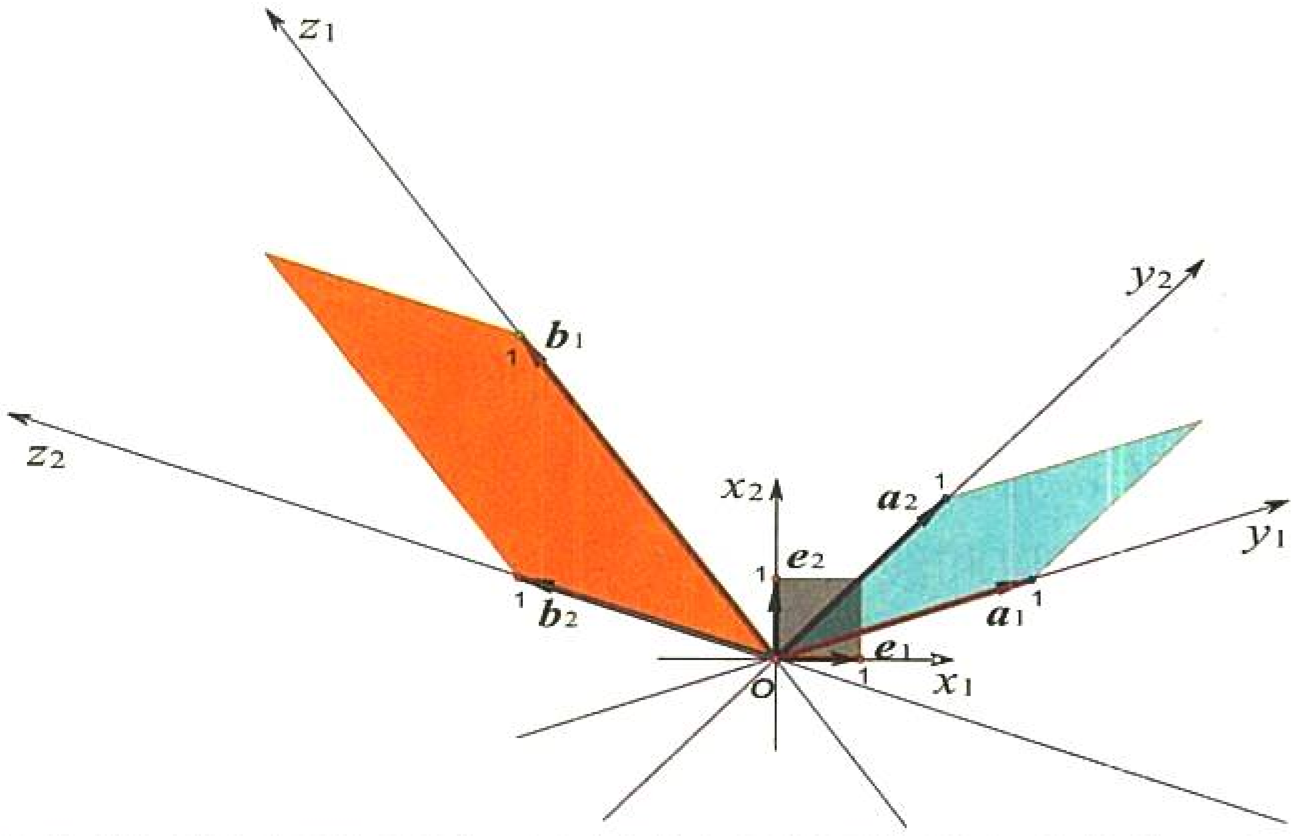


图 5-59 坐标系改变时，正方形或平行四边形都是单位面积元

在标准坐标系下或者在无处不在的世界坐标系下，这些单位元之间的关系是：:

(1) 小正方形单位元变化到小平行四边形单位元的面积比率就是变换矩阵（这里是 $\boldsymbol{A}$）的行列式。

(2) 小平行四边形单位元变化到大平行四边形单位元的面积比率就是变换矩阵（这里是 $\boldsymbol{B}$）的行列式。

(3) 因此，从小正方形单位元变化到大平行四边形单位元的面积比率就是变换矩阵（这里是 $\boldsymbol{AB}$）的行列式。

其中第(3)条就是本节的主题“两个矩阵乘积的行列式等于行列式的乘积”的几何解释。

好了，一个新的行列式的几何意义诞生了！

添加了一个单位元的概念就这样大呼小叫的，至于吗？太至于了。有了这个单位元的变换率，我们就很显然地得到了雅可比矩阵及其行列式的几何意义。

5.11 雅可比矩阵及其行列式的几何意义

因为雅可比矩阵如此重要且有趣，所以我们把它单列一节讨论，并放在矩阵的行列式的几何意义后面。

雅可比矩阵是线性代数和微积分的纽带，是把非线性问题转换为线性问题的有力工具之一。

5.11.1 雅可比矩阵及其行列式的几何意义

话说有一个函数方程组由 n 个函数组成，每个函数有 n 个自变量 $x_1, x_2, \cdots, x_n$：

$$\begin{cases} y_1 = f_1(x_1, x_2, \cdots, x_n) \\ y_2 = f_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ y_n = f_n(x_1, x_2, \cdots, x_n) \end{cases}$$

这个函数组有两个意义可以解释：一个解释它是一个映射，点 $(x_1, x_2, \cdots, x_n)$ 被映射成 $(y_1, y_2, \cdots, y_n)$；另外的一个解释就是坐标变换的意思，如果你把这个函数组代到一个以 $y_1, y_2, \cdots, y_n$ 为自变量的某方程中，即相当于把某方程的原坐标系 $\{o, y_1, y_2, \cdots, y_n\}$ 被替换成 $\{o, x_1, x_2, \cdots, x_n\}$ 坐标系。这两个解释的本质是一回事，是同一件事情从不同角度的看法。坐标系不动，一个点被变换到另一个点，这等价于说点不动，一个坐标系被代换到另一个坐标系。

下面从其坐标变换的解释角度来分析。

一般情况下，这个函数方程组不是线性方程组，它的图形多是高维曲线、曲面类的。稍详细一点说，每一个函数是个超维曲面，n 个超维曲面组合在一起交割成超维曲线。不过猛地看起来蛮像线性方程组的样子，心里于是就有了把它弄成线性方程组的冲动：弄成线性的可以使用矩阵、行列式啊什么的，可以和线性变换联系起来，多有几何意义啊。

咋弄成线性的？直接改写成矩阵形式吗，恐怕不行。嘿，不是有微积分嘛，微分就是把曲的弄成直的，积分就是把直的弄成弯的。好，对多元的非线性可微方程组进行偏微分：

$$\begin{cases} \mathrm{d}y_1 = \dfrac{\partial f_1}{\partial x_1}\mathrm{d}x_1 + \dfrac{\partial f_1}{\partial x_2}\mathrm{d}x_2 + \cdots + \dfrac{\partial f_1}{\partial x_n}\mathrm{d}x_n \\ \mathrm{d}y_2 = \dfrac{\partial f_2}{\partial x_1}\mathrm{d}x_1 + \dfrac{\partial f_2}{\partial x_2}\mathrm{d}x_2 + \cdots + \dfrac{\partial f_2}{\partial x_n}\mathrm{d}x_n \\ \vdots \\ \mathrm{d}y_n = \dfrac{\partial f_n}{\partial x_1}\mathrm{d}x_1 + \dfrac{\partial f_n}{\partial x_2}\mathrm{d}x_2 + \cdots + \dfrac{\partial f_n}{\partial x_n}\mathrm{d}x_n \end{cases}$$

到了这一步是不是和线性方程组有点相似了?！这个过程就是激动人心之化曲为直的过程。几何意义上化每个超曲面为超平面（函数 $\mathrm{d}y_i = \dfrac{\partial f_i}{\partial x_1}\mathrm{d}x_1 + \dfrac{\partial f_i}{\partial x_2}\mathrm{d}x_2 + \cdots + \dfrac{\partial f_i}{\partial x_n}\mathrm{d}x_n$ 是超维切平面方程，因此实际上就是化为超维切平面），n 个超平面组合在一起就是超维切线方程，因此就这样化曲线为直线了。代数意义上就把高次函数方程组化成了齐次线性方程组。好，那就把它写成矩阵的形式吧：

$$\begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \vdots \\ \mathrm{d}y_n \end{pmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \\ \vdots \\ \mathrm{d}x_n \end{pmatrix} \tag{5-17}$$

当里个当，主角雅可比矩阵出现了，就是式(5-17)向量方程中间的大方块。当然雅可比行列式就是雅可比矩阵的行列式。毋庸置疑，雅可比矩阵和行列式具有前面讲过的矩阵和行列式的所有意义。除此之外，还有两点特殊的地方：

一是向量的元素如 $\mathrm{d}x_1$、$\mathrm{d}y_1$ 等是微分，它们是一些极小量，而且是极小的向量；$\mathrm{d}x_i$ 是在 x_i 坐标轴上的微分向量，$\mathrm{d}y_i$ 是在 y_i 坐标轴上的微分向量。

二是雅可比矩阵里面的元素如 $\frac{\partial f_1}{\partial x_1}$、$\frac{\partial f_n}{\partial x_n}$ 等一般不是常数，而是变量，这和前面讲的矩阵不同。

恰恰是这两点，并结合矩阵的坐标系变换的意义，我们终于揭开了雅可比矩阵和行列式的最终几何意义：

雅可比矩阵把一个超平面的仿射坐标系变换成了一个超曲面坐标系；雅可比行列式就是曲面坐标系下单位微元和仿射坐标系下单位微元面积的比值。

不太明白，再换一种说法：

雅可比矩阵把一个空间里的一个平面坐标系（基）变换成了无数个极小平面坐标系（基）；无数个极小平面就是曲面的切平面；雅可比行列式就是切平面上每个坐标系下极小单位元和原坐标系下极小单位元面积的比值。

泛泛而谈让人昏昏入睡，我们找个二维的具体例子看看吧。

5.11.2　雅可比矩阵在二重积分中的应用例子

例 5.7 计算二重积分 $\iint\limits_D (y/x)^2 \mathrm{d}x\mathrm{d}y$，式中 D 为由曲线 $y=x, y=3x, xy=1, xy=5$ 所围得第一象限部分的区域。

常规的解法是在默认的直角坐标系 $\{o,x,y,z\}$ 中进行积分，其积分区域必须分为 $D=D_1+D_2+D_3$ 三个区域（见图 5-60）。先进行 $\mathrm{d}y$ 积分，然后再进行 $\mathrm{d}x$ 积分。

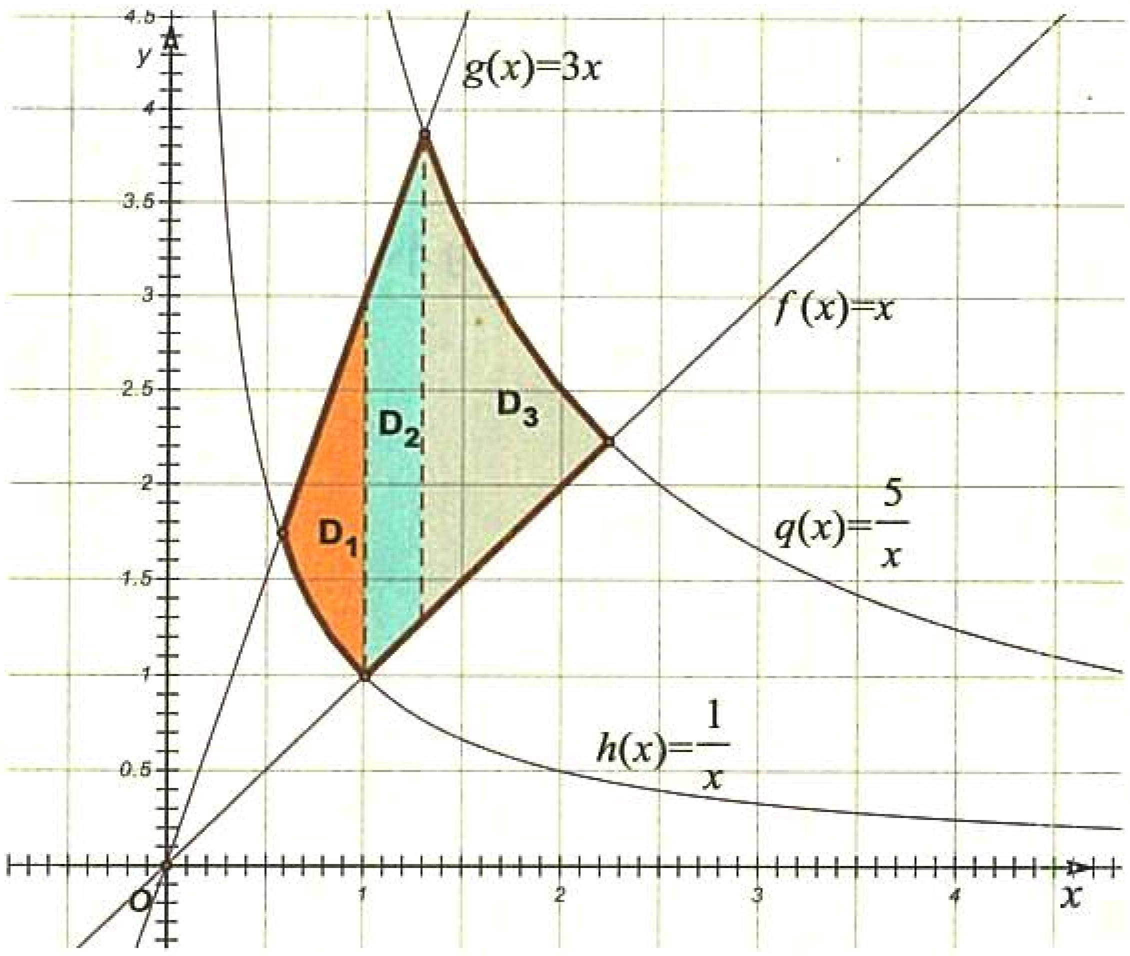


图 5-60 二重积分的区域划分

更简洁的解法是采用坐标变换法 $\begin{cases} y/x=u \\ xy=v \end{cases}$，目的是把积分区域的边界化曲为直。颠倒自变量和因变量，得到坐标变换关系为 $\begin{cases} x=\sqrt{v/u} \\ y=\sqrt{uv} \end{cases}$，这个变换函数组是非线性的，无法写成矩阵和向量的形式，对其进行偏微分，因此把变换写成雅可比矩阵形式为

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \end{pmatrix} = \begin{bmatrix} -\frac{1}{2u}\sqrt{\frac{v}{u}} & \frac{1}{2}\frac{1}{\sqrt{uv}} \\ \frac{1}{2}\sqrt{\frac{v}{u}} & \frac{1}{2}\sqrt{\frac{u}{v}} \end{bmatrix} \begin{pmatrix} \mathrm{d}u \\ \mathrm{d}v \end{pmatrix}$$

其雅可比行列式为

$$|\boldsymbol{J}|=\begin{vmatrix}-\dfrac{1}{2u}\sqrt{\dfrac{v}{u}} & \dfrac{1}{2}\dfrac{1}{\sqrt{uv}} \\ \dfrac{1}{2}\sqrt{\dfrac{v}{u}} & \dfrac{1}{2}\sqrt{\dfrac{u}{v}}\end{vmatrix}=-\frac{1}{2u}$$

那么对应的积分区域 D 的映像变为 $D'=\{(u,v)\mid 1\leqslant u\leqslant 3,\ 1\leqslant v\leqslant 5\}$，而因积分变换为

$$\iint\limits_{D}(y/x)^2\mathrm{d}x\mathrm{d}y=\iint\limits_{D'}(y(u,v)/x(u,v))^2|\boldsymbol{J}|\mathrm{d}u\mathrm{d}v=\iint\limits_{D'}u^2(-\frac{1}{2u})\mathrm{d}u\mathrm{d}v=-\int_1^3\frac{1}{2}u\mathrm{d}u\int_1^5\mathrm{d}v=-8$$

雅可比行列式的绝对值问题

在这里我们没有取雅可比行列式的绝对值带入积分中，因为这里自然地认为积分区域是有定向的，雅可比行列式 $|\boldsymbol{J}|=-\dfrac{1}{2u}$ 看起来为负值，是因为微分向量 $\mathrm{d}u$、$\mathrm{d}v$ 叉积方向和 $\mathrm{d}x$、$\mathrm{d}y$ 的叉积方向相反。但在本科生阶段，因为没有引进外微分的概念，总认为面积元 $\mathrm{d}x\mathrm{d}y$ 和 $\mathrm{d}u\mathrm{d}v$ 是正的，当然作为面积元比值的雅可比行列式也必须要求是正的了。所以你要考试的话老老实实地把雅可比行列式取绝对值吧。

好了，先把例子的计算过程摆在上面，下面我们就给出它对积分区域的坐标变换的几何解释。

首先看坐标变换前后的坐标系。图 5-61（a）是直角坐标系下的情形，单位边长的浅色方块为单位面积元，深色的以 $\mathrm{d}x$、$\mathrm{d}y$ 为边长的小方块为面积微元 $\mathrm{d}x\mathrm{d}y$，这个面积微元可以要多小就有多小。因为 $\mathrm{d}x$、$\mathrm{d}y$ 都是向量，实际上这个面积微元是向量 $\mathrm{d}x$ 和 $\mathrm{d}y$ 的叉积，因而是有向面积微元。

在经过坐标替换 $\begin{cases}y/x=u\\xy=v\end{cases}$ 后呢，原来的直角–直线坐标系被替换成了直线–曲线坐标系：直线轴是一簇经过原点的直线 $\{y/x=u\mid u\text{为不同的常数}\}$，曲线轴是一簇双曲线 $\{xy=v\mid v\text{为不同的常数}\}$，见图 5-61（b）。

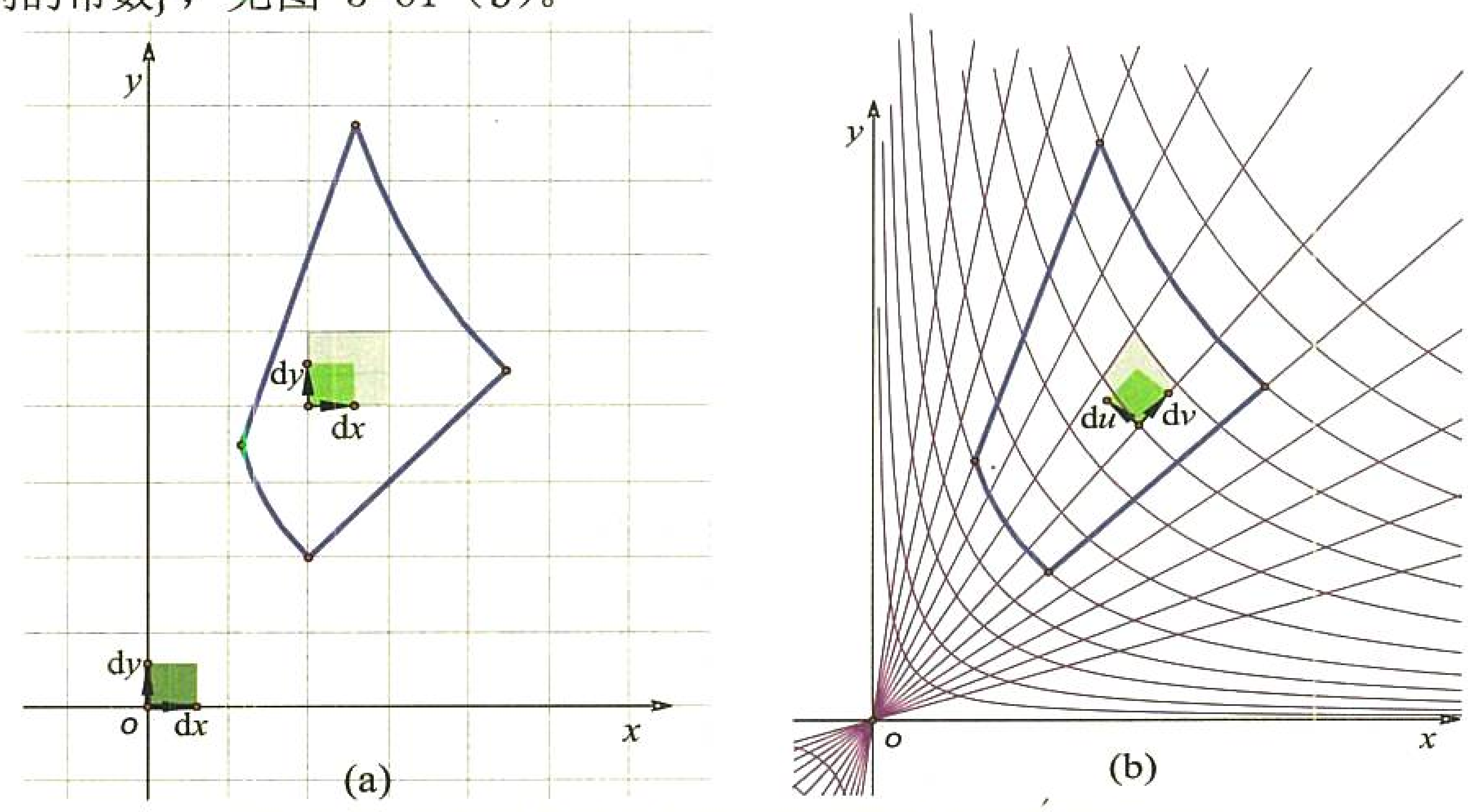


图 5-61 坐标系变换前后单位面积元和积分微元的对比

图 5-61（b）中坐标线所围成的浅色单位面积元是个曲线四边形，它的面积微元 $\mathrm{d}u\mathrm{d}v$ 是由过同一点的曲线的切线微分 $\mathrm{d}u$、$\mathrm{d}v$ 叉积张成的，这是一个深色的小平行四边形，要多小就有多小。因为 $\mathrm{d}u$、$\mathrm{d}v$ 都是向量，因而也是有向面积微元。

面积微元 $\mathrm{d}u\mathrm{d}v$ 的意思就是当这个曲线坐标轴的单位长度在取小极限时，其曲边形的面积就等于两个曲线的切线段所张成的平行四边形的面积。

好了，到了这里我们就知道了雅可比矩阵和行列式的几何意义：

二阶雅可比矩阵的几何意义就是把标准直角坐标系下的微分正方形 $\mathrm{d}x\mathrm{d}y$ 变换成了曲线坐标系下的微分平行四边形 $\mathrm{d}u\mathrm{d}v$；

二阶雅可比行列式的几何意义就是由标准直角坐标系下的微分正方形 $\mathrm{d}x\mathrm{d}y$ 所表示的面积微元变换到了曲线坐标系下的微分平行四边形 $\mathrm{d}u\mathrm{d}v$ 所表示的面积微元之比率。

那么三阶雅可比的矩阵和行列式的几何意义可以类推为：

三阶雅可比矩阵的几何意义就是把标准直角坐标系下的微分立方体 $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ 变换成了曲线坐标系下的微分平行六面体 $\mathrm{d}u\mathrm{d}v\mathrm{d}w$；

三阶雅可比行列式的几何意义就是由标准直角坐标系下的微分立方形 $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ 所表示的体积微元变换到了曲线坐标系下的微分平行六面体 $\mathrm{d}u\mathrm{d}v\mathrm{d}w$ 所表示的体积微元之比率。

到了这里雅可比矩阵和行列式的几何意义基本就清楚了。但更有意思的和微分几何有深层联系的却是所谓的曲线坐标系。大家已经看出来了，我们这里的曲线坐标系实际上被无数个仿射（直线）坐标系所替代或者模拟。因为随着积分区域内积分点的移动，点上的沿着曲线的切线在变化（方向在变换），当然切线上的微分向量 $\mathrm{d}u$、$\mathrm{d}v$ 也在变化（方向在变换），由 $\mathrm{d}u$、$\mathrm{d}v$ 构成的仿射坐标轴 $\{\mathrm{d}u,\mathrm{d}v\}$ 也在变化，也就是微分的仿射坐标系在不停地变换。

如图 5-62，如果把整个积分区域的点遍历完毕，将有无数个微小的仿射坐标系 $\{\mathrm{d}u,\mathrm{d}v\}$ 出现。当然仿射坐标的微分向量所张成的积分微元（小平行四边形）的大小或方向也在变化，也就是雅可比行列式函数 $|\boldsymbol{J}|=-\dfrac{1}{2u}$ 所给出或确定的。显然，在曲线上随着 u 值的变大点会向上移动，点的切线顺时针旋转，此点的另外一根切线方向不变，因此 $\mathrm{d}u$、$\mathrm{d}v$ 的夹角变小，所以积分微元的面积大小会变小，也就是 $|\boldsymbol{J}|$ 的绝对值会变小。

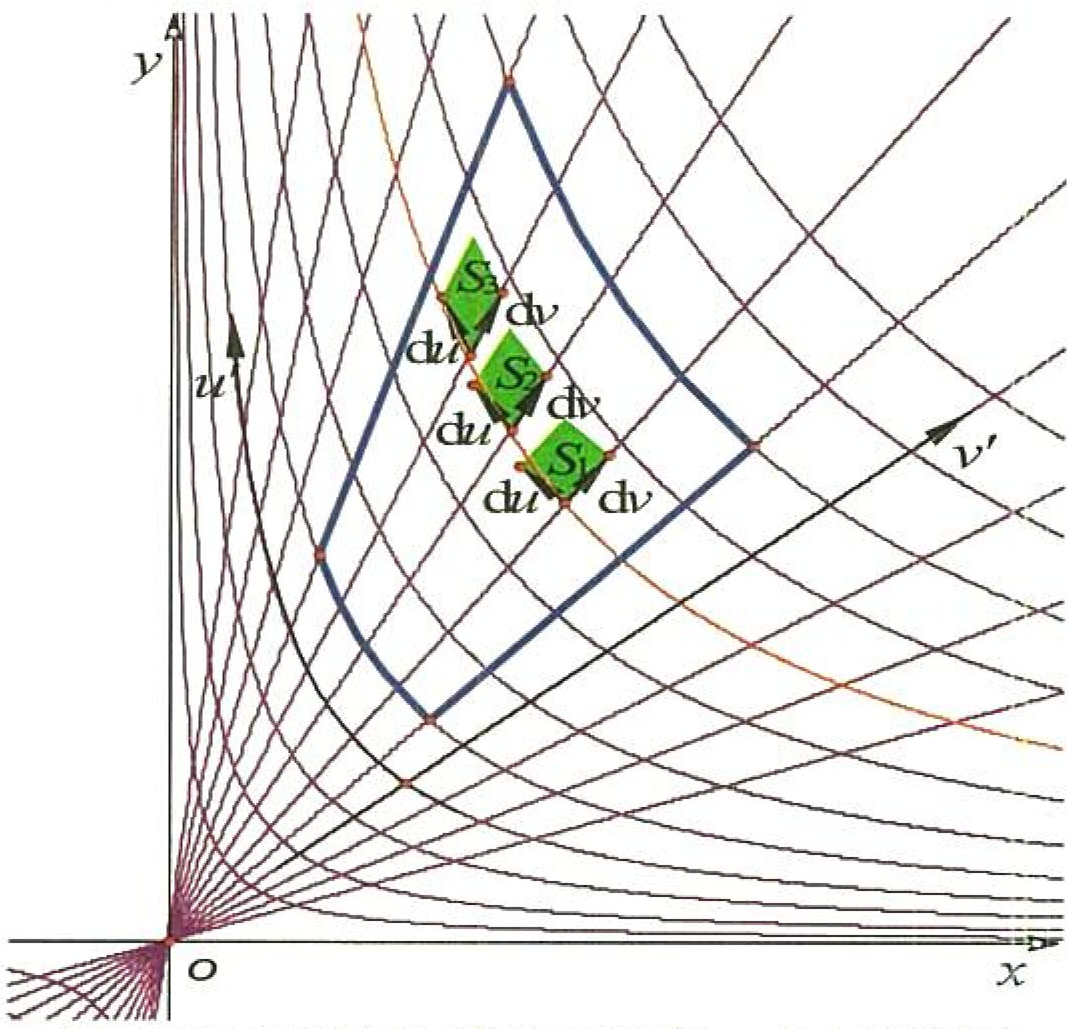


图 5-62 积分微元的面积在每一点上都不同

一句话，就是因为雅可比矩阵行列式 $|\boldsymbol{J}|=\begin{vmatrix} -\dfrac{1}{2u}\sqrt{\dfrac{v}{u}} & \dfrac{1}{2}\dfrac{1}{\sqrt{uv}} \\ \dfrac{1}{2}\sqrt{\dfrac{v}{u}} & \dfrac{1}{2}\sqrt{\dfrac{u}{v}} \end{vmatrix}$ 里面的元素是变量或函数，所

以雅可比矩阵J实际上包含有无数个具体的变换矩阵，随着点(u,v)的改变，此点附近的积分微元图形也被矩阵不停地变换着。

好了，为了更深入地理解矩阵相乘和各类矩阵的意义，我们首先探究一下矩阵对空间的旋转变换，为进一步理解更有意思的矩阵运算热热身。

5.12 矩阵对平面和空间的旋转变换

平面和空间中的旋转变换是很常见的，我们在前面的例子里也多次谈到旋转矩阵，又比如工程中我们要模拟飞机在空中的前后、左右和上下的旋转动作等。因此，弄清旋转矩阵是很有意思的。下面我们就详细讨论这个问题。

5.12.1 平面上的旋转变换

1. 平面上点的旋转变换

如图 5-63 所示，平面上任意一点$P(x,y)$对应的向量$\overline{oP}$（与原点o相连接得到），以逆时针方向绕原点在平面上旋转θ角，得到向量$\overline{oP'}$，即点$P(x,y)$在平面上以逆时针方向绕原点旋转θ角，变化到点$P'(x',y')$。

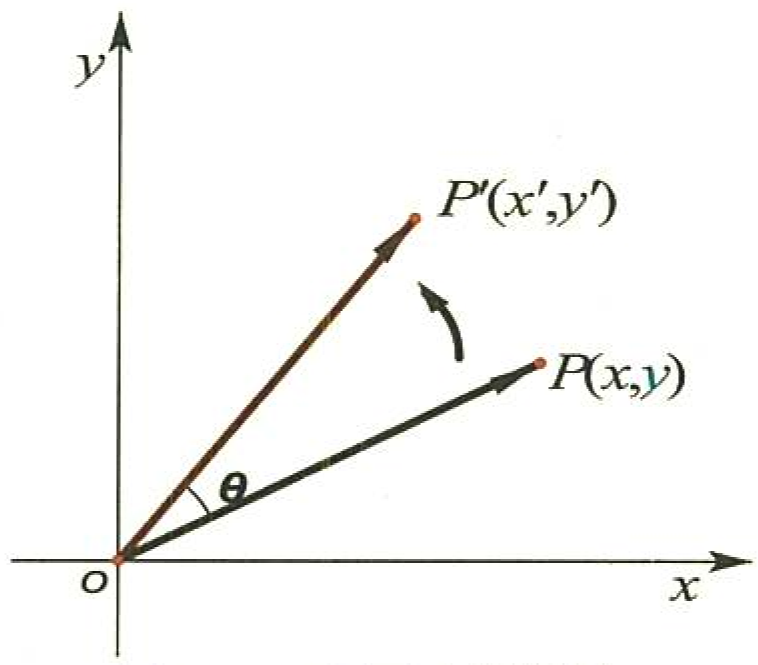


图 5-63 平面点的旋转

如果我们记这个变换为$\boldsymbol{\Gamma}$，那么有$\overline{oP'}=\boldsymbol{\Gamma}\overline{oP}$。实际上这个点的旋转变换$\boldsymbol{\Gamma}$就是前面我们介绍的旋转矩阵$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$，即点（或向量）的旋转变换为

$$\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

2. 平面上坐标的旋转变换

如果将坐标系$\{xoy\}$也以逆时针方向绕原点旋转θ角，会得到新的坐标系$\{x'oy'\}$，如图 5-64 所示。旧坐标轴上的基本单位向量$\boldsymbol{i}$和$\boldsymbol{j}$变为新坐标轴上的基本单位向量$\boldsymbol{i}'$和$\boldsymbol{j}'$，即$\boldsymbol{i}'=\boldsymbol{\Gamma}\boldsymbol{i}$，$\boldsymbol{j}'=\boldsymbol{\Gamma}\boldsymbol{j}$。实际上，此时的旋转效果是最终对坐标系$\{xoy\}$和向量$\overline{oP}$一起做了旋转$\theta$角的操作。

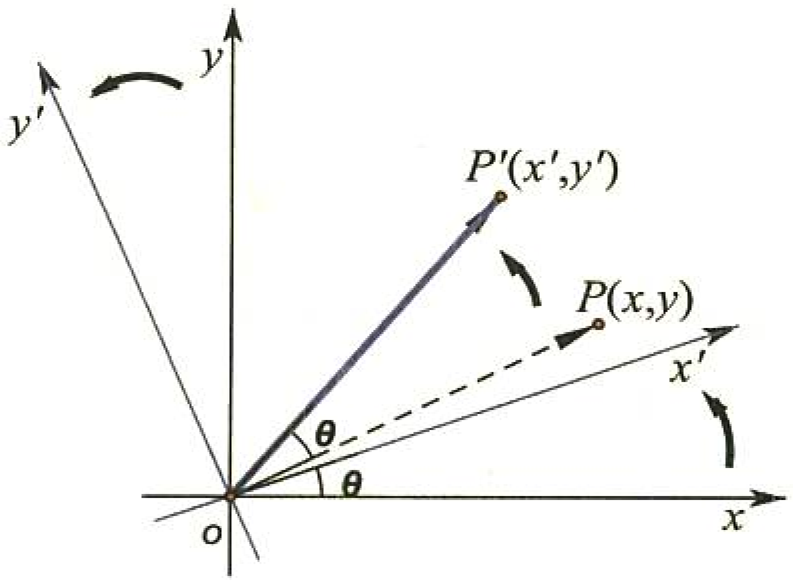


图 5-64 平面坐标系的旋转

考察基本单位向量的变化，容易检验 $\boldsymbol{i}'$ 关于旧坐标轴的坐标为（$\cos\theta, \sin\theta$），即

$$\boldsymbol{i}' = \boldsymbol{i}\cos\theta + \boldsymbol{j}\sin\theta$$

同理，有

$$\begin{aligned}\boldsymbol{j}' &= \boldsymbol{i}\cos(\theta + \frac{\pi}{2}) + \boldsymbol{j}\sin(\theta + \frac{\pi}{2}) \\ &= -\boldsymbol{i}\sin\theta + \boldsymbol{j}\cos\theta\end{aligned}$$

即

$$\begin{cases}\boldsymbol{i}' = \ \ \boldsymbol{i}\cos\theta + \boldsymbol{j}\sin\theta \\ \boldsymbol{j}' = -\boldsymbol{i}\sin\theta + \boldsymbol{j}\cos\theta\end{cases} \quad 或 \quad \begin{pmatrix}\boldsymbol{i}' \\ \boldsymbol{j}'\end{pmatrix} = \begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}\begin{pmatrix}\boldsymbol{i} \\ \boldsymbol{j}\end{pmatrix}$$

式中，$\begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}$称为坐标系（单位向量）旋转变换矩阵，记为 $\boldsymbol{T}$。

我们看到点的旋转矩阵和坐标系的同样旋转的旋转矩阵 $\boldsymbol{T}$ 不同，容易验证，它们互为转置矩阵。

另外，也可以验证它们互为逆矩阵。因为坐标系的坐标轴 （只是坐标轴，不包括平面上的点或者向量）作逆时针旋转 θ 角就相当于平面上的点或者向量（不包括坐标轴）作顺时针旋转 θ 角，如图 5-65 所示。

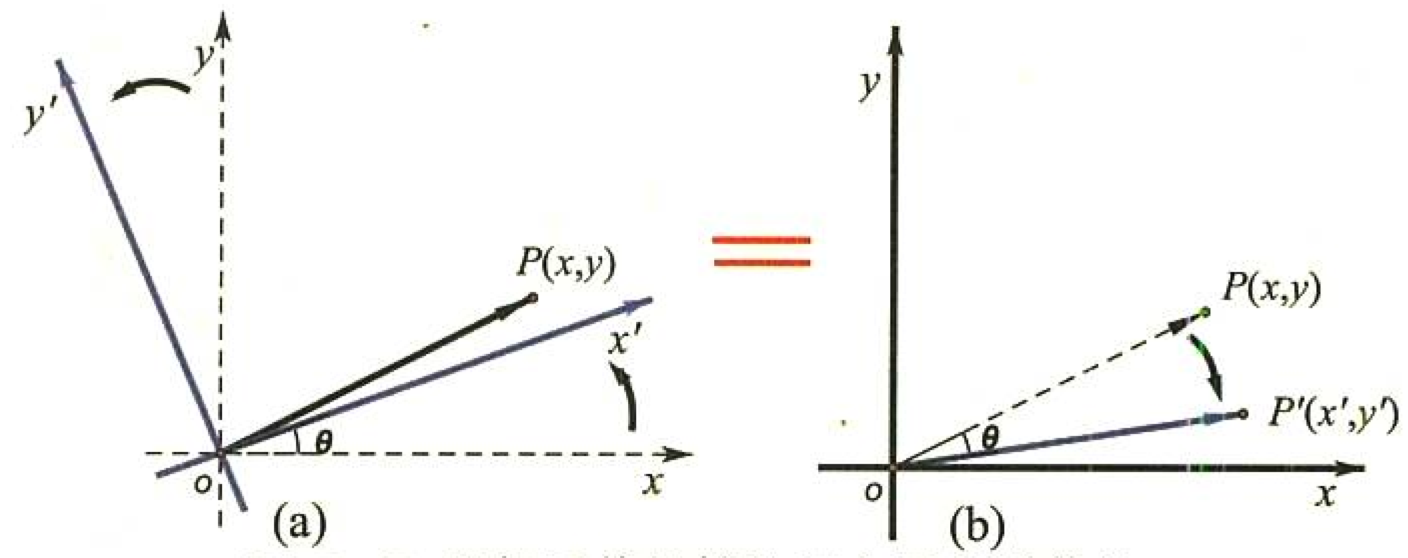


图 5-65 坐标系旋转等价于向量的逆旋转

坐标系的旋转的变换公式可以通过作图得到，这个图也可以看做这个变换矩阵的几何解释。设原坐标系是$\{xoy\}$， 这个坐标系以原点 o 为固定点进行逆时针旋转 θ 角，得到新的直角坐标系 $\{x'oy'\}$，P 点保持不动，则旋转的图像如图 5-66（a）所示。

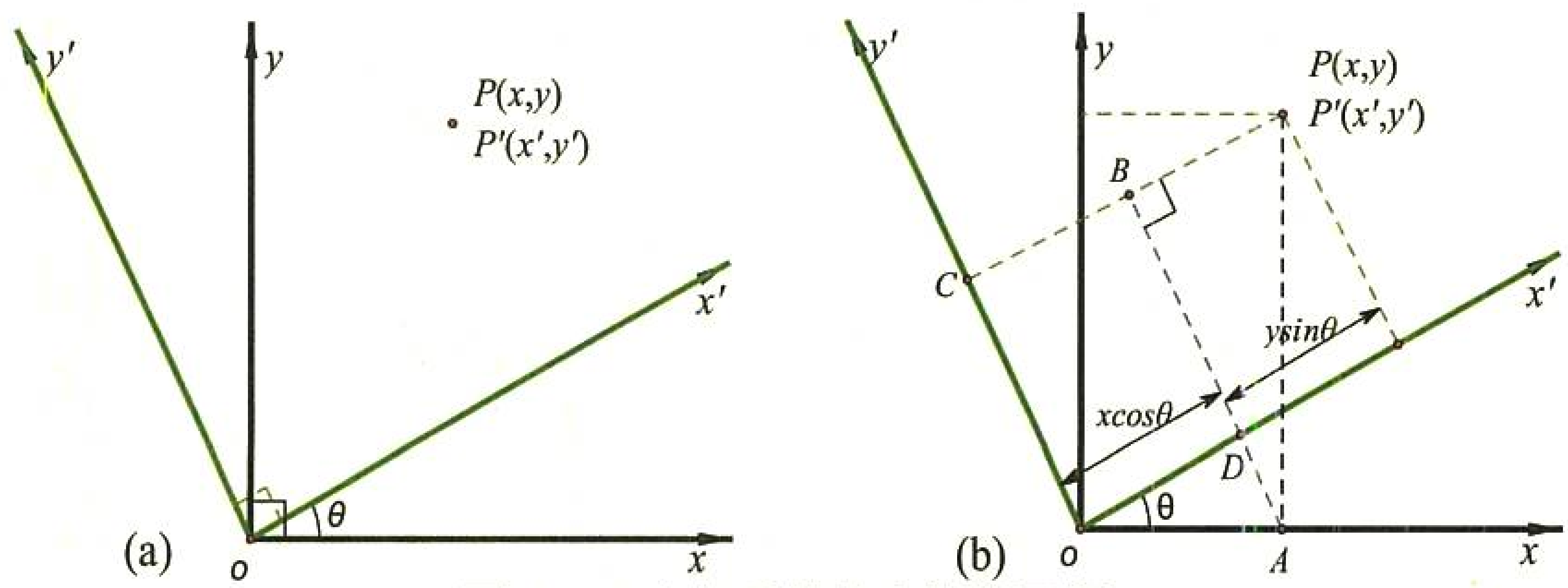


图 5-66 坐标系旋转变换的图解

为分析 P 点在两个坐标系中的坐标值之间的变换关系，从 P 点分别作四个坐标轴的垂线，然后作 AB 垂直于 PC 线段，得到图 5-66（b）所示的图形，则 P 点在 $\{x'oy'\}$ 坐标系下的值分别是

$$\begin{cases} x' = PB + BC \\ y' = AB - AD \end{cases}$$

把式中线段 PB 等的长度分别用三角式子替换，即可得到前面的旋转变换的公式。

5.12.2 空间的旋转变换

三维空间的情况完全类似，如图 5-67 所示，将空间中任意一点 $P(x,y,z)$ 对应的向量 $\overrightarrow{oP}$（与原点 o 相连接得到）以逆时针方向绕某一个直线 L（过原点）旋转 θ 角，得到向量 $\overrightarrow{oP'}$，即点 $P(x,y,z)$ 变化到点 $P'(x',y',z')$。

在开始的时候，如果将整个空间作为一个刚体绕直线 ℓ 旋转 θ 角，那么点 $P(x,y,z)$ 当然变化到点 $P'(x',y',z')$，而旧坐标系 $\{oxyz\}$ 变换到新的坐标系 $\{ox'y'z'\}$，旧坐标轴上的基本单位向量 $\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 变为新坐标轴上的基本单位向量 $\boldsymbol{i}'$、$\boldsymbol{j}'$、$\boldsymbol{k}'$。

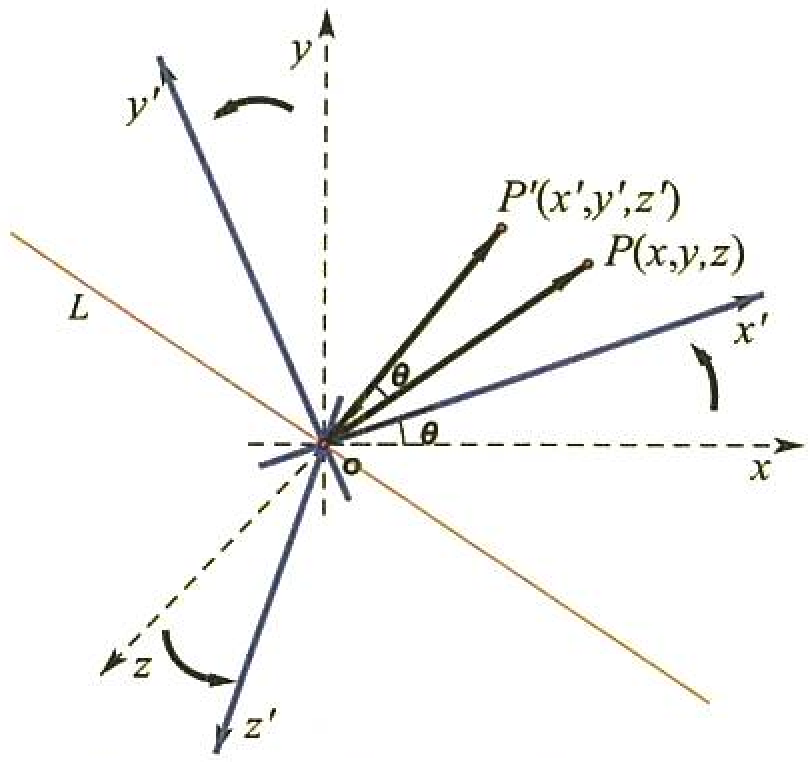


图 5-67 三维空间的旋转变换

因此，点 P' 关于新坐标系 $\{ox'y'z'\}$ 的位置关系恰如点 P 关于旧坐标系 $\{oxyz\}$ 的位置关系，从而有：

$$\overrightarrow{oP'} = x'\boldsymbol{i} + y'\boldsymbol{j} + z'\boldsymbol{k} = x\boldsymbol{i}' + y\boldsymbol{j}' + z\boldsymbol{k}'$$

若令 x'、y'、z' 关于坐标系 $\{0xyz\}$ 的方向余弦分别是 a_1、b_1、c_1，a_2、b_2、c_2 和 a_3、b_3、c_3，那么坐标系轴的旋转变换为

$$\begin{cases} \boldsymbol{i}' = a_1\boldsymbol{i} + b_1\boldsymbol{j} + c_1\boldsymbol{k} \\ \boldsymbol{j}' = a_2\boldsymbol{i} + b_2\boldsymbol{j} + c_2\boldsymbol{k} \\ \boldsymbol{k}' = a_3\boldsymbol{i} + b_3\boldsymbol{j} + c_3\boldsymbol{k} \end{cases} \quad 或 \quad \begin{pmatrix} \boldsymbol{i}' \\ \boldsymbol{j}' \\ \boldsymbol{k}' \end{pmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{pmatrix} \boldsymbol{i} \\ \boldsymbol{j} \\ \boldsymbol{k} \end{pmatrix}$$

下面看看点的坐标 x'、y'、z' 和 x、y、z 的变换关系。由于

$$\begin{aligned} \overrightarrow{oP} = x'\boldsymbol{i} + y'\boldsymbol{j} + z'\boldsymbol{k} = x\boldsymbol{i}' + y\boldsymbol{j}' + z\boldsymbol{k}' \\ &= x(a_1\boldsymbol{i} + b_1\boldsymbol{j} + c_1\boldsymbol{k}) + y(a_2\boldsymbol{i} + b_2\boldsymbol{j} + c_2\boldsymbol{k}) + z(a_3\boldsymbol{i} + b_3\boldsymbol{j} + c_3\boldsymbol{k}) \\ &= (a_1x + a_2y + a_3z)\boldsymbol{i} + (b_1x + b_2y + b_3z)\boldsymbol{j} + (c_1x + c_2y + c_3z)\boldsymbol{k} \end{aligned}$$

通过比对等式两边有

$$\begin{cases} x' = a_1x + a_2y + a_3z \\ y' = b_1x + b_2y + b_3z \\ z' = c_1x + c_2y + c_3z \end{cases} \quad \text{或} \quad \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

式中的点坐标的变换矩阵和坐标轴的变换矩阵（方向余弦矩阵）是一对转置矩阵。实际上，我们还有

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

这个实际上就是坐标轴旋转矩阵的坐标表示式。这说明这一对转置矩阵也是一对逆矩阵。坐标的旋转也就是正交变换，正交矩阵的定义就是逆矩阵等于转置矩阵。

转置矩阵的几何意义

本节我们看到，在平面或空间的旋转变换中，空间里元素（如向量或点）的旋转矩阵和对应的空间坐标系（坐标轴）旋转矩阵既是逆矩阵也是转置矩阵。这有助于我们理解转置矩阵在几何上的意义，可以简单地认为：空间和空间元素之间的相对旋转运动的几何描述是转置矩阵的几何意义。其实空间和空间元素之间的关系应是一种对偶关系，在后面的 5.14.2 节里也会从对偶空间的角度讨论转置矩阵的几何意义。

5.13 矩阵的等价、相似与合同关系

5.13.1　矩阵等价、相似及合同的关系对比

据说，整个线性代数里矩阵之间有三种最典型的关系：矩阵相似（similar）、矩阵等价（equivalent）和矩阵合同(congruent)。具体的定义如下：

（1）$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 等价⇔存在可逆矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$，使得 $\boldsymbol{B}=\boldsymbol{PAQ}$；

（2）$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似⇔ 存在可逆矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$；

（3）$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 合同⇔存在可逆矩阵 $\boldsymbol{C}$，使得 $\boldsymbol{B}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{AC}$。

注意这三种关系的联系和差别。据定义可以知道，这三种矩阵关系都是等价关系。其中等价关系是最弱的一个关系：两个矩阵相似，或两个矩阵合同，那这两个矩阵一定是等价的，但是反过来不成立。相似与合同矩阵之间不能够互相推导。但是如果两个实对称矩阵是相似的，那肯定是合同的；反之也成立。如果从整体上来看矩阵之间的三种关系，我们会想到用集合来表示三者之间的关系(见图 5-68)。集合的表示更加清晰地展现了矩阵等价、相似、合同三者之间的关系。

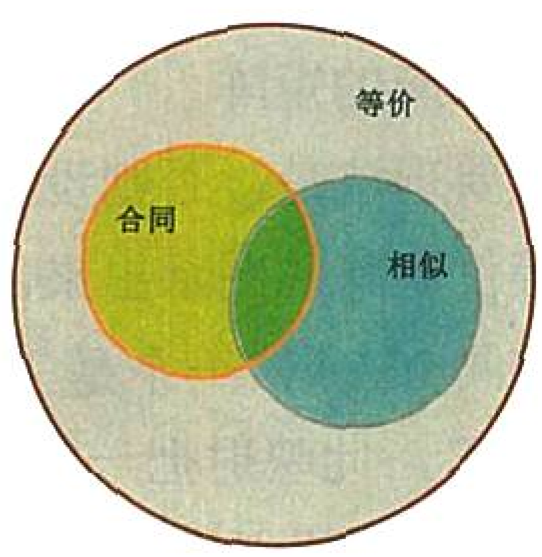


图 5-68 矩阵的等价、相似、合同的关系图

图中显示，相似的矩阵不一定合同，合同的矩阵不一定相似。但相似和合同有交集，就是

有既相似又合同的矩阵。若转换矩阵 $\boldsymbol{P}^{-1}=\boldsymbol{P}^{\mathrm{T}}$，这两定义变成相同，则 $\boldsymbol{A}$、$\boldsymbol{B}$ 两个矩阵既是相似的又是合同的，但具备这种性质的矩阵 $\boldsymbol{P}$，只有正交矩阵一类，因为这正是正交矩阵的定义。

接着，实对称矩阵一定存在正交转换矩阵 $\boldsymbol{P}$，满足相似和合同的定义式，因此，如果两个实对称矩阵是相似的，那肯定是合同的。

关于这三个矩阵的关系的几何意义已有结论，包括我们前面刚讨论过的相似矩阵的几何意义也一起合并在这里，就是：

两个有限维向量空间之间的同一个线性映射，其在这两个向量空间上的不同基下所对应的矩阵之间的关系就是等价关系。

一个有限维向量空间上的同一个线性变换（或称线性算子），其在不同基下所对应的矩阵之间的关系是相似关系。

一个有限维向量空间上的同一个双线性函数或内积，其在两个基下的度量矩阵是相合关系。

下面分别研究它们几何意义之间的区别和联系。

5.13.2 等价矩阵几何意义

等价矩阵的概念可以看成从同解（等价）线性方程组得来的。

我们知道，对一个线性方程组经过下列若干个变换：（1）**变换方程的顺序**；（2）**用一个非零的数乘以方程**；（3）**用一个数乘以方程加到另一个方程上**。方程组变为另一个方程组，这些变换都不影响它的解，变换前后的方程组都是同解方程组，也是等价方程组。

把方程组改写成矩阵的形式，上面的三种变换就是曾经讲过的初等行变换。因此我们可以得到这样的一个结论：

一个矩阵经过初等行变换后得到的任意一个矩阵都与原矩阵等价。

以上的结论大家都没有问题，但是，我们也知道另外一个结论就是：**一个矩阵经过初等列变换后得到的任意一个矩阵都与原矩阵等价。**

这个结论不能从同解方程组得到，因为我们解线性方程组只能用初等行变换而不能用初等列变换，如果对方程组的矩阵进行初等列变换将会把未知量搞乱（实质上是改变了坐标系），显然就得不到所给线性方程组的解了（因为所谓的解是在同一个坐标系下的数值）。

因此，对矩阵进行初等变换（包括初等行变换和初等列变换），是等价变换，是对方程组矩阵求解变换进行更大的推广。在这种情况下，只有矩阵的秩是不变的量。

看来不能从解线性方程组得到理解了，要理解等价矩阵还是需要从线性空间的角度得到彻底理解。前面说，等价矩阵是在两个有限维向量空间之间的同一个线性映射的表示，是线性映射在这两个向量空间上的不同基下所对应的矩阵。

理解起来蛮抽象的，但我们必须要理清楚，否则没法进一步理解相似和合同的来由。好了不要怕麻烦，下面我们就画出两个等价矩阵的图示出来，见图 5-69。

图中给出了两个一般性的线性空间 U、V，一个是三维的，一个是二维的。在从三维空间到二维线性空间定义了一个线性映射 $\boldsymbol{\Gamma}$，此映射把一个立方块变换成了一个平行四边形面。

此线性映射 $\boldsymbol{\Gamma}$ 在空间 U 的基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\}$ 和 V 的基 $\{\boldsymbol{y}_1,\boldsymbol{y}_2\}$ 下的矩阵为 $\boldsymbol{A}$，$\boldsymbol{\Gamma}$ 在 U 的另外一个基 $\{\boldsymbol{s}_1,\boldsymbol{s}_2,\boldsymbol{s}_3\}$ 和 V 的另外一个基 $\{\boldsymbol{t}_1,\boldsymbol{t}_2\}$ 下的矩阵为 $\boldsymbol{B}$。一个映射表示为两个矩阵只因基的不同。

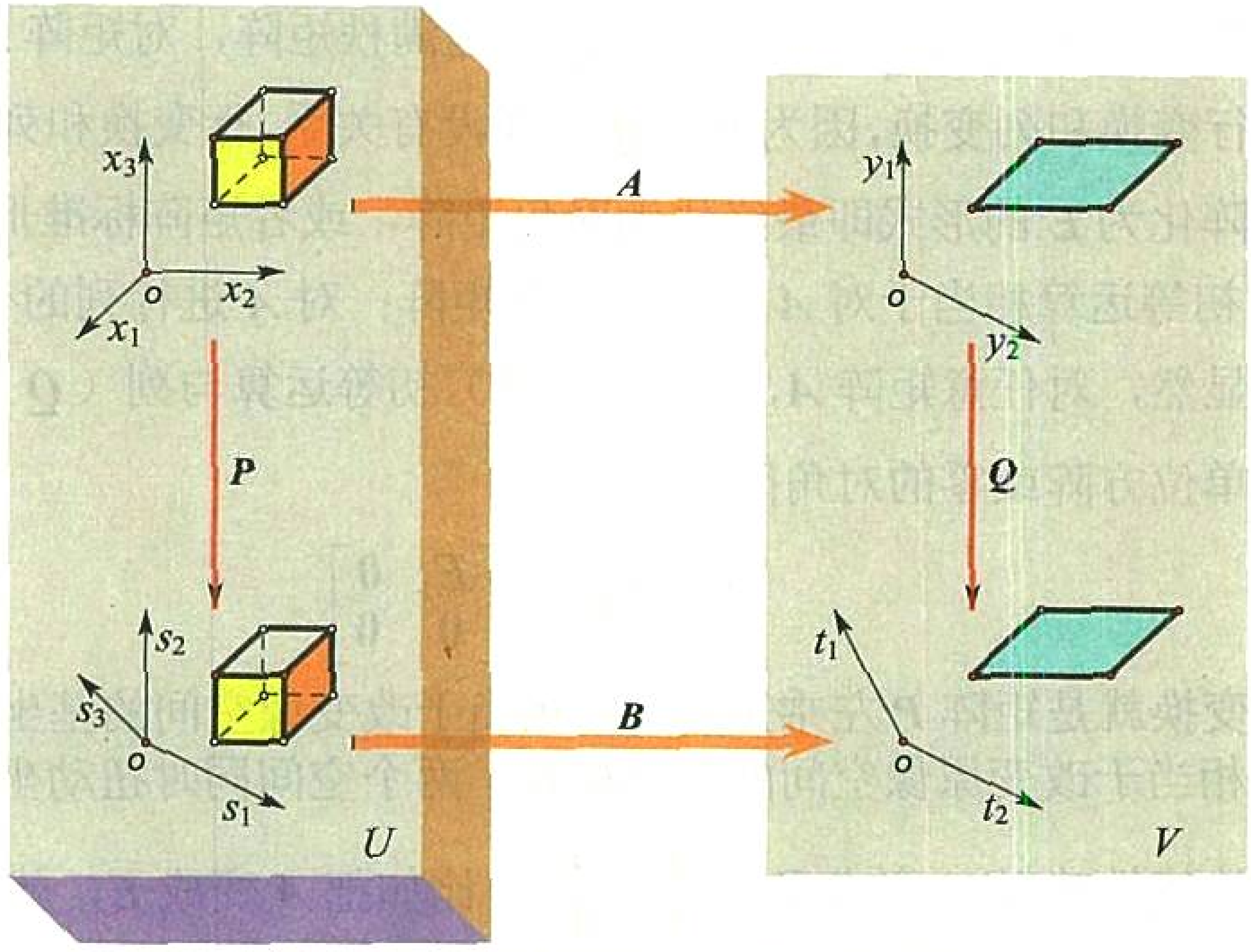


图 5-69 两个等价矩阵 A 和 B 的几何意义

另外，空间 U 下的基转换矩阵（或称为过渡矩阵）为 $\boldsymbol{P}$，它把基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\}$ 转换为 $\{\boldsymbol{s}_1,\boldsymbol{s}_2,\boldsymbol{s}_3\}$；空间 V 下的基转换矩阵为 $\boldsymbol{Q}$，它把基 $\{\boldsymbol{y}_1,\boldsymbol{y}_2\}$ 转换为 $\{\boldsymbol{t}_1,\boldsymbol{t}_2\}$。

还有，把映射前后的向量定义一下：设任一个向量 $\boldsymbol{\alpha}\in U$，向量 $\boldsymbol{\alpha}$ 在基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\}$ 和 $\{\boldsymbol{s}_1,\boldsymbol{s}_2,\boldsymbol{s}_3\}$ 下的坐标式为 $\boldsymbol{x}$ 和 $\boldsymbol{s}$；映射后的向量 $\boldsymbol{\Gamma}(\boldsymbol{\alpha})\in V$，向量 $\boldsymbol{\Gamma}(\boldsymbol{\alpha})$ 在基 $\{\boldsymbol{y}_1,\boldsymbol{y}_2\}$ 和 $\{\boldsymbol{t}_1,\boldsymbol{t}_2\}$ 下的坐标式为 $\boldsymbol{y}$ 和 $\boldsymbol{t}$。

好了，矩阵和向量们都到位了，我们开始研究它们之间的关系吧！

由图示和定义知道，各个空间内部的向量之间的变换关系为

$$\boldsymbol{x}=\boldsymbol{P}\boldsymbol{s},\ \boldsymbol{y}=\boldsymbol{Q}\boldsymbol{t}$$

由于 $\boldsymbol{A}$ 是从基 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\}$ 到基 $\{\boldsymbol{y}_1,\boldsymbol{y}_2\}$ 下的映射矩阵，因此有

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$$

把 $\boldsymbol{x}=\boldsymbol{P}\boldsymbol{s}$，$\boldsymbol{y}=\boldsymbol{Q}\boldsymbol{t}$ 代入上式，得到

$$\boldsymbol{Q}\boldsymbol{t}=\boldsymbol{A}\boldsymbol{P}\boldsymbol{s}$$

即

$$\boldsymbol{t}=(\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{P})\boldsymbol{s}$$

上式的意思就是从向量 $\boldsymbol{s}$ 变换到 $\boldsymbol{t}$ 的矩阵是 $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{P}$，又知道前面的矩阵定义是 $\boldsymbol{t}=\boldsymbol{B}\boldsymbol{s}$，对比两式，可说明线性映射 $\boldsymbol{\Gamma}$ 的两个表示矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 之间的关系为

$$\boldsymbol{B}=\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{P}$$

因为 $\boldsymbol{P}$、$\boldsymbol{Q}^{-1}$ 都是可逆的方阵，换个写法就和前面的等价矩阵的定义式 $\boldsymbol{B}=\boldsymbol{P}\boldsymbol{A}\boldsymbol{Q}$ 相同了。别忘了 $\boldsymbol{P}$、$\boldsymbol{Q}$ 都是基转换矩阵，看来同一个线性映射 $\boldsymbol{\Gamma}$ 的两个表示矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 之间的关系还是与各个基及其变换有关的。

至此等价矩阵的几何意义披露无遗了（看不懂上面推导的请回顾 4.2.4 节内容）。

在这里大家应该明白几个小问题：

（1）为什么矩阵 A、B 可以不是方阵？因为是不同空间之间的线性映射矩阵。

（2）为什么矩阵 P、Q 必须是方阵而且可逆？因为是同一空间里的基过渡矩阵。

（3）矩阵既可以看做向量的变换也可以看做是基的变换，两者的表达式中自变向量和因变向量的位置相反。

进一步地，我们试图理解以下常见的说法：

根据等价矩阵的定义 $B=PAQ$，其中 P 和 Q 都是满秩矩阵，对矩阵 A 作等价变换，相当于对矩阵 A 进行了行变换和列变换，因为 P 和 Q 之间没有关系，行变换和列变换各自独立进行，所以可以把任何矩阵化为 E 的形式即最终化为对角矩阵，或者矩阵标准形。

对 A 进行行的初等运算相当于对 A 左乘相应的矩阵；对 A 进行列的初等运算相当于对 A 右乘相应的矩阵。显然，对任意矩阵 A，经过行（P）初等运算与列（Q）初等运算都可以变为一个对角线上是单位方阵或零的对角阵：

$$B=PAQ=\begin{bmatrix} E_k & 0 \\ 0 & 0 \end{bmatrix}$$

现在看来，行变换就是矩阵 P 左乘矩阵 A，相当于改变像空间的基坐标系；列变换就是矩阵 Q 右乘矩阵 A，相当于改变原像空间的基坐标系。两个空间同时扭动坐标轴，扭来改去，总能找到一种改法，让线性映射 A 变成 E 或 $\begin{bmatrix} E_k & 0 \\ 0 & 0 \end{bmatrix}$。比如把 A 变成 E，就是说两边空间的向量坐标完全一样了。

任一矩阵都可通过一系列的初等变换化非对角线上的元素为零，从而成为对角阵。因此任一矩阵都等价于一个对角阵，其对角线上非零元素的个数正好是原矩阵的秩。所以说秩是等价变换的不变量。由于对等价变换的限制少，故适用的范围也广。然而除了秩不变外，矩阵的其他性质在变换以后就很难反映出来了。

5.13.3　相似与等价矩阵几何意义的对比

和我们讲的相似矩阵的几何解释比较起来，等价矩阵怎么看着跟相似矩阵的几何意义差不多呢？确实意思很接近。相似矩阵是同一个向量空间里两个坐标基下的同一个线性映射。相似变换就是特殊的等价变换。

当然也可以认为相似变换是两个空间的线性映射，我暂时同意这个观点，画一画框图看看：

把上面的等价变换的框图中的基过渡矩阵 Q 改成 P，结果满足相似矩阵的定义式 $B=P^{-1}AP$，成为如图 5-70 所示的结果。

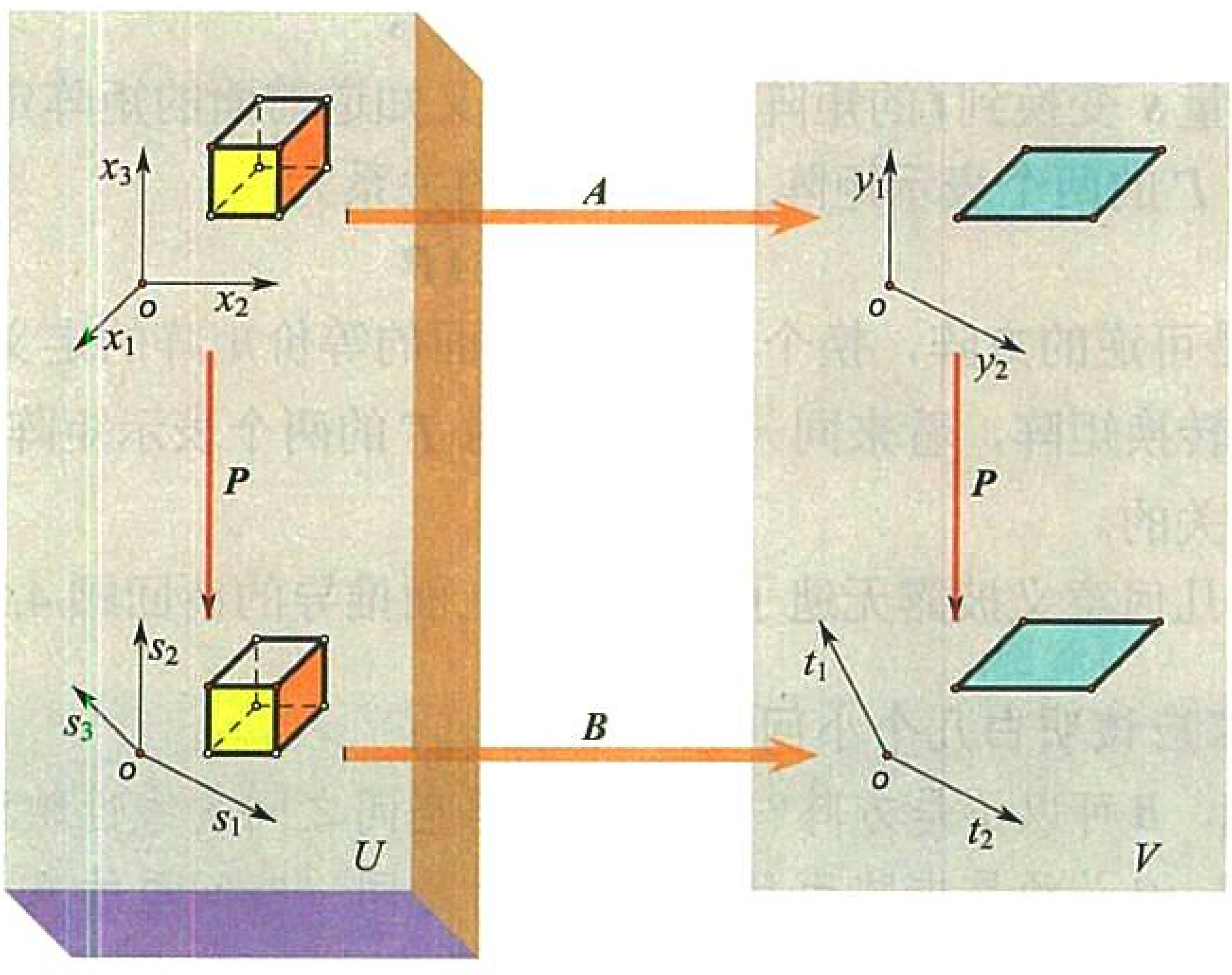


图 5-70 相似矩阵是特殊的等价矩阵

问题来了。

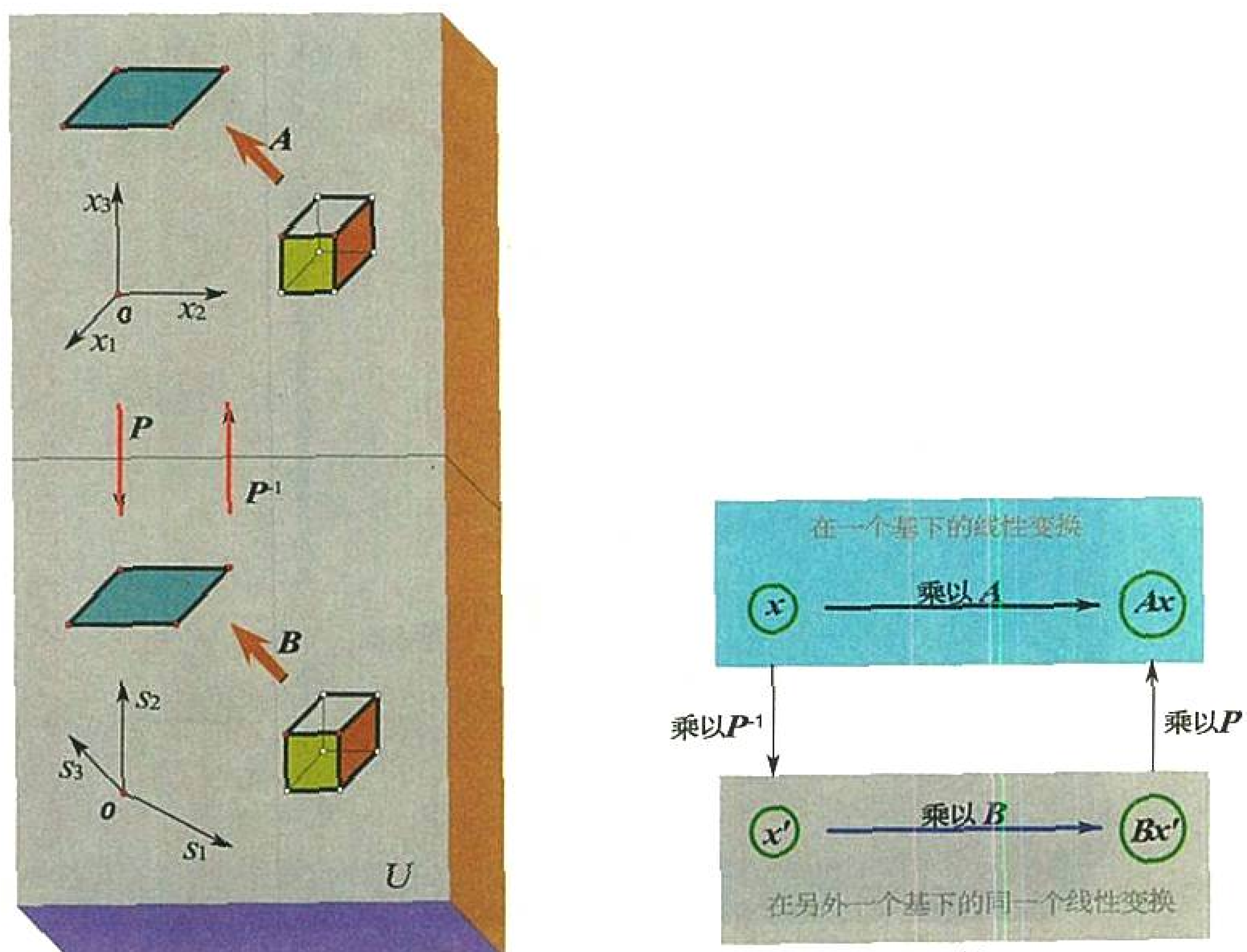
既然两个空间的基过渡矩阵都是方阵 $\boldsymbol{P}$，这说明两个空间的维数是相同的，而不应该是图中所画的那样。要么都是二维，要么都是三维的空间。

好的，我们引入同构的概念。

同构是线性代数中重要的概念。如果两个线性空间上的映射变换既是单射又是满射，就称这两个向量空间同构。两个向量空间同构，那么就有线性映射使这两个线性空间的向量（或点）一一对应，而且保持线性不变。这时往往将这两个向量空间看做同一个。对于向量空间，同构也是等价关系。

因此我们应该将两个向量空间合二为一。所以说，相似变换是同一个向量空间的变换，是两个基上的同一个线性变换。

顺着现在的框图画法，我们重新画一下相似矩阵的几何意义框图，见图 5-71（a）。试着和图 5-71（b）比较一下，有没有更多的启发呢？



(a)　(b)

图 5-71 相似矩阵的几何意义

5.13.4　合同与等价矩阵几何意义的对比

相似矩阵描述的是在不同参照系的同一个变换，动作规则是相同的。类似地，合同矩阵描述的是在不同参照系下的同一个内积的度量矩阵。前面我们在定义内积的推广式子时知道：

$$(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{y},\ \boldsymbol{S}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$$

其中方阵 $\boldsymbol{S}$ 就是度量矩阵，度量矩阵是由基向量所构成的过渡方阵 $\boldsymbol{P}$ 与其转置的乘积得到的。

好了，既然度量矩阵 $\boldsymbol{S}$ 是由基的过渡矩阵所决定的，那么每更换一次基坐标系就会有一个新的度量矩阵比如 $\boldsymbol{T}$ 出来。这些度量矩阵 $\boldsymbol{S}$、$\boldsymbol{T}$ 对应着同一个内积。有人会问，这些度量矩阵肯定有关系，这个关系是啥？这个问题好，太有关系了，这个关系就是合同关系：$\boldsymbol{S}$ 合同于 $\boldsymbol{T}$。

为了看出来合同关系也是等价关系，我们也沿袭等价关系框图的画法给出空间中两个合同矩阵的示意图，见图 5-72。

图中显示，在一个基为$\{x_1, x_2, x_3\}$的三维空间里，首先有两个向量a、b及其内积$a \cdot b$的值。如果这个初始的基取所谓的世界坐标系，那么向量a、b的内积值就是一个通用的国际标准值。这个值在对基坐标系合同变换时将保持数值不变。

不确定？参照图 5-72，简单地推导一下就可以确认了：

显然，图中显示了三个基坐标系互相转换的关系，而合同矩阵作为内积度量矩阵，也是和基联系在一起的。为了简化推导，我们这里把基向量都看成行向量。因为度量矩阵的定义是矩阵及其转置的积，当然列向量变成行向量，度量矩阵仍然不变。

在基$\{x_1, x_2, x_3\}$到基$\{y_1, y_2, y_3\}$的转换时，转换过渡矩阵定义为P_1，那么内积度量矩阵定义为A，且有$A = P_1P_1^{\mathrm{T}} = P_1^{\mathrm{T}}P_1$。同时在基$\{x_1, x_2, x_3\}$到另一个基$\{z_1, z_2, z_3\}$的转换时，转换过渡矩阵定义为$P_2$，那么内积度量矩阵定义为$B$，且有$B = P_2P_2^{\mathrm{T}} = P_2^{\mathrm{T}}P_2$。

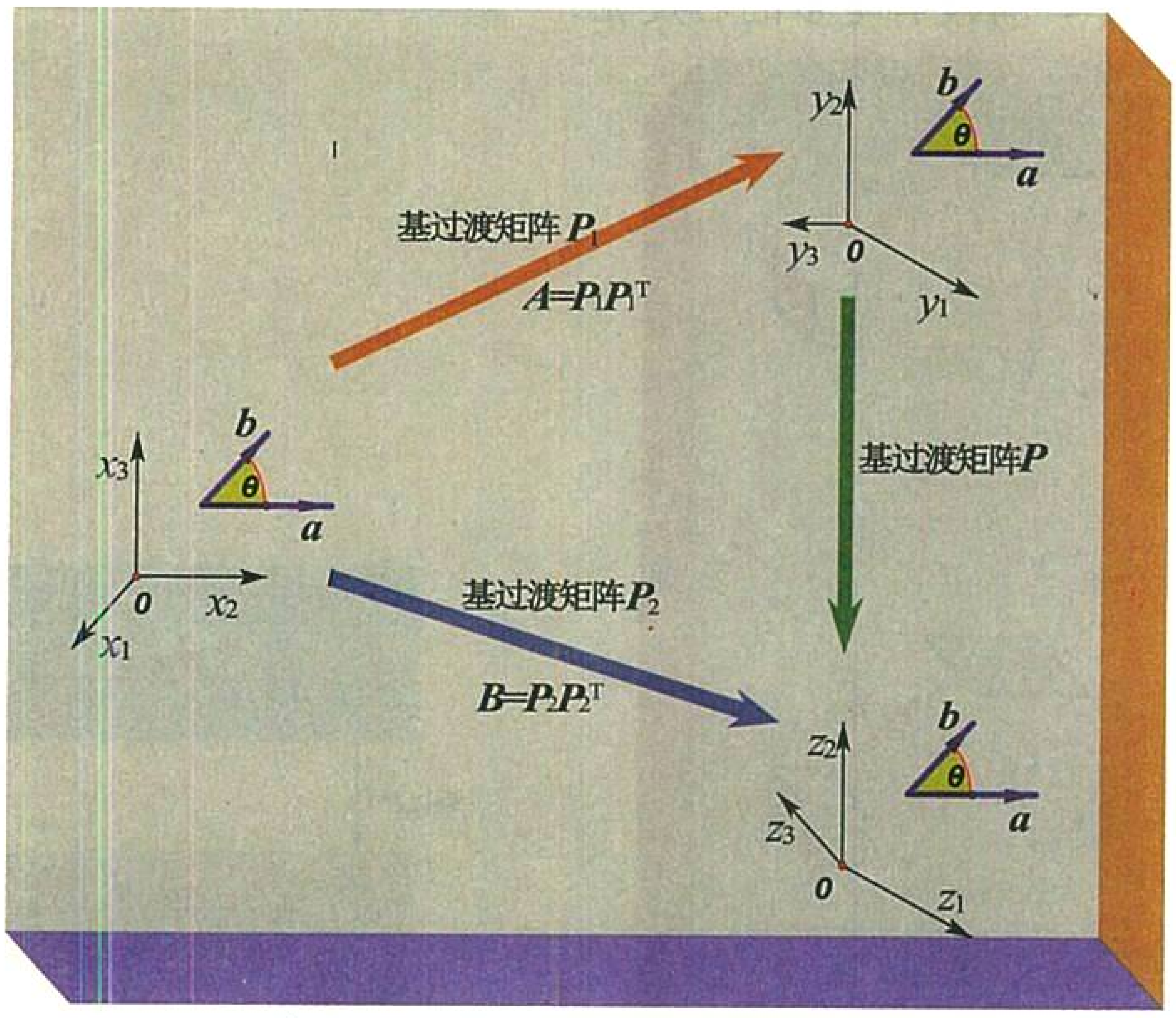


图 5-72 合同矩阵的几何意义

如果基$\{y_1, y_2, y_3\}$到基$\{z_1, z_2, z_3\}$的过渡矩阵为P，则有$P_2 = P_1P^{\mathrm{T}}$。

好，至此我们把这三个基之间的互换关系都定义好了，下面推导A和B之间的关系了。

把$P_2 = P_1P^{\mathrm{T}}$代入$B = P_2^{\mathrm{T}}P$，有

$$B = P_2^{\mathrm{T}}P_2 = (P_1P^{\mathrm{T}})^{\mathrm{T}}P_1P^{\mathrm{T}} = PP_1P_1^{\mathrm{T}}P^{\mathrm{T}} = P(P_1P_1^{\mathrm{T}})P^{\mathrm{T}} = PAP^{\mathrm{T}}$$

由上式看来，度量方阵A、B确实是合同矩阵。

这里咱们就简单讨论下，到二次型的时候还要讨论合同变换。

5.14 其他各类矩阵的几何意义

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。如果我们对矩阵的作用进行分类，就会发现一些特殊的矩阵，比如前面讨论过的旋转矩阵，只对向量进行旋转而没有做伸缩操作。单位矩阵是对一个向量不做任何操作的矩阵，组成矩阵的向量是标准的单位向量；单位矩阵也可以看做一个对向量作 0 角度旋转的旋转矩阵。本节就这些具有一定代

数和几何特质的矩阵做一些几何意义上的集中讨论及总结。

5.14.1 逆矩阵的几何意义

方阵 $\boldsymbol{A}$ 的逆记做 $\boldsymbol{A}^{-1}$，也是一个矩阵。当 $\boldsymbol{A}$ 与 $\boldsymbol{A}^{-1}$ 相乘时，结果是单位矩阵，表示为公式形式：

$$\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{E}$$

并非所有的矩阵都有逆。一个明显的例子是，若矩阵的某一行或列上的元素都为 0，用任何矩阵乘以该矩阵，结果都是一个零矩阵。如果一个矩阵有逆矩阵，那么称它为**可逆**的或**非奇异**的。如果一个矩阵没有逆矩阵，则称它为不可逆的或奇异矩阵。奇异矩阵的行列式为 0，非奇异矩阵的行列式不为 0，所以检测行列式的值是判断矩阵是否可逆的有效方法。此外，对于任意可逆矩阵 $\boldsymbol{A}$，当且仅当 $\boldsymbol{v}=\boldsymbol{0}$ 时，$\boldsymbol{A}\boldsymbol{v}=\boldsymbol{0}$。

在一个有乘法的代数系统中我们自然会考虑是否可以定义除法，如果能够定义除法，那么运算将会更完善，而且只有在具有除法的系统中方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 才有解。由于矩阵有乘法，那么能否定义除法呢?要定义除法必须存在具有数“1”相同性质的元，而且每个元必须存在具有“倒数”性质的元。这也许就是定义“逆矩阵”的原因。那么有“逆矩阵”的概念，为什么矩阵又没有除法呢?这有如下原因：

(1) 不是每一个非零矩阵都有逆矩阵。

(2) 矩阵的乘法不适合交换律。即使一个矩阵有逆矩阵，我们仍然不能定义除法。例如 $\boldsymbol{A}/\boldsymbol{B}$，因为一般情况下 $\boldsymbol{A}\boldsymbol{B}^{-1}\neq\boldsymbol{B}^{-1}\boldsymbol{A}$，是 $\boldsymbol{A}/\boldsymbol{B}=\boldsymbol{A}\boldsymbol{B}^{-1}$ 还是 $\boldsymbol{A}/\boldsymbol{B}=\boldsymbol{B}^{-1}\boldsymbol{A}$ 呢？ $\boldsymbol{A}\boldsymbol{B}^{-1}$ 和 $\boldsymbol{B}^{-1}\boldsymbol{A}$ 哪一个为 $\boldsymbol{A}/\boldsymbol{B}$ 的结果呢？我们无法定义。

单纯从代数运算的角度理解矩阵的逆对于学生有一定困难。因此，下面从矩阵表示的几何变换的角度出发，利用矩阵的几何意义引导学生直观理解矩阵逆的含义。

给出一个 n 阶矩阵，画出列向量构成的一个平行多面体，如果这个多面体能够在 n 维向量空间中围出一个 n 维的空间，那么这个矩阵具有逆阵。

换句话讲，这个可逆的 n 方阵不会把一个 m 维图形变换后降低成 m-1 维或更低维的图形。没有丢失原几何图形的信息，还可以逆向地变回来。如果一个矩阵把三维的立体图形变成了二维的平面图形，肯定有信息丢失，把原图形玩残了；如果变换成了一维的图形，就成了一条线了，打死都恢复不了三维的原图。那么这个矩阵必然是不可逆的。

因此，当我们一听到说某个矩阵是可逆的，心里就说，哦，这是个不会把别人玩残的操作，实在不行还可以把别人变回来。

矩阵的逆在几何上非常有用，因为它使得我们可以进行“反向”或“相反”变换——能“撤销”原变换的变换。就像孙猴子七十二般变化练成以后，不论变成什么都可以再变回来，这说明猴子使用的变换矩阵是可逆矩阵。如果向量 $\boldsymbol{a}$ 用矩阵 $\boldsymbol{A}$ 进行变换得到 $\boldsymbol{A}\boldsymbol{a}$，接着用其逆阵 $\boldsymbol{A}^{-1}$ 进行变换，将会得到原向量。这很容易通过代数方法验证：

$$\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{a})=(\boldsymbol{A}^{-1}\boldsymbol{A})\boldsymbol{a}=\boldsymbol{E}\boldsymbol{a}=\boldsymbol{a}$$

例如，把一个图形沿某一方向压缩一半后，再沿相反方向伸长一倍就恢复了原样；把一个

图形顺时针旋转 90°，再逆时针旋转 90° 也就恢复了原样。所以，压缩矩阵$\begin{bmatrix}1 & 0\\ 0 & 0.5\end{bmatrix}$的逆矩阵是伸长矩阵$\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$，顺时针旋转 90° 矩阵$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$的逆矩阵是逆时针旋转 90° 矩阵$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$，二者作用于一个几何图形就等于恒等变换。对于一个图形连续两次作同样的反射变换就变回原来的图形，因此，反射变换的逆变换是它自身。

对于一个几何图形先压缩（用矩阵 $\boldsymbol{P}$ 表示）再旋转（用矩阵 $\boldsymbol{R}$ 表示）后，要变回原样应该先逆向旋转（$\boldsymbol{R}^{-1}$）回来再拉长（$\boldsymbol{P}^{-1}$），即

$$(\boldsymbol{PR})^{-1} = \boldsymbol{R}^{-1}\boldsymbol{P}^{-1}$$

投影变换把一个正方形变成一条线段，这条线段再无法变回成一个正方形，所以，投影变换没有逆变换。也就是说，有些矩阵没有逆矩阵。

求二阶方阵逆的技巧

对于一个二阶方阵 $\boldsymbol{A}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$，其逆阵的公式为 $\boldsymbol{A}^{-1}=\dfrac{1}{|\boldsymbol{A}|}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}$，记忆口诀：

主对调，副变号（主对角线元素相互调换，副对角线元素改变符号），

行列式除搞定了（最后除以原矩阵的行列式）。

5.14.2 转置矩阵的几何意义

美国数学家 C. Strang（他较早写过著名教材《线性代数及其应用》一书，见后附参考文献）说过，转置矩阵就像烙馅饼一样，把它翻个个而已。我们翻烙饼的目的就是要得到饼的另一面。类似地，矩阵的一面是行，另一面是列，如果要从列得到行照样要翻转一下。烙饼的两个面看起来是共生的，有正面就有反面；矩阵的列与行也是一样，多个列排起来自然有多个行出现。转置矩阵的实质就是矩阵的行和列的关系或者说是行向量和列向量的关系。

1. 转置的一般性理解

习惯上，我们是处理矩阵的列，把一列看成一个向量，行是坐标值；如果要把矩阵的行处理成向量并保持它的意义的话就要翻转一下，这样原来的看成的列向量转置成行向量，其实际表示的物理意义没有改变。但要注意，要么一直保持列向量的观点要么一直保持行向量的观点（因为行向量和列向量是属于不同的线性空间的元素）。

实际上单个的向量也是一样，比如两个列向量 $\boldsymbol{a}=\begin{pmatrix}a_1\\ a_2\\ a_3\end{pmatrix}$，$\boldsymbol{b}=\begin{pmatrix}b_1\\ b_2\\ b_3\end{pmatrix}$，如果求它们的内积，按照我们行乘以列的计算规矩，只能把其中一个转置后再计算：

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=\begin{pmatrix}a_1 & a_2 & a_3\end{pmatrix}\begin{pmatrix}b_1\\ b_2\\ b_3\end{pmatrix}=$$

两个矩阵也是一样的。本来大家都是把矩阵的列看成向量，如果也要求两个矩阵的每个列向量内积，只能把其中一个矩阵转置后相乘，内积的度量矩阵 $\boldsymbol{S}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ 就是这样来的。

看起来转置矩阵和向量内积有着本质的渊源，这个事儿稍后再聊。另外，转置矩阵也是线

性变换，这个是显然的，转置后也是矩阵嘛，是矩阵就是线性变换呀。那么矩阵转置前后的线性变换有没有关系？当然有：

如果转置前的矩阵是把一个列向量变换成另一个列向量，则转置矩阵就是把一个行向量变换成另一个行向量。例如：

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}=\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix},\quad \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix}\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}=\begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix}$$

第一个等式是矩阵把列向量 $\boldsymbol{b}$ 变换成了列向量 $\boldsymbol{c}$ ，第二个等式是矩阵把行向量 $\boldsymbol{b}$ 变换成了行向量 $\boldsymbol{c}$ 。其中，两个三阶的变换方阵就是转置关系了。

其实忽略行、列的区别，这两个运算等式的数学运算实质是等价的(你可以具体计算看看)。因此我们可以这样认为，为了区别行和列向量，同时要保持矩阵与向量的算术计算规则，就有了矩阵的转置变换了。

回忆一下，我们推导基过渡矩阵的时候分别把空间中的向量看成行向量和列向量后得到的两个基过渡矩阵就是互相的转置。当时有这样的说明：

$\boldsymbol{P}$ 和 $\boldsymbol{P}^{\mathrm{T}}$ 互为转置矩阵，这取决于你对基向量 α_i 和 β_i 看做行向量还是列向量。

上面对转置的讲解感觉有点肤浅哈。转置矩阵更多的时候是与“对偶”的数学概念相联系的，比如行向量和列向量是对偶的，行空间和列空间也是对偶的等。故我们不妨先理解下点线对偶的几何意义，以此帮助理解转置矩阵对偶性质的几何意义（其实对偶也与行、列有关）。

2. 对偶原理的几何意义

在平面几何中，我们研究图形的性质，其中最基本的一种性质就是结合性。比如说，点与直线间有一种简单的位置关系，我们说“一点在一直线上”，可是这件事实也可以换成另外的一种说法“一直线通过一点”。在这里“点”与“直线”两个名词互相交换了一下位置，像这样的两个命题在几何上就称之为“自行对偶”的命题，“点”与“直线”称为对偶元素。自行对偶命题的几何图形可以是相同的。

还有一个对偶的命题例子：“两点在一直线上”和“两直线交于一点”，这是两种不同的位置关系，但只要把“点”换成“直线”，“直线”换成“点”，再把关系词改动一下，就可以从前面的一种关系得出后面的一种关系。像这样的命题，几何学中就称之为“互为对偶”命题。这个互为对偶命题的几何图形是不相同的。

上述的对偶命题看不出来太神奇的地方。其实，在射影空间里，如果一个命题成立，那么它的对偶命题也成立，这就是对偶原理。这个对偶原理就比较神奇了——数学定理之间居然有某种对称性。比如帕斯卡定理与布利安桑定理是一对对偶定理；笛沙格定理与其逆定理也是一对对偶定理。我们具体看看帕斯卡定理与布利安桑定理的对偶性。

帕斯卡定理：内接于二次曲线的六边形的三对对边的交点共线。

布利安桑定理：外切于二次曲线的六边形的相对顶点的三条对角线共点。

这两个定理的图形如图 5-73 所示，对偶元素是：内接六边形的边对偶于外切六边形的顶

点；三个交点对偶于三个对角线；交点共同的一根直线 JKL 对偶于对角线共同的一个交点 P。

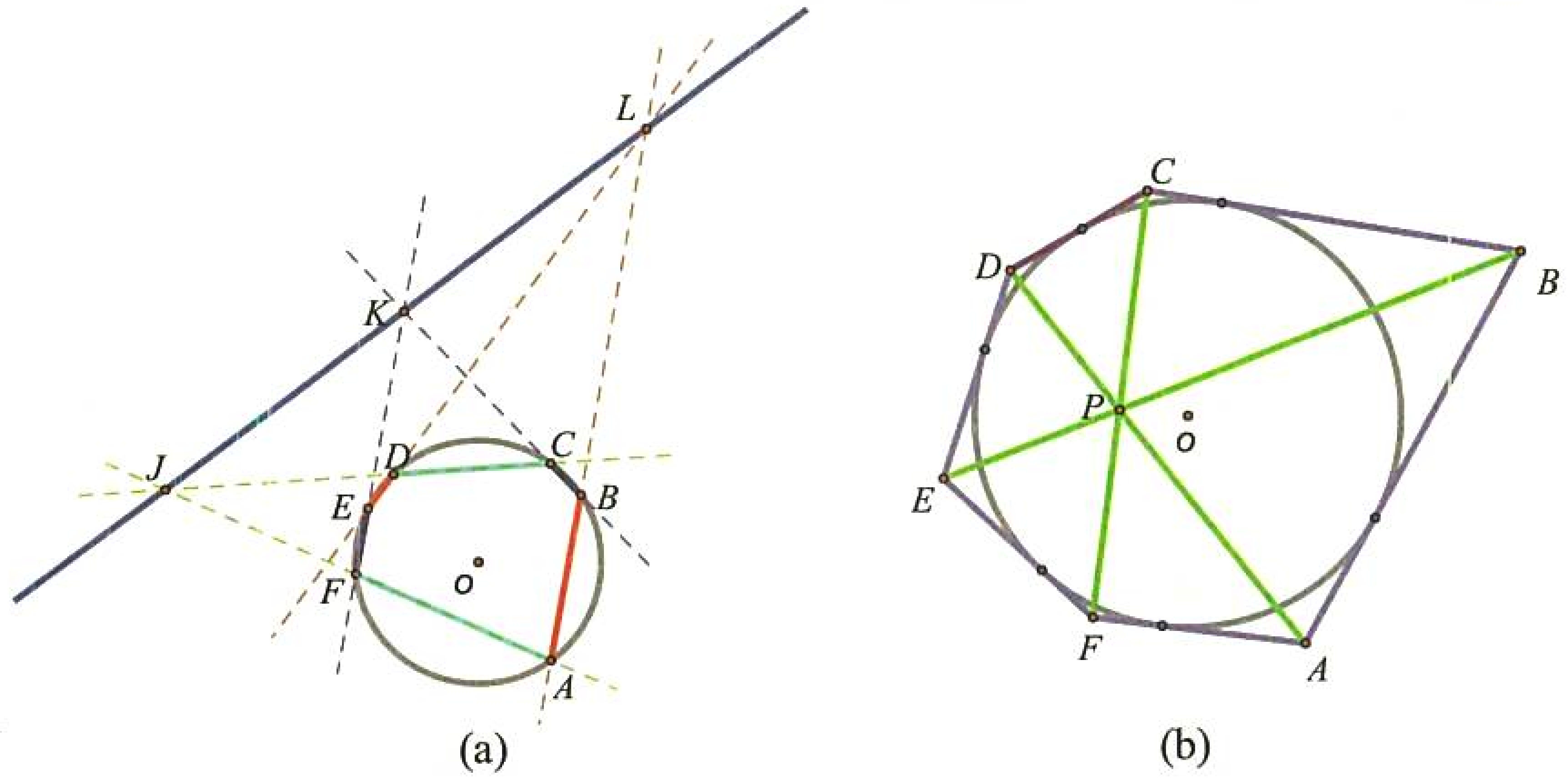


图 5-73　帕斯卡定理与布利安桑定理是一对对偶定理

帕斯卡定理与布利安桑定理虽然是一对对偶定理，但它们的发现并不是同时代的，竟然相差了一百多年。帕斯卡定理发表于 1640 年，而它的对偶定理在 1806 年才由布利安桑利用配极关系发现。当对偶定理发现后，数学家就可以利用对偶定理诱导出新的几何定理，如利用笛沙格定理诱导出其对偶定理。

如何理解对偶性的现实意义呢？

让我们先来考虑一个简单的问题。当我们问路的时候，别人通常会知在第几个路口向左（右）拐（走多少米），这是一种很常见的表达方式，事实上还有其他的表达方式。当为地平面取笛卡儿坐标系的时候，我们可以将人和建筑物（包括目的地）都给定坐标，然后以坐标的方式来描述，这样能达到同样的效果。当然，其理解性和使用性都不如前者。从这个问题，我们可以看出，实质上到达目的地有两种途径，一种是按方向和距离，另一种是按坐标。方向加距离是线段（向量），坐标是点，点线对偶，这也在客观上证明了两者的作用一致性。

为从代数式上揭示对偶原理，几何学中引入了所谓的线和点的齐次坐标的概念。

在欧氏几何里几何图形被认为是点的轨迹，是把点作为图形的基本元素。而在射影几何里认为平面图形也是直线（切线）的包络。点 (u,v) 可以看做线集 $\{y=c(x-u)+v,\ c\text{任意值}\}$ 里无穷直线的交点。比如，我们不把一个椭圆曲线考虑成一个移动着的点的轨迹，而是把它看成一条移动着的线的轨迹，就像素描画里用铅笔作画所勾勒出的草图线条。因此，直线作为点的对偶元素也是一种基本元素，进而有了线坐标的概念。

比如一个齐次代数式:

$$ux+vy=0 \tag{5-18}$$

这里有四个变量，如果把 (u,v) 看做平面上的线坐标，把 (x,y) 看做点坐标，那么此齐次式就是一个点线结合的方程。

把 (x,y) 看做点坐标容易理解；把 (u,v) 看做线坐标的理由是一个定值的坐标 (u,v) 所确定的方程是一条直线，换句话说就是一个 (u,v) 值确定一条直线。

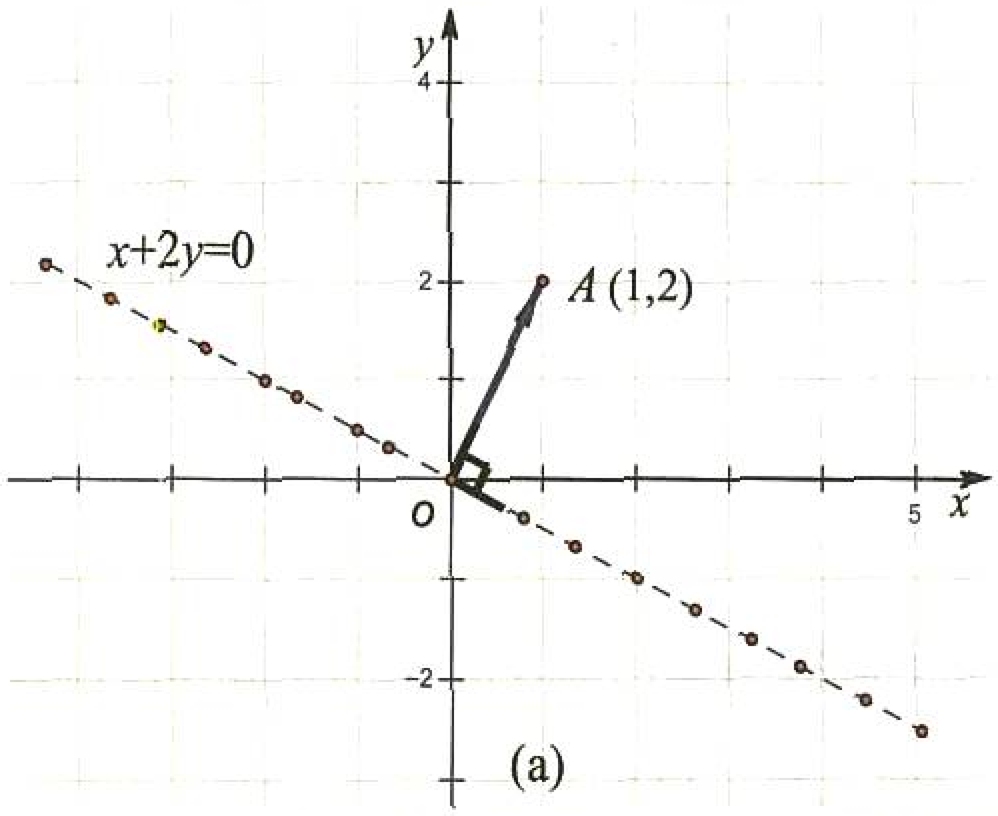
对于这个方程，不同情况下可以理解为点的方程或直线的方程，就是说点的方程和直线的方程具有同等的地位，或者说它们是完全对称的。于是，由代数式推导出的关于点的几何图形

的性质对于对偶的线的几何图形应该同样也具有；反之，关于线的几何图形的性质，对于对偶的点的图形也应成立。

实际上，**二维平面上，线坐标和点坐标之间几何图形的关系就是法向量和直线的关系。三维平面上，线坐标和点坐标之间几何图形的关系就是法向量和平面的关系。**

例如，二维平面上（见图 5-74（a）），设线坐标(u,v)的一个值为（1，2），那么就有无穷多点的轨迹$x+2y=0$（轨迹为一根直线）与之对应。向量（1，2）与直线$x+2y=0$垂直，是直线的法向量。一个法向量确定一根直线，因此可以说法向量是此直线的线坐标。

反之，任取一个点坐标(x,y)的一个值（3，-2），那么就有无穷多法向量的轨迹（也为一根直线）$3u-2v=0$与之对应（见图 5-74（b））。点（3，-2）与原点的连线与向量直线$3u-2v=0$也垂直。一个点确定一根向量直线，因此可以说此点是此直线的点坐标。



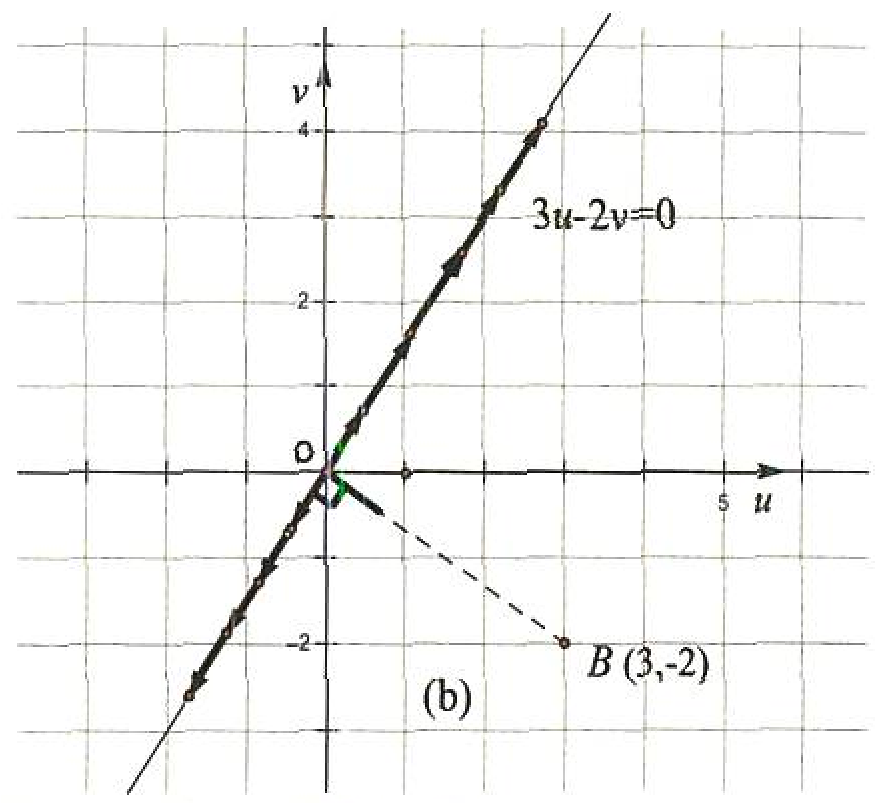


图 5-74　法向量 oA 与点的直线图形及点坐标 B 与向量的直线图形

进一步地，法向量（线）和点如果都处理成一样的向量的形式，那么前面的齐次式就是普通的两个向量内积的关系式，是内积为 0 的关系式，是正交的向量之间的关系。因而，具有正交关系的向量之间具有对偶的关系。

其实，不正交的向量之间也可以是对偶的，即使上述的齐次式等于一个数值而不只是 0。这时线坐标所对应的直线是不过原点的平行线。比如$x+2y=1$是与$x+2y=0$平行的且上移一个单位的直线。相应地，点坐标也有类似的几何图形。

上面的讨论其实是说，一个内积定义可以构筑一对相对偶的几何元素的集合（两个几何元素可以用向量表示），内积为零则构筑一对对偶的向量空间。

对偶原理应用广泛，据说在布尔代数、数理逻辑和偏序集理论、范畴论中都有应用。另外，在泛函分析中为研究一个函数空间的结构往往转而研究其对偶空间或共轭空间，在博弈论中研究对偶策略，在线性规划中考察对偶规划，在积分变换中研究互逆变换等，都是把一对数学结构按对偶化的方法联系起来，以便更好地解决问题。

3. 对偶空间与转置矩阵

对偶原理所揭示的对偶性与矩阵的转置有什么关系呢？

我们已知道，完全采用向量的观点，对偶的齐次代数式是两个向量内积的定义式，把线坐标记为行向量(u,v)，那么点坐标应该记为列向量$\begin{pmatrix}x\\y\end{pmatrix}$，这样才有等价的代数式：

$$(u,v)\begin{pmatrix}x\\y\end{pmatrix}=ux+vy=k \tag{5-19}$$

因此，我们可以说，由内积运算联系起来的行向量和列向量是天然对偶的。

如果我们认为，行向量属于一个线性空间，列向量属于另一个线性空间，则可以通过一个线性映射来把这两个空间构建成一对对偶空间。在基坐标下，线性映射可以用矩阵表示，矩阵其实表示了多重的内积运算，因此线性映射——矩阵把两个向量空间构建成一对对偶空间。

实际上，一个内积运算式即可构建一个对偶空间，只是这个对偶不是一一对应的，一般是一个向量对偶一个向量集合，多个内积运算式联合起来即可实现对偶向量一一对应。矩阵是多个内积运算的联合体，一个可逆方阵即可建立一个一一对偶的向量空间——行空间和列空间。

以上是矩阵对偶性的表述。那么矩阵转置和这种对偶性是什么呢？

简单地说，矩阵的转置就是对偶空间上的反向映射（或变换）。

稍仔细地说，有两个相对偶的线性空间 M 和 N。M 空间上有一个线性变换——矩阵 $\boldsymbol{A}$，它把 M 空间上的一个向量 $\boldsymbol{a}$ 变换成向量 $\boldsymbol{b}$；如果 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的对偶向量分别是 $\boldsymbol{a}^*$ 和 $\boldsymbol{b}^*$，那么对应的转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 就把对偶空间 N 上的 $\boldsymbol{b}^*$ 变换成 $\boldsymbol{a}^*$，即如下公式：

$$\begin{cases}\boldsymbol{A}\boldsymbol{a}=\boldsymbol{b}\\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^*=\boldsymbol{a}^*\end{cases} \tag{5-20}$$

千言万语，不如小例一两枚：

平面向量空间里，有一个向量 $\boldsymbol{a}=\begin{pmatrix}1\\1\end{pmatrix}$，在一个矩阵 $\boldsymbol{A}=\begin{bmatrix}1&3\\2&4\end{bmatrix}$ 的作用下得到向量：

$$\boldsymbol{b}=\boldsymbol{A}\boldsymbol{a}=\begin{bmatrix}1&3\\2&4\end{bmatrix}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}4\\6\end{pmatrix}$$

好了，在内积式 $(x,y)\begin{pmatrix}u\\v\end{pmatrix}=k$ 的关联下，向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的对偶极线分别为

$$(x,y)\begin{pmatrix}1\\1\end{pmatrix}=k,\quad (x,y)\begin{pmatrix}4\\6\end{pmatrix}=k$$

即

$$x+y=k,\quad 4x+6y=k$$

极线就是对偶点的集合，把这两根对偶极线用点坐标分别表示为 $(x,\ k-x)$，$(x,\ (k-4x)/6)$。

与把线坐标作为列向量的做法相对应：把点坐标看做行向量处理则有行向量：

$$\boldsymbol{a}^*=(x,\ k-x),\ \boldsymbol{b}^*=(x,\ (k-4x)/6)\text{。}$$

好。因为列向量是与矩阵左乘，那么行向量就与矩阵右乘。行向量 $\boldsymbol{b}^*$ 与矩阵 $\boldsymbol{A}$ 右乘得到：

$$\boldsymbol{b}^*\boldsymbol{A}=(x,\ (k-4x)/6)\begin{bmatrix}1&3\\2&4\end{bmatrix}=((k-x)/3,\ (2k+x)/3) \tag{5-21}$$

式(5-21)可以写成矩阵转置的形式，当然是在把 $\boldsymbol{a}^*$ 和 $\boldsymbol{b}^*$ 看作列向量的情况下：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^*=\begin{bmatrix}1&2\\3&4\end{bmatrix}\begin{pmatrix}x\\(k-4x)/6\end{pmatrix}=\begin{pmatrix}(k-x)/3\\(2k+x)/3\end{pmatrix} \tag{5-22}$$

经过计算我们将会神奇地发现，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^*$ 的坐标值落在了 $\boldsymbol{a}^*$ 的直线 $x+y=k$ 上。

图 5-75 是当 k=-20 时的图形。向量 $\boldsymbol{a}=\begin{pmatrix}1\\1\end{pmatrix}$ 在矩阵 $\boldsymbol{A}$ 的作用下变换为向量 $\boldsymbol{a}=\begin{pmatrix}4\\6\end{pmatrix}$，图中向量对应的极线 $x+y=-20$ 也变换为 $4x+6y=-20$。那么在对偶空间里，头端在极线上的向量也

在转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 作用下从 $4x+6y=-20$ 上变换到直线 $x+y=-20$ 上（延长线上）。

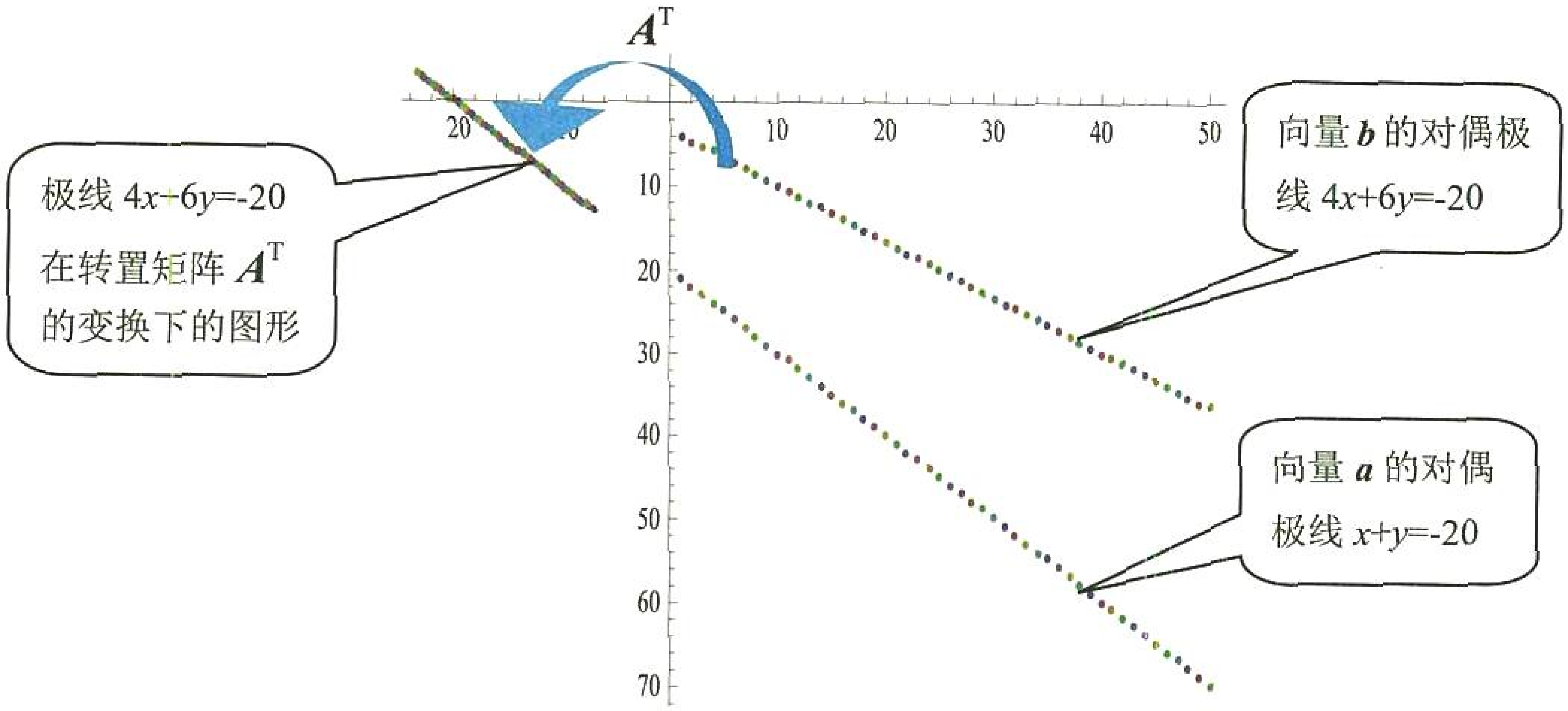


图 5-75　k=-20 时，向量的对偶极线在转置矩阵作用下进行的逆变换

我们换一下 k 的值，得到了如图 5-76 所示的变换图形。

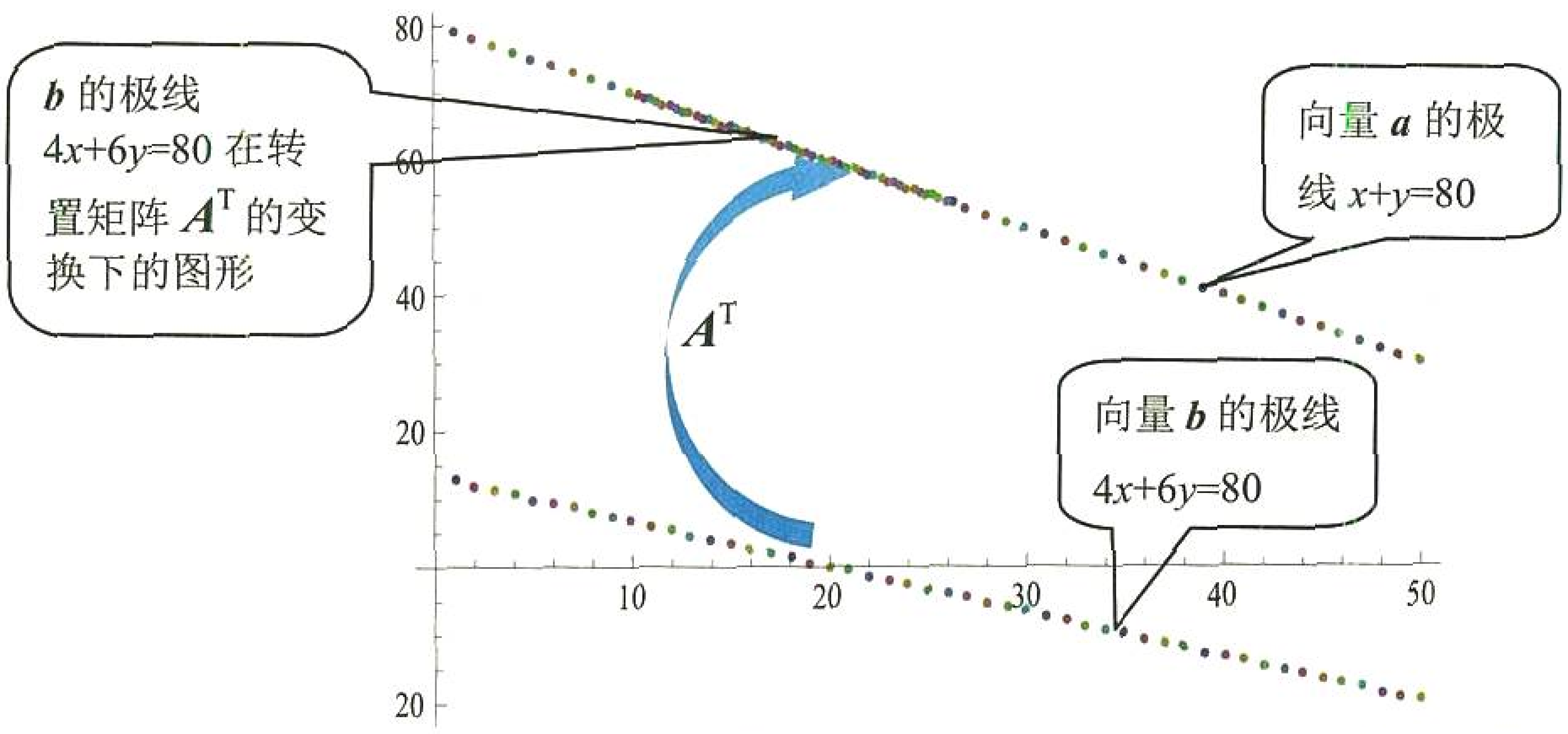


图 5-76 k=80 时，向量的对偶极线在转置矩阵作用下进行的逆变换

当然，对偶变换不仅仅局限为二维空间，在三维空间也存在对偶变换。在三维空间中，原空间的一点 $P(a,b,c)$ 通过三维对偶变换成为对偶空间中的一个平面 $z=ax+by+cz$。我们来一个三维空间里的例子，矩阵及其转置矩阵的变换图形之间的关系将会更清晰。

有一个三维向量 $\boldsymbol{a}=\begin{pmatrix}1\\1\\1\end{pmatrix}$，在一个矩阵 $\boldsymbol{A}=\begin{bmatrix}1&2&3\\4&5&6\\8&7&9\end{bmatrix}$ 的作用下得到向量：

$$\boldsymbol{b}=\boldsymbol{A}\boldsymbol{a}=\begin{bmatrix}1&2&3\\4&5&6\\8&7&9\end{bmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}6\\15\\24\end{pmatrix}$$

好了，在内积式 $(x,y,z)\begin{pmatrix}u\\v\\w\end{pmatrix}=k$ 的关联下，向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 的对偶极面分别为

$$(x,y,z)\begin{pmatrix}1\\1\\1\end{pmatrix}=k,\quad (x,y,z)\begin{pmatrix}6\\15\\24\end{pmatrix}=k$$

即

$$x+y+z=k,\ \ 6x+15y+24z=k$$

上式称之为极面，也就是点坐标的集合。把这两个对偶极面的点坐标用列向量形式表示并分别记为

$$\boldsymbol{a}^{*}=\begin{pmatrix} x \\ y \\ k-x-y \end{pmatrix},\ \boldsymbol{b}^{*}=\begin{pmatrix} x \\ y \\ (k-6x-15y)/24 \end{pmatrix}$$

变向量 $\boldsymbol{b}^{*}$ 在 $\boldsymbol{A}$ 的转置矩阵作用下有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^{*}=\begin{bmatrix} 1 & 4 & 8 \\ 2 & 5 & 7 \\ 3 & 6 & 9 \end{bmatrix}\begin{pmatrix} x \\ y \\ (k-6x-15y)/24 \end{pmatrix}=\begin{pmatrix} (k-3x-3y)/3 \\ (7k+6x+15y)/24 \\ (9k+18x+9y)/24 \end{pmatrix} \tag{5-23}$$

经过计算我们也会神奇地发现，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^{*}$ 的坐标值落在了 $\boldsymbol{a}^{*}$ 的平面 $x+y+z=k$ 上。换句话说，向量 $\boldsymbol{b}$ 的对偶极面 $\boldsymbol{b}^{*}$ 在转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 的变换下得到一个新平面，新平面落在原向量 $\boldsymbol{a}$ 的对偶平面上，见图 5-77。图 5-77（a）为当 k=-100 时的图形，图 5-77（b）为当 k=0 时的图形。当 k=0 时，对偶的两个平面都过原点，都是线性空间。

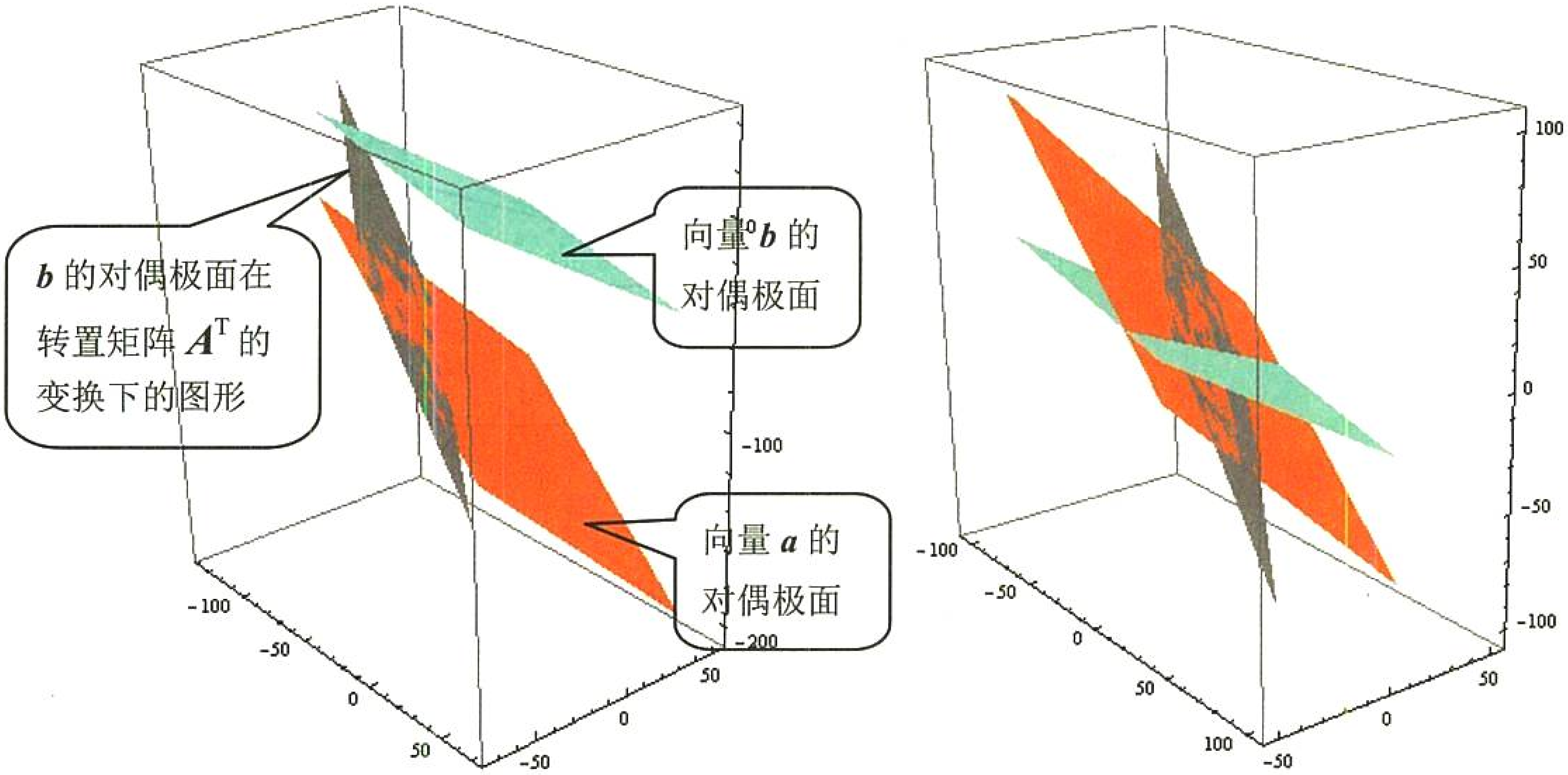


图 5-77 三维空间里，向量的对偶极面在转置矩阵作用下进行的逆变换

深入一点

在运筹学理论中，每一线性规划问题，都伴随另一线性规划问题。二者密切相关，互为对偶，其中一个称为原问题，另一个称为对偶问题，它们是对一个研究对象从不同的角度提出的两个极值问题。因此，在这两个线性规划模型中，约束条件的系数矩阵互为转置的关系。

当 $V'\times V$ 上有非退化的双线性型 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{y}$ 时，一个双线性型对应一个矩阵 $\boldsymbol{G}$，我们也称 V' 是 V 的对偶空间。线性空间 V 的每个线性变换 $\boldsymbol{T}$，诱导出对偶空间 V' 的一个线性变换 $\boldsymbol{T}^{*}$，称为 $\boldsymbol{T}$ 的伴随变换；如果变换用矩阵表述，或者说线性变换 $\boldsymbol{T}$ 在 V 的某组基下方阵表示为 $\boldsymbol{A}$，则伴随变换 $\boldsymbol{T}^{*}$ 在 V' 的对偶基下方阵表示为 $\boldsymbol{A}^{\mathrm{T}}$，即 $\boldsymbol{A}$ 的转置矩阵。所以此伴随变换也称为转置变换。

张贤科的《高等代数学》中引言说，……对偶有直观的几何意义（正交补），较容易懂。

4. 转置矩阵性质的讨论

好了，下面我们说说转置矩阵的性质。

转置矩阵和逆矩阵的关系好像比伴随矩阵的关系更铁。大家都知道，转置矩阵的性质和逆矩阵的性质在公式形式上几乎一致，比如 $(\boldsymbol{AB})^{\mathrm{T}}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$ 、$(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ 、$(\boldsymbol{A}^{\mathrm{T}})^{-1}=(\boldsymbol{A}^{-1})^{\mathrm{T}}$，另外矩阵逆的逆等于原矩阵，同样矩阵转置的转置也等于原矩阵：$(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{A}$。

为什么这么像呢？你只要看看逆矩阵计算公式的变形形式就明白了。

前面我们知道利用伴随矩阵求逆矩阵的公式 $\boldsymbol{A}^{-1}=\dfrac{\boldsymbol{A}^*}{|\boldsymbol{A}|}$，如果我们再定义一个余子式矩阵 $\boldsymbol{A}^{\oplus}$，它是由矩阵 $\boldsymbol{A}$ 的每个元素的代数余子式所构成的，余子式矩阵 $\boldsymbol{A}^{\oplus}$ 的每个元素是用 $\boldsymbol{A}$ 的元素 a_{ij} 的代数余子式替代自己而构成的矩阵。显然，我们有关系式 $\boldsymbol{A}^*=(\boldsymbol{A}^{\oplus})^{\mathrm{T}}$ 。

把此式代到逆矩阵的公式里，就有

$$\boldsymbol{A}^{-1}=\frac{(\boldsymbol{A}^{\oplus})^{\mathrm{T}}}{|\boldsymbol{A}|}=\left(\frac{\boldsymbol{A}^{\oplus}}{|\boldsymbol{A}|}\right)^{\mathrm{T}}$$

把 $\dfrac{\boldsymbol{A}^{\oplus}}{|\boldsymbol{A}|}$ 看成一个矩阵，这个公式所包含的代数意义就是**一个矩阵 $\boldsymbol{A}$ 的逆等于另一个矩阵的转置**。另一个矩阵是由原来的矩阵 $\boldsymbol{A}$ 计算后得到的。

嘿，这里用代数意义比用几何意义强多了。白管什么意义抓住真意就是好义。一切水落石出了，原来矩阵的逆就是另一矩阵的转置。怪不得哥俩像是一对双胞胎呢。

正因为转置矩阵和逆矩阵有些渊源，因此也就和正交矩阵攀上了亲戚。这个问题在下面的正交矩阵的几何意义继续探讨。

转置矩阵和伴随矩阵的相似性

类似地，因为有了 $\boldsymbol{A}^*=(\boldsymbol{A}^{\oplus})^{\mathrm{T}}$，它们除了和逆运算相似外，转置矩阵和伴随矩阵两者之间也有互换的关系：

$$(\boldsymbol{A}^*)^{\mathrm{T}}=(\boldsymbol{A}^{\mathrm{T}})^*$$

5.14.3 伴随矩阵的几何意义

$$\boldsymbol{A}^*\boldsymbol{A}=|\boldsymbol{A}|\boldsymbol{E}$$

式中，$\boldsymbol{A}$ 的伴随乘以 $\boldsymbol{A}$ 等于 $\boldsymbol{A}$ 的行列式乘以单位阵，伴随矩阵因此提出。只要见到伴随矩阵就用这个式子处理就行。伴随矩阵 $\boldsymbol{A}^*$ 总是像伴侣一样伴随着矩阵 $\boldsymbol{A}$，它就是解决 $\boldsymbol{A}$ 的行列式和 $\boldsymbol{A}$ 矩阵的关系而构造的。用它求矩阵的逆非常方便：

$$\boldsymbol{A}^{-1}=\frac{\boldsymbol{A}^*}{|\boldsymbol{A}|}$$

大伙儿仔细看看，逆矩阵和伴随矩阵之间只是相差一个常数，两个是非常类似的概念，也可以把伴随矩阵理解成总是“伴随”着逆矩阵。怪不得伴随矩阵和逆矩阵的性质也很接近呢，比如 $(\boldsymbol{AB})^*=\boldsymbol{B}^*\boldsymbol{A}^*$ 和 $(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$，还有 $(\boldsymbol{A}^*)^{-1}=(\boldsymbol{A}^{-1})^*$ 等。不过矩阵逆的逆等于原矩阵，

而矩阵伴随的伴随不等于原矩阵，而是原矩阵的倍数：$(\boldsymbol{A}^*)^*=|\boldsymbol{A}|^{n-1}\boldsymbol{A}$。

下面我们使用拉普拉斯展开式及其几何意义简洁地推导一下公式 $\boldsymbol{A}^*\boldsymbol{A}=|\boldsymbol{A}|\boldsymbol{E}$。

大伙儿知道，伴随矩阵 $\boldsymbol{A}^*$ 里面的元素构成是代数余子式，代数余子式的几何意义我们在行列式一章中讨论过。为便于讨论，再拿来回顾一下：

三阶行列式$|\boldsymbol{A}|$的拉普拉斯展开式为

$$|\boldsymbol{A}|=|\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}|=\begin{vmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{vmatrix}=c_1\begin{vmatrix}a_2&b_2\\a_3&b_3\end{vmatrix}+c_2\left(-\begin{vmatrix}a_1&b_1\\a_3&b_3\end{vmatrix}\right)+c_3\begin{vmatrix}a_1&b_1\\a_2&b_2\end{vmatrix}$$

其展开的几何意义如图 5-78 所示。

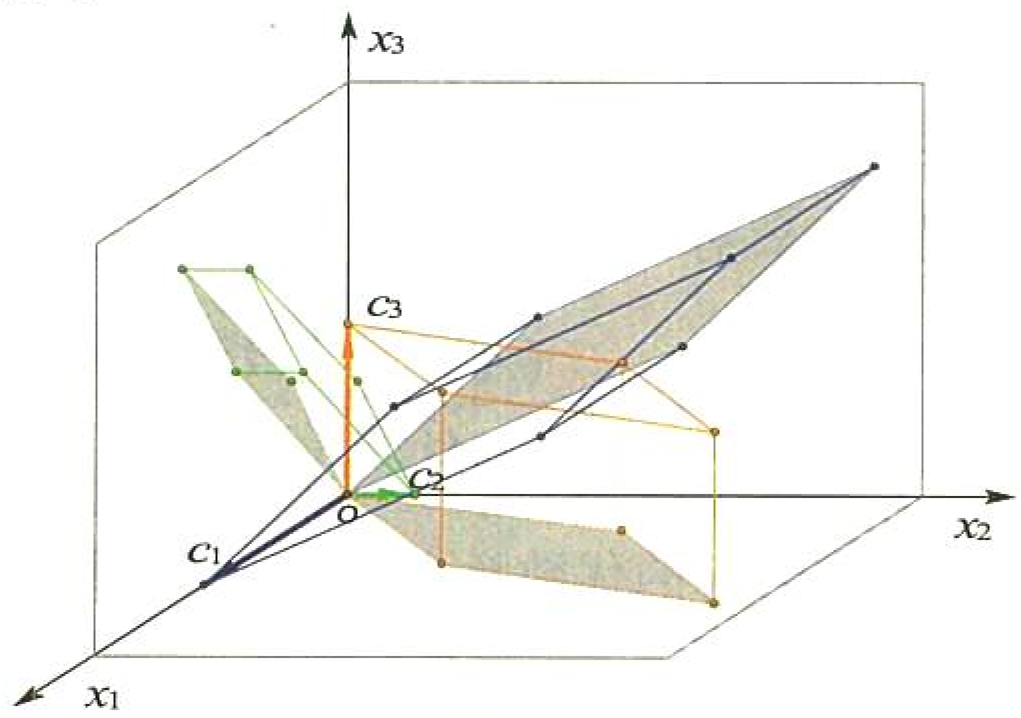


图 5-78 代数余子式是坐标平面上的有向投影

继续把等式改写如下：

$$|\boldsymbol{A}|=\left(\begin{vmatrix}a_2&b_2\\a_3&b_3\end{vmatrix},-\begin{vmatrix}a_1&b_1\\a_3&b_3\end{vmatrix},\begin{vmatrix}a_1&b_1\\a_2&b_2\end{vmatrix}\right)\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}$$

类似地，也可以按第一列和第二列向量分别展开如下：

$$|\boldsymbol{A}|=\left(\begin{vmatrix}b_2&c_2\\b_3&c_3\end{vmatrix},-\begin{vmatrix}b_1&c_1\\b_3&c_3\end{vmatrix},\begin{vmatrix}b_1&c_1\\b_2&c_2\end{vmatrix}\right)\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}$$

$$|\boldsymbol{A}|=\left(-\begin{vmatrix}a_2&c_2\\a_3&c_3\end{vmatrix},\begin{vmatrix}a_1&c_1\\a_3&c_3\end{vmatrix},-\begin{vmatrix}a_1&c_1\\a_2&c_2\end{vmatrix}\right)\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}$$

我们把上面三个式子组合成矩阵形式为

$$\begin{bmatrix}|\boldsymbol{A}|&0&0\\0&|\boldsymbol{A}|&0\\0&0&|\boldsymbol{A}|\end{bmatrix}=\begin{bmatrix}\begin{vmatrix}b_2&c_2\\b_3&c_3\end{vmatrix}&-\begin{vmatrix}b_1&c_1\\b_3&c_3\end{vmatrix}&\begin{vmatrix}b_1&c_1\\b_2&c_2\end{vmatrix}\\-\begin{vmatrix}a_2&c_2\\a_3&c_3\end{vmatrix}&\begin{vmatrix}a_1&c_1\\a_3&c_3\end{vmatrix}&-\begin{vmatrix}a_1&c_1\\a_2&c_2\end{vmatrix}\\\begin{vmatrix}a_2&b_2\\a_3&b_3\end{vmatrix}&-\begin{vmatrix}a_1&b_1\\a_3&b_3\end{vmatrix}&\begin{vmatrix}a_1&b_1\\a_2&b_2\end{vmatrix}\end{bmatrix}\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\a_3&b_3&c_3\end{bmatrix}=\boldsymbol{A}^*\boldsymbol{A}$$

上式稍作变形，正是 $\boldsymbol{A}^*\boldsymbol{A}=|\boldsymbol{A}|\boldsymbol{E}$。

经过上面的推导，我们就不再恐怖一些关于伴随矩阵的怪异的证明式子了，比如$|\boldsymbol{A}^*|=|\boldsymbol{A}|^{n-1}$等。另外，我们也大致知道了伴随矩阵 $\boldsymbol{A}^*$ 的几何构成：就是行列式$|\boldsymbol{A}|$所表示的平行六面体图形在 $\boldsymbol{A}$ 的每个元素方向上的投影面积。

5.14.4　正交矩阵的几何意义

1. 正交的概念

正交（orthogonal）是几何的垂直概念的推广。作为一个形容词，只有在一个确定的内积空间中才有意义。若内积空间中两向量的内积为 0，则称它们是正交的。

根据内积为 0 的定义来辨别向量是否正交虽然是笛卡儿直角坐标系的垂直概念的推广，但与我们的现实中的垂直的感受是不一样，即使它们的数学本质是一致的。下面我们举个平面上的例子，见图 5-79。

在笛卡儿直角坐标系上三个向量的坐标分别是 $\boldsymbol{a}=(2,1),\ \boldsymbol{b}=(1,1),\ \boldsymbol{c}=(-2,4)$。从图 5-79 上看到：

向量 $\boldsymbol{a}$ 和 $\boldsymbol{c}$ 是正交的，它们的内积 $\boldsymbol{a}\cdot\boldsymbol{c}=2\times(-2)+1\times 4$ 是 0，其夹角 $\theta=\arccos 0=\pi/2$；

而向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 不是正交的，它们的内积 $2\times 1+1\times 1=3$，其夹角 $\theta=\arccos\dfrac{3}{\sqrt{5}\sqrt{2}}\neq\pi/2$。

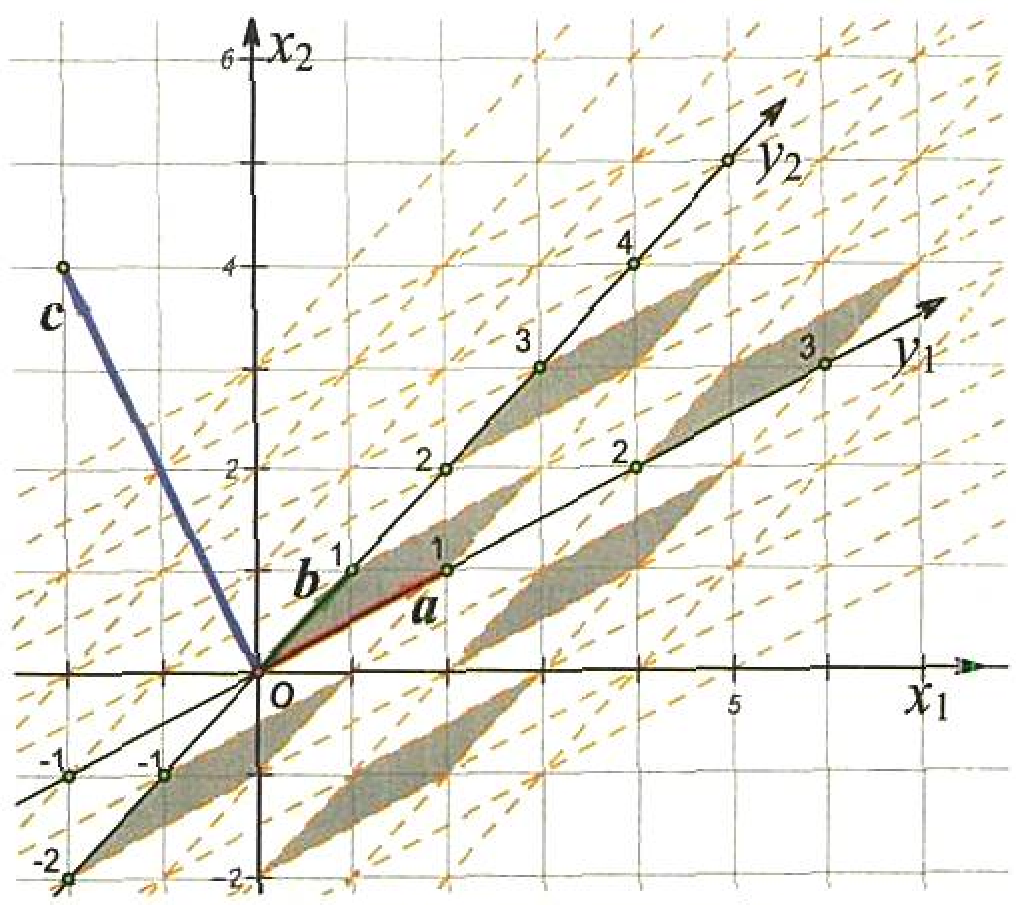


图 5-79　正交概念的几何意义

但是，如果我们以 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 为基建立新坐标系 $\{y_1oy_2\}$，原向量在新坐标系下的坐标分别为（读图）

$$\boldsymbol{a}=(1,0),\ \boldsymbol{b}=(0,1)$$

显然，向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 又变成了正交的，因为他们的内积 $\boldsymbol{a}\cdot\boldsymbol{b}=1\times 0+0\times 1=0$，其夹角 $\theta=\arccos 0=\pi/2$。

明明不是直角怎么就变成了直角，非正交的向量变成了正交的（同时正交的变成了非正交的）向量呢？这个计算让我们深思。

其实这个不难理解，两个向量的内积定义本来就包含了一个向量在另一个向量上的投影，如果投影为 0，当然内积也为 0（注意这个投影是与当前坐标轴平行的投影，笛卡儿系投影线构成了正方形网格，新基下投影线构成了平行四边形网格）。在一个坐标轴上投影为 0 说明投影向量对轴参数没有影响，或没有关联，这正是正交的物理意义。比如，在低于光速运动的情况下时间量正交于三维物质空间，因而可以构成四维的正交时空坐标系。

关于正交的概念认识似乎有点混乱，认为斜坐标系不是标准正交的，包括数学家 G. Strang 在他的名著《线性代数及其应用》中也说“如果把每个基都看成标准正交基将导致错误……”

（全面的说法咱后面说）。

既然这样，有人会说，为什么还要把一个基进行施密特正交化呢？

答：因为我们不自觉地把高维的笛卡儿坐标系作为世界坐标系，这个标准正交基的正交的概念是水平地面和重力的夹角，是四分之一正圆弧。当我们在这个笛卡儿基下另取一个基并标准正交化时，实际上是对两个基进行了一个正交变换：第一个标准正交基经过正交变换变换为另一个标准正交基。这个正交变换主要是基的一些旋转或镜像，在这个变换过程中所有的向量包括基向量的坐标表达式仍然是笛卡儿基下的坐标值，内积的度量矩阵没变，都是单位方阵 $\boldsymbol{E}$，正交的计算和判断标准没有改变。

前面说过，认识线性变换有两个方面：变换空间里的向量，空间基坐标系不变；或变换基坐标系而向量不变。

当我们采用第一个看法时，空间只存在一个基坐标系，参照系没变如何变换向量也没什么问题。当我们采用第二个看法时，空间存在两个以上的基坐标系，这个时候就会容易混乱参照系，不把后妈当妈了。原始的基坐标系绝对是标准正交基，亲妈也；以后取的基常常不当坐标系看，仍然当向量看——称之为基向量，这个基向量还不是参照系，参照系仍然是亲妈，作为后妈的基向量要照亲妈的正交的标准端正自己的行为——称之为施密特正交化。

再娶个后妈再施密特正交化……无论向量们的老爸娶多少后妈，正交的衡量标准仍然是第一个妈创建的标准——内积度量矩阵总是单位阵。这样一来，原来的向量正交的换了基坐标系仍然正交，不正交的仍然不正交。

其实，前面讲过，**要保持正交的仍然正交，不正交的仍然不正交有两个方法（本质上一回事）：一个是把基正交化，用老办法计算内积就可以了；一个是不把基正交化，改变内积的计算——中间插一个度量矩阵。**

拼命保内积计算的一致就是考虑了内积的物理意义——功/能量在里面，这是物理学上的有力应用，没有什么问题。如果从几何学上的应用角度出发——不保内积保外积也是非常好的应用，保外积的应用就是保面积保体积，亦即我们讨论过的雅可比矩阵或外微分的东东了。

在前面讨论雅可比矩阵时提到过曲线坐标系，其实，只要我们**首次**取定了曲线作为坐标系，这时候自然是正交曲线坐标系，即使在笛卡儿坐标系下看起来不那么垂直（我们不保正交的一致性了）。同时在新坐标系下的坐标微分向量 $\mathrm{d}x$ 和 $\mathrm{d}y$ 当然正交，它们可以直接相乘得到积分面积微元 $\mathrm{d}x\mathrm{d}y$（雅可比行列式乘以面积微元等于原坐标系下积分面积微元，这个没有变，保的就是这个）。

好，言归正传，继续我们的保内积的正交变换的讨论。

2. 正交变换和正交矩阵的几何意义

正交变换是保持任意向量的长度不变或者保持度量不变的线性变换。正交变换是欧氏空间中一类重要的变换。设 $\boldsymbol{\sigma}$ 是欧氏空间 V 的一个变换，则下列条件是等价的：

（1）$\boldsymbol{\sigma}$ 是空间 V 上的正交变换；

（2）$\boldsymbol{\sigma}$ 保持向量的内积不变；

（3）$\boldsymbol{\sigma}$ 保持向量的长度和夹角不变；

（4）对任意向量$\xi,\eta \in V$，有$|\sigma(\xi)+\sigma(\eta)|=|\xi+\eta|$；

（5）保持向量的长度不变且满足条件：对任意向量$\xi,\eta \in V$，有$\sigma(\xi+\eta)=\sigma(\xi)+\sigma(\eta)$；

（6）保持向量的距离不变且对任意向量$\xi \in V$，有$\sigma(-\xi)=-\sigma(\xi)$。

正交变换对图形的变换主要是旋转和镜像。在二维平面空间里，镜像变换可以看成旋转π个角度的旋转变换。但在三维空间里，镜像不能够看成旋转变换，比如一个人的面孔无论如何旋转，也不能与镜子里的面孔重合。左手如何旋转也不能变成右手，但镜子就可以把你的左手变成右手。这也是三维直角坐标系有左手系和右手系的区别原因之所在。

正交变换具有合同与相似变换共同的优点。合同变换与相似变换各有所长。前者仅适用于对称阵，保持了矩阵的对称性、正定性、秩等性质不变；后者适用于一般方阵，保持了矩阵的秩与特征值不变。能兼有两者优点的是正交变换。

正交变换的矩阵$\boldsymbol{P}$满足$\boldsymbol{P}^{-1}=\boldsymbol{P}^{\mathrm{T}}$，实际上是合同变换与相似变换的一种结合，因而具有两者的优点。

这里$\boldsymbol{A}$的特征向量经正交规范化后实为原向量的线性组合，故仍为特征向量。因此，化成对角阵后，对角线上的元正好就是$\boldsymbol{A}$的全部特征值。因此正交变换也保留了原矩阵的特征值不变的的性质。

正交变换法最大的优点是某个问题经正交变化为标准形后，其几何图形保持不变。

正交变换的主要性质是它不改变几何图形的度量。正交变换对应的矩阵就是正交矩阵，正交矩阵是一种很重要的特殊方阵，其基本属性是$\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{E}$或$\boldsymbol{Q}^{-1}=\boldsymbol{Q}^{\mathrm{T}}$。简单地说，一个正交矩阵就是一个具有标准正交列向量组的方阵。实际上，它的行向量组也是一个标准正交组，这个结论可由$\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{E}$直接推出。从几何上说，列向量相互垂直，行向量也相互垂直。

在$\mathbf{R}^2$空间上，如果说一个二阶矩阵$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$是正交矩阵，那么就有$a_1^2+a_2^2=1$，$b_1^2+b_2^2=1$，$a_1^2+b_1^2=1$和$a_1b_1+a_2b_2=0$等条件。二阶正交矩阵只有如下的 2 种类型：

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix}$$

因为平面上的正交变换可以归结为只有旋转变换，平面上的旋转只有逆时针和顺时针两个方向，因此第一个矩阵是逆时针旋转，第二个矩阵是顺时针旋转。

第一个矩阵的行列式等于 1，是个标准的旋转矩阵；第二个矩阵又可以通过第一个矩阵和关于x轴的反射变换矩阵的积来表示：

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

因此二阶实正交矩阵的几何意义也可以认为是绕平面上原点的旋转变换或者是旋转加以坐标轴为对称轴的镜像反射变换。这种说法可以兼顾三维空间的旋转。

三维空间$\mathbf{R}^3$中的正交矩阵实际上也是两种类型：一个是右手坐标系下的旋转矩阵，一个是左手坐标系下的旋转矩阵；它们在三维空间的标准正交基一节（见 4.2.6 节）中作为基向量已讨论过，我们这里把它们组合成矩阵：

$$\begin{bmatrix} -\sin\varphi & \cos\theta\cos\varphi & \sin\theta\cos\varphi \\ \cos\varphi & \cos\theta\sin\varphi & \sin\theta\sin\varphi \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sin\varphi & \cos\theta\cos\varphi & \sin\theta\cos\varphi \\ -\cos\varphi & \cos\theta\sin\varphi & \sin\theta\sin\varphi \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

其中 $0\leqslant\theta\leqslant\pi$，$0\leqslant\varphi\leqslant 2\pi$；或者 $0\leqslant\theta\leqslant 2\pi$，$0\leqslant\varphi\leqslant\pi$。

这两种矩阵所表示的几何意义就是三阶正交矩阵的几何意义。如果取 $0\leqslant\theta\leqslant 2\pi$，$\varphi=\pi/2$（见图 4-37），那么分别有下面简化的旋转矩阵：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

第一个矩阵是以 x 坐标轴为旋转轴的旋转变换矩阵，第二个矩阵是以 x 坐标轴为旋转轴的旋转变换后再伴随一个垂直此旋转轴的平面（即 xoy ）为镜面的反射变换矩阵。

另外，如果 n 维欧氏空间中有两个标准正交基，那么两个标准正交基的过渡矩阵就是正交矩阵。

正定矩阵不是正交矩阵

因为这两个概念只有一字之差，很容易混淆或记不清，掌握这两种矩阵的判别方法是(充要条件)：正交矩阵的列(行)向量是两两正交的单位向量；正定矩阵的特征值全为正(或各阶主子式全为正)。并且正定矩阵一定是对称矩阵，这一点往往在证明题中是隐含的条件。

5.14.5 分块矩阵的代数及几何意义

在计算阶数较大的矩阵特别是较大的稀疏矩阵的时候（这在大型的方程组计算时是常有的事，比如在设计波音 777 或黑丝带 J20 的机身表面线型轮廓时，所用到的空气流体动力学的方程组，方程数和变量数达几百万个，矩阵的阶数也是百万级的，但这个矩阵内部的元素好多都是零元素，非零的元素很稀疏），为了尽量简化计算，降低计算量，我们常常把矩阵进行分块。对矩阵进行分块相当于把一个大矩阵看成由一些小矩阵所组成。在对这个分块矩阵进行计算(主要是矩阵的乘法）时，就把小矩阵当成大矩阵的元素来进行处理。例如：

$$\left[\begin{array}{cc:cc} a & 1 & 0 & 0 \\ 0 & a & 0 & 0 \\ \hdashline 1 & 0 & b & 1 \\ 0 & 1 & 1 & b \end{array}\right] = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{E} & \boldsymbol{B} \end{bmatrix}$$

分块后，矩阵的表示果然简洁、结构清晰。但为什么可以把小矩阵当成一个元素来处理呢？有人说这揭示了矩阵结构的深刻特性。实际上我们既可以用矩阵的眼光来看分块和元素，这些都是矩阵；也可以把这些都看做向量，元素是向量、行和列是元素和的向量，矩阵也是行或列的和的向量。所以我们可以从如下两个方面认识矩阵分块的意义。

1. **从向量及内积角度看分块（几何意义）**

两个矩阵相乘 $\boldsymbol{AB}$，这个乘法规则包含着一个嵌套的结构：左矩阵 $\boldsymbol{A}$ 的行向量和右矩阵 $\boldsymbol{B}$ 的列向量逐个按顺序相乘运算，而每组行向量乘以列向量的运算里面又是每个元素逐个按顺序相乘；最后一步是相加得到内积（每个内积再逐个按序放在坐标轴上逆变成向量的过程）。

嵌套的计算顺序暗示着矩阵元素和行/列向量具有一致的特性，因此如果用向量的眼光来统

一看矩阵，矩阵的每个元素实际上也是向量，是位于或平行于坐标轴的向量；行向量的第 i 个元素乘以列向量的第 i 个元素，就是同一个轴上两个向量的乘积，也是内积运算。

矩阵可以分列/分行是因为一列或一行的元素（向量）可以相加构成一个新向量，称之为列向量和行向量，它们之间的乘积当然是内积的运算。

如果左矩阵里面两行的行向量组成一块，就是两个行向量相加构成一个行向量块，向量块也是向量，那么行向量块就可以与右矩阵里面的列向量块进行更高一级的内积运算。

从整个矩阵上看也是左向量组合和右向量组合的内积运算，其结果——矩阵也就是内积的排列。既然如此，矩阵可以任意切割和组合，这就是分块的原因。

还是不能明了？看看下面的解释就释然了。

2.　从线性变换或线性方程组的角度看分块（代数意义）

为了更坚信矩阵可以分块，我们从线性变换的角度进行解读。因为矩阵乘法可以从线性代换得到推导，分块当然也可以。

三阶的线性变换（或叫线性方程组，叫变换更贴切些）如下：

$$\begin{cases} y_1 = a_1x_1 + a_2x_2 + a_3x_3 \\ y_2 = b_1x_1 + b_2x_2 + b_3x_3 \\ y_3 = c_1x_1 + c_2x_2 + c_3x_3 \end{cases}$$

写成向量及矩阵形式为

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

把向量和矩阵分块为（先把矩阵任意分，向量再配合着分）

$$\left(\begin{array}{c} y_1 \\ y_2 \\ \hdashline y_3 \end{array}\right) = \left[\begin{array}{c:cc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ \hdashline c_1 & c_2 & c_3 \end{array}\right] \left(\begin{array}{c} x_1 \\ \hdashline x_2 \\ x_3 \end{array}\right)$$

按照拆开的格式再把上式写回方程组的形式：

$$\begin{cases} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} x_1 + \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \\ y_3 = c_1x_1 + \begin{bmatrix} c_2 & c_3 \end{bmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \end{cases} \text{或} \begin{cases} \begin{cases} y_1 = a_1x_1 + a_2x_2 + a_3x_3 \\ y_2 = b_1x_1 + b_2x_2 + b_3x_3 \end{cases} \\ y_3 = c_1x_1 + c_2x_2 + c_3x_3 \end{cases}$$

噢好，原来矩阵分解成块就是方程组分成了更小的方程组；原来的三维向量分解成了二维或一维向量了。继续看两个矩阵相乘的分解：

两个三阶的线性变换（或叫线性方程组）如下：

$$\begin{cases} \begin{cases} z_1 = e_1y_1 + e_2y_2 + e_3y_3 \\ z_2 = f_1y_1 + f_2y_2 + f_3y_3 \\ z_3 = g_1y_1 + g_2y_2 + g_3y_3 \end{cases} \\ \begin{cases} y_1 = a_1x_1 + a_2x_2 + a_3x_3 \\ y_2 = b_1x_1 + b_2x_2 + b_3x_3 \\ y_3 = c_1x_1 + c_2x_2 + c_3x_3 \end{cases} \end{cases} \text{或者} \begin{cases} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{bmatrix} e_1 & e_2 & e_3 \\ f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \end{cases}$$

把它们的矩阵分别分块为如下的格式：

$$\left\{\begin{array}{l}
\begin{pmatrix} z_1 \\ z_2 \\ \hline z_3 \end{pmatrix} = \left[\begin{array}{c:cc} e_1 & e_2 & e_3 \\ f_1 & f_2 & f_3 \\ \hdashline g_1 & g_2 & g_3 \end{array}\right] \begin{pmatrix} y_1 \\ \hdashline y_2 \\ y_3 \end{pmatrix} \\
\begin{pmatrix} y_1 \\ \hdashline y_2 \\ y_3 \end{pmatrix} = \left[\begin{array}{cc:c} a_1 & a_2 & a_3 \\ \hdashline b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array}\right] \begin{pmatrix} x_1 \\ x_2 \\ \hdashline x_3 \end{pmatrix}
\end{array}\right.$$

按照拆开的格式再把上式写回方程组的形式：

$$\left\{\begin{array}{l}
\left\{\begin{array}{l}
\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} e_1 \\ f_1 \end{bmatrix} y_1 + \begin{bmatrix} e_2 & e_3 \\ f_2 & f_3 \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} \\
z_3 = g_1 y_1 + \begin{bmatrix} g_2 & g_3 \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \end{bmatrix}
\end{array}\right. \\
\left\{\begin{array}{l}
y_1 = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + a_3 x_3 \\
\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_3 \\ c_3 \end{bmatrix} x_3
\end{array}\right.
\end{array}\right.$$

看起来，一个矩阵可以任意方式分块，但另外一个必须配合分块，否则下面的方程组无法代入上面的方程组。

方程组的代入过程，就是两个矩阵相乘的过程。左、右两个矩阵相乘就是右边的矩阵所表示的线性方程组代入左边的矩阵所表示的线性方程组（参见 6.8.1 节）。

由方程组代入关系得到如下的关系式，这正是矩阵分块运算的规则：

$$\left[\begin{array}{c:cc} e_1 & e_2 & e_3 \\ f_1 & f_2 & f_3 \\ \hdashline g_1 & g_2 & g_3 \end{array}\right] \left[\begin{array}{cc:c} a_1 & a_2 & a_3 \\ \hdashline b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array}\right] = \begin{bmatrix} \begin{bmatrix} e_1 \\ f_1 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \end{bmatrix} + \begin{bmatrix} e_2 & e_3 \\ f_2 & f_3 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}, & \begin{bmatrix} e_1 \\ f_1 \end{bmatrix} a_3 + \begin{bmatrix} e_2 & e_3 \\ f_2 & f_3 \end{bmatrix} \begin{bmatrix} b_3 \\ c_3 \end{bmatrix} \\ g_1 \begin{bmatrix} a_1 & a_2 \end{bmatrix} + \begin{bmatrix} g_2 & g_3 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}, & g_1 a_3 + \begin{bmatrix} g_2 & g_3 \end{bmatrix} \begin{bmatrix} b_3 \\ c_3 \end{bmatrix} \end{bmatrix}$$

一般地，我们可以任意地对矩阵进行分块，但分块时要保证矩阵分块后同一行的小矩阵内的元素具有相同的行数；同一列的小矩阵内的元素具有相同的列数。实际上，只要用横线和竖线贯穿整个矩阵就可以了。如：

$$\left[\begin{array}{cc:cc:c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{array}\right]$$

矩阵分块后，其加法、乘法和数乘等计算与矩阵未分块的计算法则相同。下面我们逐一看看矩阵分块方法及运算的注意事项：

加法：两个分块矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相加，只有 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的分块规则相同才能相加。在作加法时，先把 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 中的每一分块看做一个元素（一个整体）对应相加，然后再把 $\boldsymbol{A}+\boldsymbol{B}$ 中的块看做两个矩阵的和。

数乘：一个常数 k 乘以一个分块矩阵 $\boldsymbol{A}$，同样是常数首先乘以作为一个元素的分块，然后才是常数与小矩阵的乘积。

乘法：两个分块矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相乘，首先要求矩阵 $\boldsymbol{A}$ 的分块 A_{ij} 中元素的列数和矩阵 $\boldsymbol{B}$ 的分块 B_{jk} 中元素的行数相同才能进行运算。这里有个简单的记法就是，**两个矩阵的分块线构成的“十”或“井”字图形是平面上的斜轴对称或镜像图形。**

例如：

$$\boldsymbol{AB}=\left[\begin{array}{c:c:ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{array}\right]\left[\begin{array}{cc:c} b_{11} & b_{12} & b_{13} \\ \hdashline b_{21} & b_{22} & b_{23} \\ \hdashline b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \\ b_{51} & b_{52} & b_{53} \end{array}\right]$$

显然，我们对矩阵 $\boldsymbol{A}$ 的列的分块原则用在了矩阵 $\boldsymbol{B}$ 行的分块上，而对矩阵 $\boldsymbol{A}$ 的行的分块原则用在了矩阵 $\boldsymbol{B}$ 列的分块上。

转置：与矩阵的数乘法则类似，分块矩阵的转置，先对每个小的分块作为一个整体元素对矩阵转置，然后还要对每个小的分块再作一次转置。

5.14.6　三角矩阵几何意义

三角矩阵是形如

$$\begin{bmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

的矩阵。三角矩阵的意义可以通过把一个一般矩阵化为三角的过程给出几何解释，过程清楚了，那么三角矩阵的意义就不言而喻了。

我们来研究三维几何空间中的三个不共面向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 的表示（三个向量不共面即线性无关）。

如图 5-80(a)所示，先看向量 $\boldsymbol{a}$。在 $\boldsymbol{a}$ 的方向上任选一个非零向量 $\boldsymbol{\alpha}$ 来表示 $\boldsymbol{a}$，即 $\boldsymbol{a}=a_1\boldsymbol{\alpha}$。把 $\boldsymbol{a}$ 表示成有序数组或坐标形式为 $\boldsymbol{a}=(a_1,0,0)$，其中 $a_1\neq 0$，后边分量取 0 是因为 $\boldsymbol{a}$ 的表示式中不需要任何其他向量。

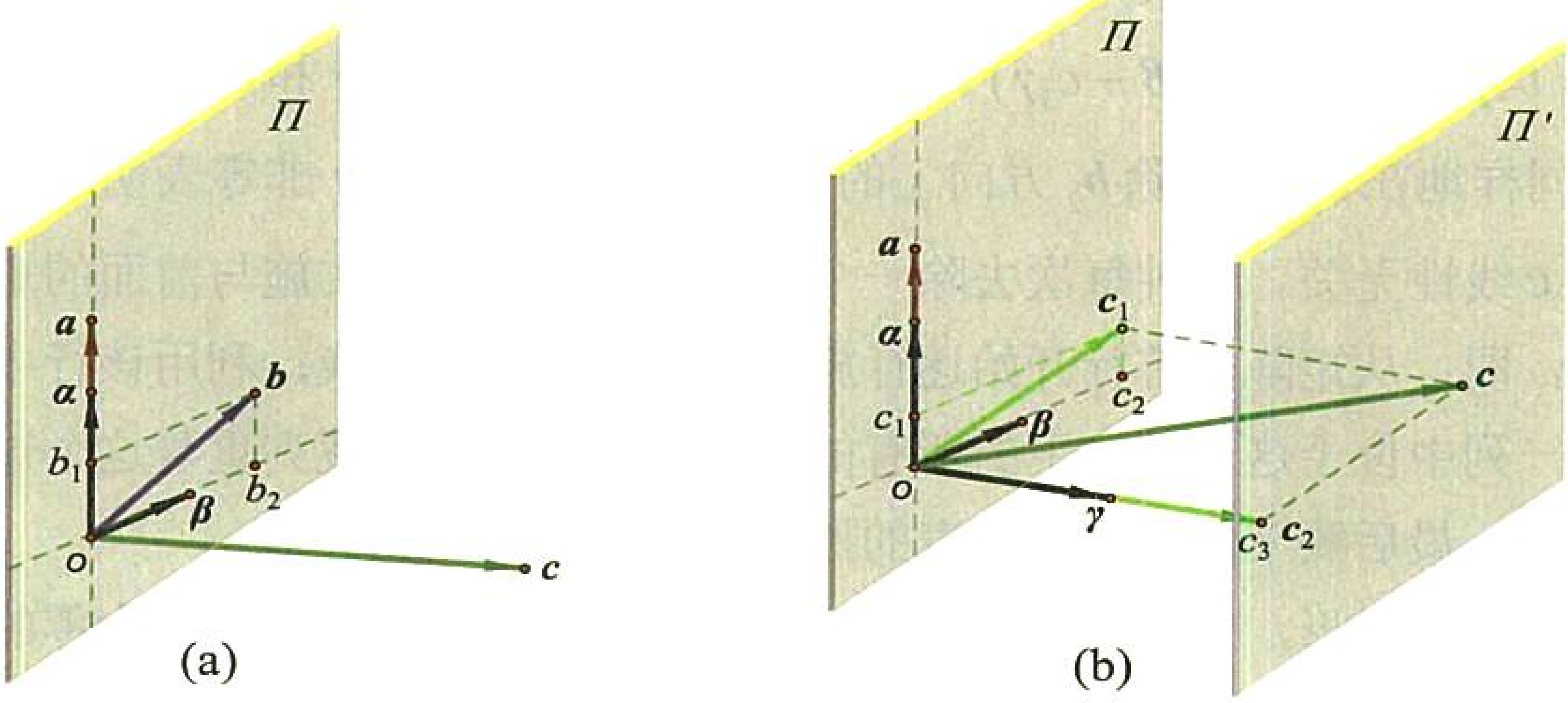


图 5-80 矩阵三角化的几何解释

再看向量 $\boldsymbol{b}$。设 $\boldsymbol{\alpha}$ 与 $\boldsymbol{b}$ 所确定的平面为 Π，将 $\boldsymbol{b}$ 按平行四边形法则分解成两个分向量；

一个分向量与$\boldsymbol{\alpha}$共线，记为$\boldsymbol{b}=b_1\boldsymbol{\alpha}$（不能保证$b_1\neq 0$，因为如果$\boldsymbol{\alpha}$与$\boldsymbol{b}$正交时$b_1=0$）；第二个分向量的终点必然落在通过$\boldsymbol{b}$的终点且与$\boldsymbol{\alpha}$平行的直线上。在第二个分向量所在直线上任取一个非零向量$\boldsymbol{\beta}$，则第二个分向量可以记为$\boldsymbol{b}_2=b_2\boldsymbol{\beta}$，其中$b_2\neq 0$。于是$\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{b}_2=b_1\boldsymbol{\alpha}+b_2\boldsymbol{\beta}$，将其表示成坐标形式为$\boldsymbol{b}=(b_1,b_2,0)$。

再看向量$\boldsymbol{c}$，如图 5-80(b)所示。$\boldsymbol{c}$可以分解为两个向量$\boldsymbol{c}_1$和$\boldsymbol{c}_2$。$\boldsymbol{c}_1$在平面Π上，它可以用$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$表示为$\boldsymbol{c}_1=c_1\boldsymbol{\alpha}+c_2\boldsymbol{\beta}$（$c_1$和$c_2$可能全部为零，如$\boldsymbol{c}$与平面$\Pi$正交时）。而$\boldsymbol{c}_2$的终点必然落在通过$\boldsymbol{c}$的终点且与$\Pi$平行的平面上。在$\boldsymbol{c}_2$所在直线上任取非零向量$\boldsymbol{\gamma}$，则$\boldsymbol{c}_2=c_3\boldsymbol{\gamma}$，其中$c_3\neq 0$。至此，向量$\boldsymbol{c}$最终可表示为$\boldsymbol{c}=c_1\boldsymbol{\alpha}+c_2\boldsymbol{\beta}+c_3\boldsymbol{\gamma}$，坐标形式为$\boldsymbol{c}=(c_1,c_2,c_3)$，其中$c_3$不等于零。

至此，三个向量都表示完毕，它们在基$\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$上表示为$\boldsymbol{a}=(a_1,0,0)$，$\boldsymbol{b}=(b_1,b_2,0)$，$\boldsymbol{c}=(c_1,c_2,c_3)$。三个向量作为行向量写成矩阵形式就是下三角阶梯矩阵：

$$\begin{bmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

其中对角线元素a_1、b_2和c_3全部不为零。

对角线元素a_1、b_2和c_3全部不为零，这是因依次增加的向量都不能被之前的向量线性表示而产生的一个不为零的分量。如果把三个向量$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$的表示式按$\boldsymbol{c}$、$\boldsymbol{b}$、$\boldsymbol{a}$的顺序重新排列，按$\boldsymbol{\gamma}$、$\boldsymbol{\beta}$、$\boldsymbol{\alpha}$的顺序表示成矩阵，其矩阵形式是上三角阶梯矩阵：

$$\begin{bmatrix} c_3 & c_2 & c_1 \\ 0 & b_2 & b_1 \\ 0 & 0 & a_1 \end{bmatrix}$$

显然，**对一个矩阵化为阶梯矩阵的过程，实质就是选择一个合适的基进行坐标变换的过程。在新基上，向量的坐标排列表现为一个阶梯形。**

我们常讲的阶梯矩阵就是三角矩阵，大家常常使用矩阵阶梯化来求解最大向量无关组。

阶梯形成之处的元素仍是a_1、b_2、c_3，可见呈阶梯形的非零元素反映了行向量在该位置上有独立分量。我们可以用这些独立分量来识别线性无关的向量组。仍用先前的例子，把$\boldsymbol{a}$作为第一个线性无关的向量，并用它来判定与$\boldsymbol{b}$和$\boldsymbol{c}$的线性相关性；在$\boldsymbol{b}$和$\boldsymbol{c}$上去除$\boldsymbol{a}$方向上的分量后，得到$\boldsymbol{b}_2\ (=b_2\boldsymbol{\beta})$和$\boldsymbol{c}_2\ (=c_2\boldsymbol{\beta}+c_3\boldsymbol{\gamma})$，这两个向量非零，表明$\boldsymbol{b}$和$\boldsymbol{c}$都独立于$\boldsymbol{a}$。再选取$\boldsymbol{b}_2$（它代表$\boldsymbol{b}$），同样地在$\boldsymbol{c}_2$上去除$\boldsymbol{b}_2$方向上的分量，得到$c_3\boldsymbol{\gamma}$，$c_3\boldsymbol{\gamma}$非零表明$\boldsymbol{c}$独立于$\boldsymbol{a}$和$\boldsymbol{b}$，于是$\boldsymbol{a}$、$\boldsymbol{b}$和$\boldsymbol{c}$线性无关。这种每次去除一个分量的方法用矩阵实施与前面的矩阵元素逐列化零是同一过程，即：从矩阵的首行开始逐行施行第三种初等行变换，利用该行中列标最靠前的非零数，将这一列中位于这个数之下的所有的元素全部化为零。逐列重复这个过程，直到行变换不能进行为止，最后剩下的非零行代表的原始向量就是最大无关组。

为了保证得到阶梯形，还得适时把非零元素最靠前的行上移或把已化为零的行下移，这种移动的实质是交换若干次行向量，因此需要引入第一种初等行变换。但是，如果要上移矩阵的行向量，只能在矩阵的前一列元素化零完成之后和后一列元素化零开始之前进行，上移这个向

量的目的是将它选为下一个目标向量并用它去除随后的向量中的分量，保证变换后的每一行都是它前面的行(如果存在)的线性组合。为了数值计算的方便，我们还需要用非零数去乘矩阵的某一行，目的是把该行元素都表示为整数，或者是把该行的某个元素化为 1 这样的特别值，这就需要引入第二种初等行变换。如果不是刻意追求阶梯形和人工计算的方便，仅用第三种初等行变换足以解决求最大无关组的问题。

5.14.7　对角矩阵的几何意义

对角矩阵可以使用上面的化三角矩阵的过程进行解释，只是基的选择有些不同而已。对角矩阵的基向量的选择特殊一些，它是选择与被对角化的向量同方向的向量作为基向量，使三角变成了对角，矩阵变得更加简单。下面我们简单地说明一下。

和三角化的选基类似，在向量 $\boldsymbol{a}$ 的方向上任选一个非零向量 $\boldsymbol{\alpha}$ 来表示 $\boldsymbol{a}$（见图 5-81），即 $\boldsymbol{a}=a_1\boldsymbol{\alpha}$。把 $\boldsymbol{a}$ 表示成有序数组或坐标形式为 $\boldsymbol{a}=(a_1,0,0)$，其中 $a_1\neq 0$。

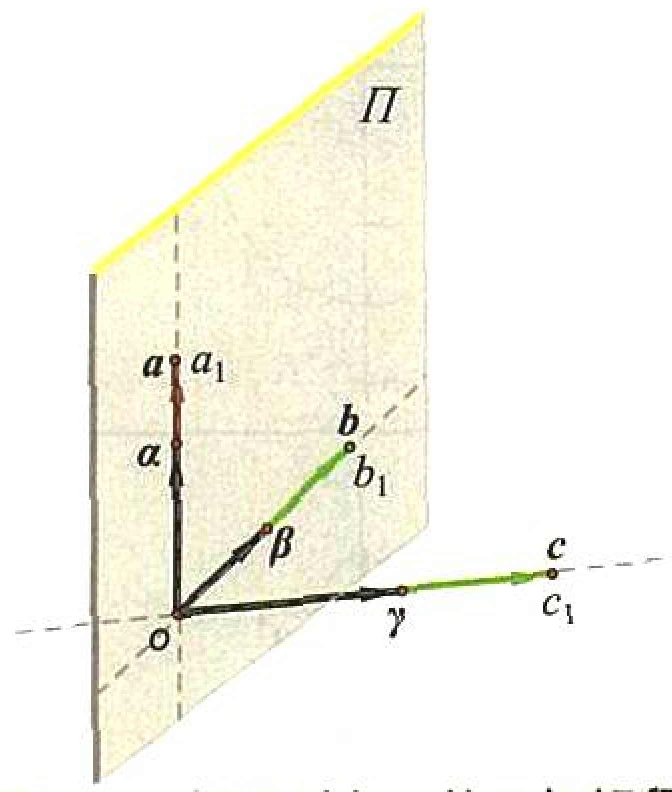


图 5-81　矩阵对角化的几何解释

以后的两个向量基的选择拷贝上面的做法，分别选择和向量 $\boldsymbol{b}$ 、$\boldsymbol{c}$ 方向相同的基向量，那么就有 $\boldsymbol{b}=b_1\boldsymbol{\beta}$ 和 $\boldsymbol{c}=c_1\boldsymbol{\gamma}$。将其表示成坐标形式为 $\boldsymbol{b}=(0,b_1,0)$，$\boldsymbol{c}=(0,0,c_1)$。写成矩阵形式就是对角矩阵形式。

如果选取的基向量就是原向量组（假设为线性无关），那么得到的矩阵就是单位矩阵。如果据此把任意一个矩阵直接写成单位矩阵也没有什么意思了，原矩阵和单位矩阵之间的关系就是前面讲的等价关系并有等式成立：

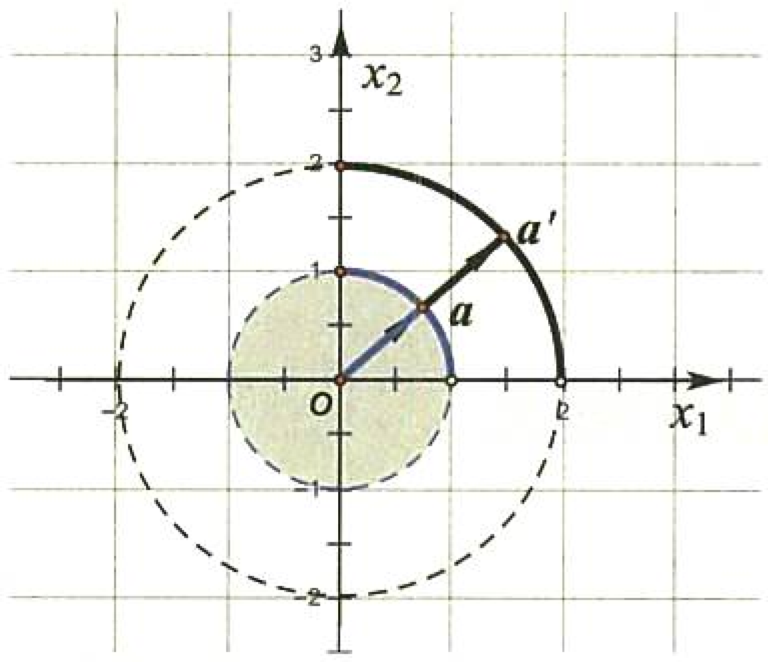
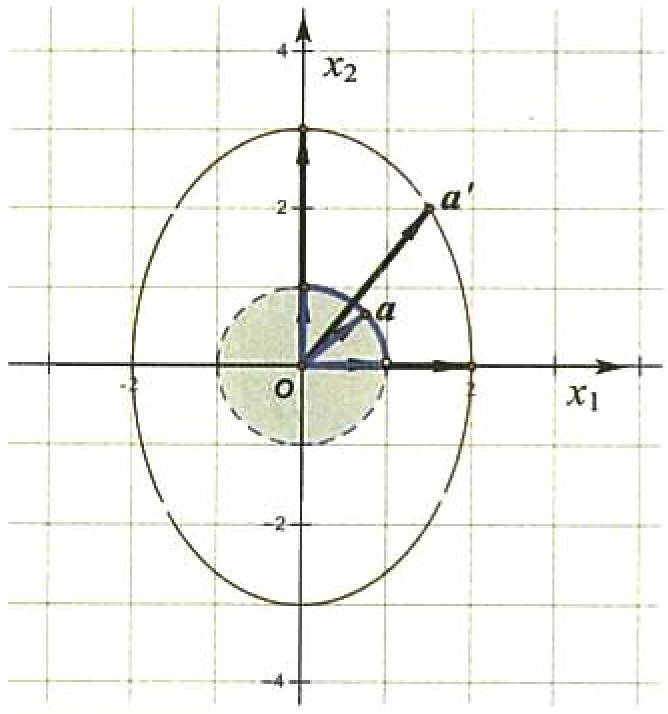
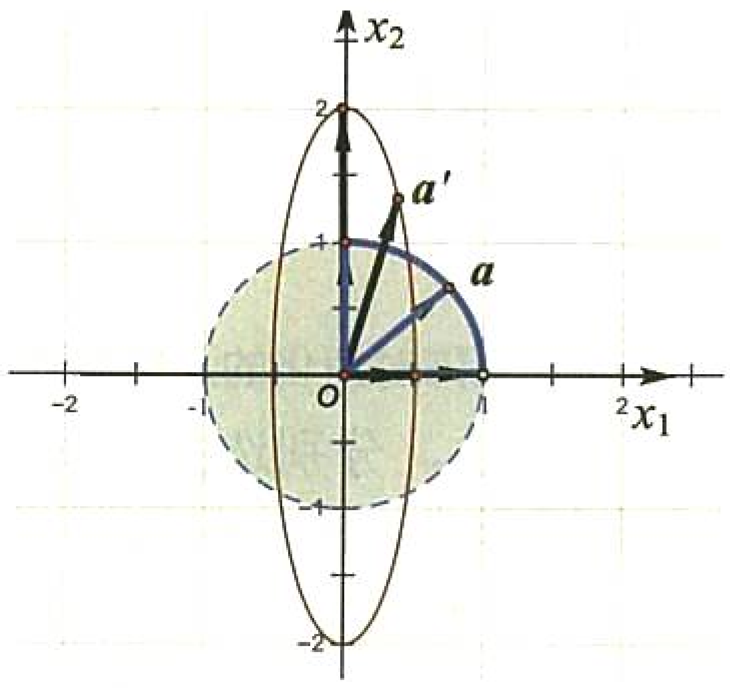
$$\boldsymbol{B}=\boldsymbol{PAQ}=\begin{bmatrix}\boldsymbol{E}_k & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$$

我们得到的对角矩阵或单位矩阵显然正交。化矩阵对角矩阵或标准形的实质是选取了正交基后得到的等价矩阵。因为我们分别选取了同向的基向量，所以每个原向量在这个新坐标系下在另外的坐标轴上的平行投影为零，看起来彼此正交。

相似矩阵对角化里有个对角矩阵的概念，在前面等价矩阵的几何意义一节中也讲过任一矩阵都等价于一个对角矩阵。在后面的二次型中，也有个合同对角化的问题，其实都是选取基的操作。

对角矩阵具有很优越的特性，在 3D 图形处理中用于进行扩大/缩小操作。作为线性变换的对角矩阵，我们给出几个图形变换效果，如表 5-7 所示。

表 5-7 对角矩阵所表示的线性变换的图形

顺序	对角矩阵	对向量圆的变换	注解
1	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$		对角线上的元素相等，对对向量和图形进行均匀放大或缩小
2	$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$		对角线上的元素不等，对向量和图形进行非均匀的缩放。如 x_1 轴方向的放大倍数小，x_2 轴方向的放大倍数大。因此单位圆变换成了放大或缩小的椭圆形
3	$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$		

5.14.8 平移矩阵的几何意义

计算机图形学（如动漫、电影特技、游戏制作以及现代药物设计中模拟生物分子模型的化学反应等）中的数学与矩阵运算联系紧密。屏幕上的图形需要移动操作，但是屏幕上的物体平移并不直接对应于一个矩阵。因为平移变换不是线性变换。那么如何用矩阵来表示一个平移呢？

在回答这个问题之前，有人敏锐地指出，平移很简单嘛，一个向量加上一个平移量（也是向量）不就是把向量的顶点平移了吗！比如图 5-82 所示平面上的点 p 平移到 p': $\boldsymbol{p} \to \boldsymbol{p}'$，用向量表述如下：

$$\boldsymbol{p} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \boldsymbol{p}' = \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}, \quad \boldsymbol{t} = \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{pmatrix}$$

那么，有 $\boldsymbol{p}' = \boldsymbol{p} + \boldsymbol{t}$，即

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{pmatrix} = \begin{pmatrix} x_1 + \mathrm{d}x_1 \\ x_2 + \mathrm{d}x_2 \end{pmatrix}$$

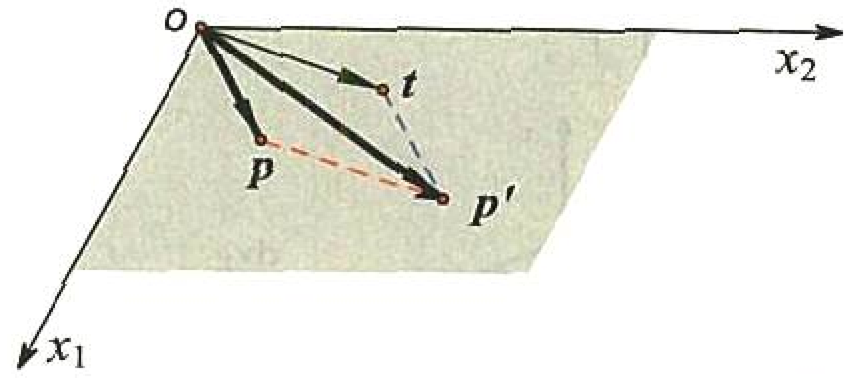


图 5-82 用向量加法可以实现点的平移

如果把一个图形平移，只需把图形各个顶点的向量都加上平移向量就可以了。

没错。但是这样做不方便，如果一个图形有成千上万个顶点，这样处理的话就不太数学了。如果能把平移用一个矩阵乘法来表示的话，我们对一个图形先旋转再平移的操作只要把图形矩阵乘以旋转矩阵然后再乘上位移矩阵即可；根据矩阵乘法的意义，旋转和平移也可以合并成一个矩阵。这是一个好主意。

那么如何得到平移矩阵呢？解决这一困难的标准办法是引入所谓的齐次坐标。齐次坐标就是把点坐标扩增一个坐标值：$(x_1,x_2)\to(x_1,x_2,1)$，二维平面$\{x_1ox_2\}$扩展成为三维空间中一个特定的平面$x_3=1$。这样一扩展，二维向量就变成了三维向量。其行向量的平移矩阵常用形式为

$$T=\begin{bmatrix}1&0&0\\0&1&0\\\mathrm{d}x_1&\mathrm{d}x_2&1\end{bmatrix}$$

其几何意义是把平面上的平移变成了立体上的切变，这样我们的目的就达到了：平面上对图形进行了平移。

如图 5-83 所示中平面上的点$\boldsymbol{p}\to\boldsymbol{p}'$的平移变成了三维空间中向量$\boldsymbol{a}$到$\boldsymbol{a}'$的切变$\boldsymbol{a}\to\boldsymbol{a}'$。而这个切变$\boldsymbol{a}\to\boldsymbol{a}'$是可以用三阶矩阵来表示的。平移是切变在$\{x_1ox_2\}$平面上的投影。或者说，二维空间上的平移可以用三维空间上的切变来投影得到。

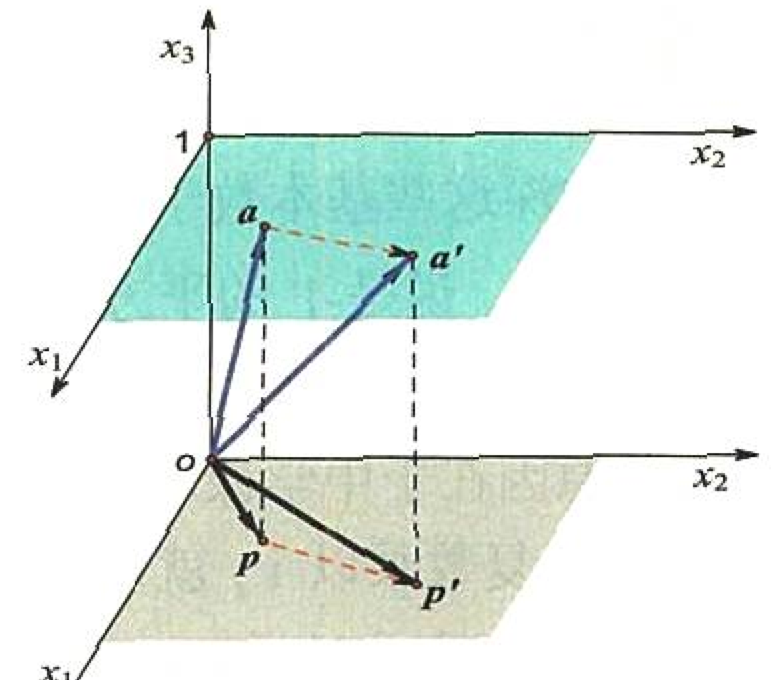


图 5-83 空间里向量的切变投影成点的平移

我们验证一下（列向量的平移），原 p 点坐标修改为齐次坐标为$\boldsymbol{a}=(x_1,x_2,1)^{\mathrm{T}}$，那么经平移变换后为

$$\boldsymbol{Ta}=\begin{bmatrix}1&0&\mathrm{d}x_1\\0&1&\mathrm{d}x_2\\0&0&1\end{bmatrix}\begin{pmatrix}x_1\\x_2\\1\end{pmatrix}=\begin{pmatrix}x_1+\mathrm{d}x_1\\x_2+\mathrm{d}x_2\\1\end{pmatrix}$$

同样，三个向量的顶端$\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$构成的一个三角形也被平移了（见图 5-84）：

$$\begin{bmatrix}1&0&\mathrm{d}x_1\\0&1&\mathrm{d}x_2\\0&0&1\end{bmatrix}\begin{bmatrix}a_1&b_1&c_1\\a_2&b_2&c_2\\1&1&1\end{bmatrix}=\begin{bmatrix}a_1+\mathrm{d}x_1&b_1+\mathrm{d}x_1&c_1+\mathrm{d}x_1\\a_2+\mathrm{d}x_2&b_2+\mathrm{d}x_2&c_2+\mathrm{d}x_2\\1&1&1\end{bmatrix}$$

三角形平移前后的图形都在平面 $x_3 = 1$ 上。

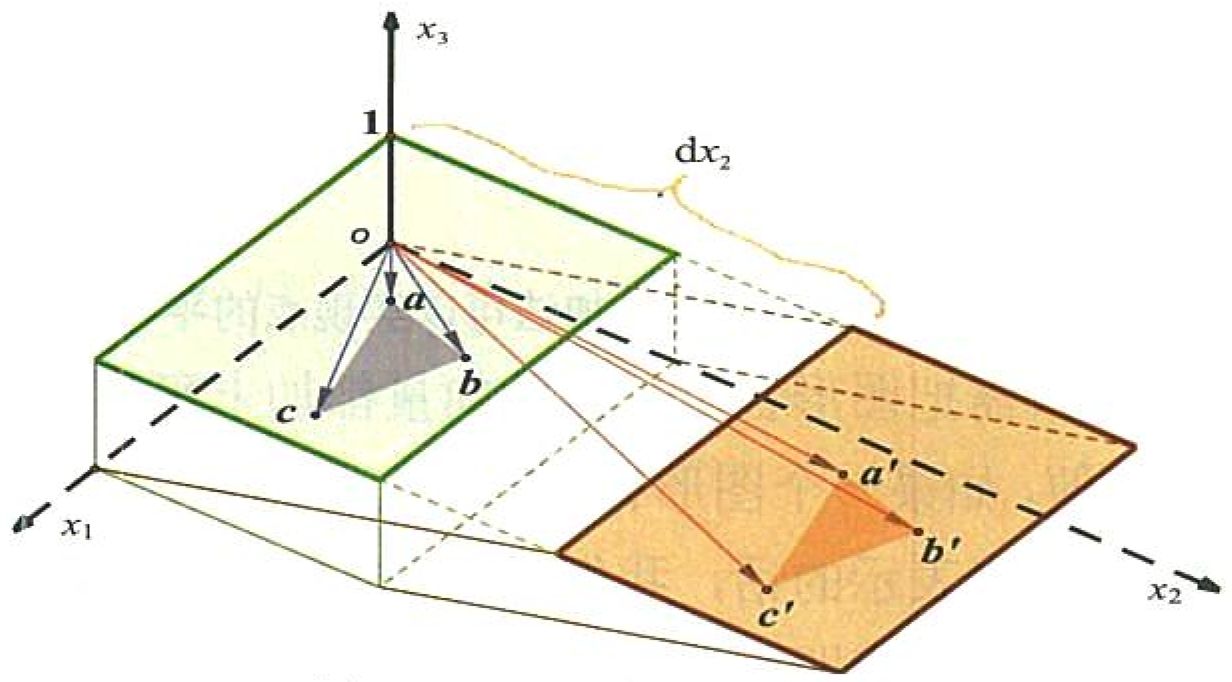


图 5-84 空间里三角形在 $x_3 = 1$ 平面上的平移

以此类推，三维空间中物体的平移使用了四维的向量组成的切变矩阵；四阶矩阵正是计算机图形学中主要的变换矩阵形式。

这个平移矩阵之所以可以实现，其实是利用了向量和点的不同处。向量切变时，向量都是通过原零点的，在矩阵（线性变换）乘法下向量永远都不会发生平移，但向量的端点却在一个平面上发生了平移运动。

n 维平面上没有 n 阶平移矩阵吗？

n 维空间的几何图形要平移的话需要 n+1 阶的平移矩阵，换句话说，n 阶矩阵可以平移 n-1 维空间里的所有几何图形。比如二维矩阵可平移一个点而不能平移一个线段，三维矩阵可平移整个平面上的点、线段、整条直线及平面复杂图形而不能平移一个立方体。那么 n 维空间真的没有 n 阶平移矩阵吗？

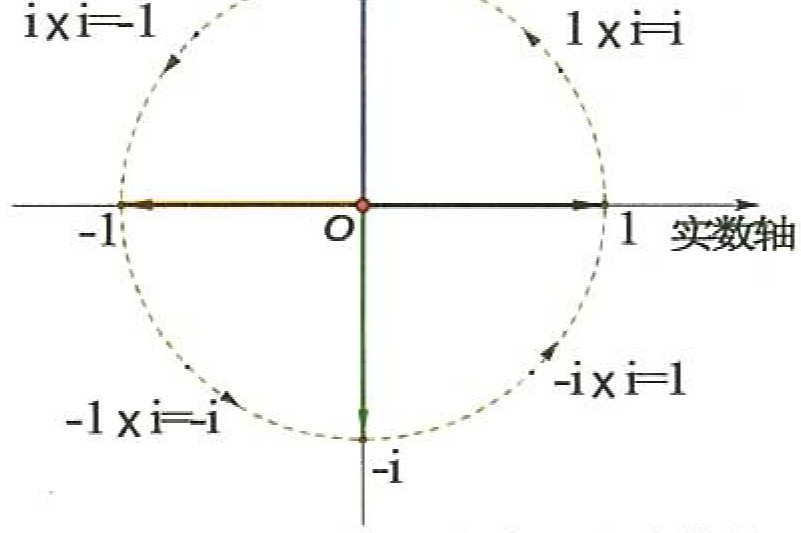
严格说来，有平移矩阵，它就是 n 阶单位方阵。单位方阵把所有空间内的向量都可以平移 0 个距离。可以证明，除了单位方阵外确实没有一般意义上的平移矩阵了。

5.14.9　复数的矩阵表示

讨论矩阵与复数的关系将进一步加深这些基本数学概念的理解。

在中学数学中规定了一个符号虚数 i，说是-1 的平方根，满足条件 $i^2 = -1$；学生大多还是不知道这个 i 为何物。引入复平面，将 i 解释为沿 y 轴正方向的单位向量，这可看做其几何意义，仍然没有解释清楚为什么 $i^2 = -1$。其原因在于中学课本里没有引入 i 具有旋转运动的概念。

虚数 i 就是逆时针旋转 90°。一个复数乘以 i，就是把这个复数所表示的向量矢逆时针旋转 90°。比如对 1 进行 i 连乘，序列是 $1, i, i^2, i^3, i^4, i^5, \cdots$ 那么就是把实数轴上的单位向量沿着单位圆进行逆时针旋转（见图 5-85），步进是 π/2 的弧度。



$$\begin{aligned} 1 &= 1, \\ i &= i, \\ i^2 &= -1, \\ i^3 &= -i, \\ i^4 &= 1, \\ i^5 &= i \\ &\cdots \end{aligned}$$

图 5-85 虚平面上 i 连乘就是逆时针连续旋转 π/2 弧度

好了，虚数有了直角的旋转的意义，这就容易和旋转矩阵联系起来。而在这里，方阵$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$就是 i 的一个模型，因为它也同样表示绕原点旋转直角的变换，也就是“向左转”。大家都知道“向左转再向左转，等于向后转”，也就是说i^2就是旋转 180°，就是将所有的向量乘以-1。将向左转 90° 这个变换记做 i，很自然就有$i^2=-1$。

那么如何用 i 实现旋转任意一个角度呢？

虚数 i 本身只能旋转 90°，如果要旋转一个α角度，就要虚实相济——虚数和实数在一起就可以实现：$\cos\alpha+\mathrm{i}\sin\alpha$，见图 5-86（a）。

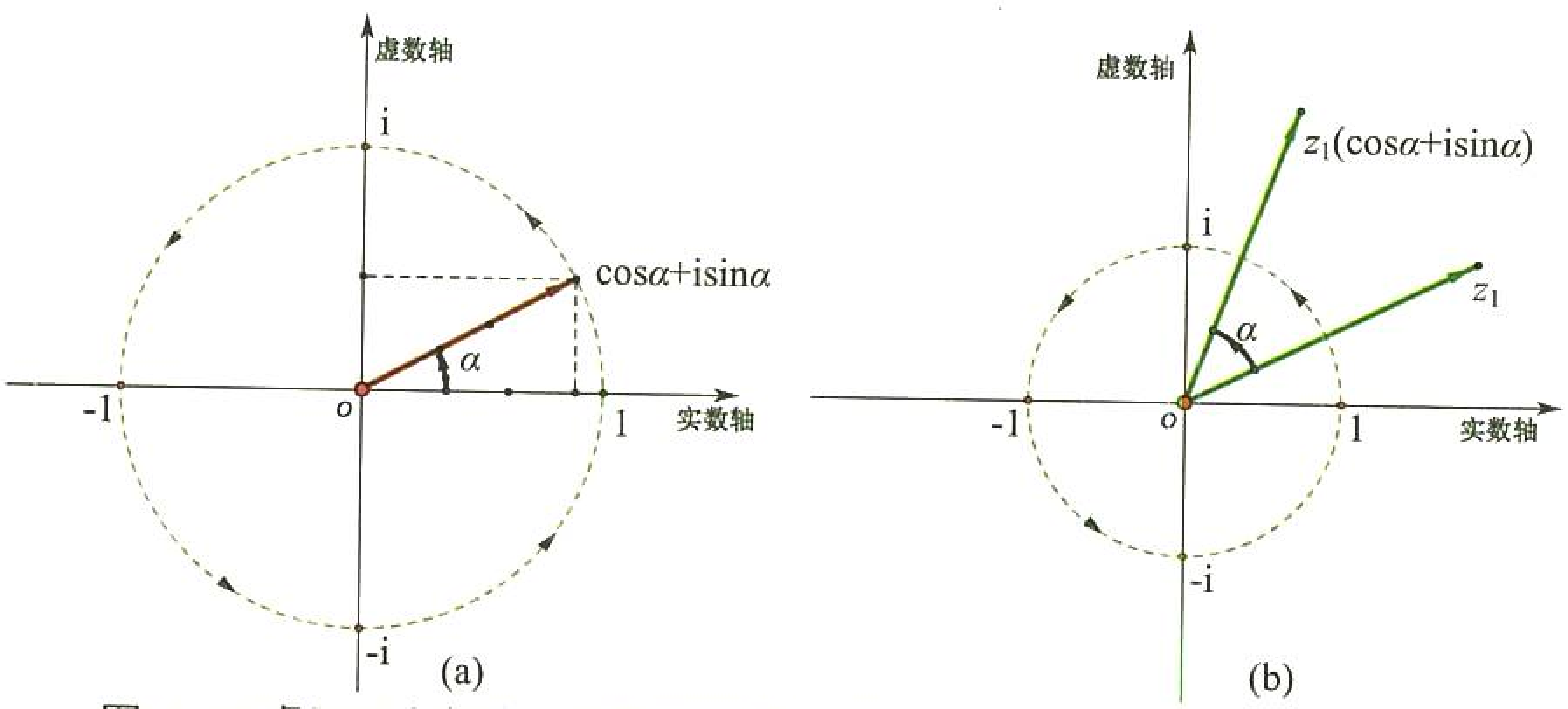


图 5-86 虚平面上任意一个复数向量z_1被$\cos\alpha+\mathrm{i}\sin\alpha$旋转$\alpha$弧度

虚平面上任意一个复数z_1乘以单位旋转复数$\cos\alpha+\mathrm{i}\sin\alpha$，就相当于把这个$z_1$逆时针旋转了一个$\alpha$角度，见图 5-86（b）。显然，复数$\cos\alpha+\mathrm{i}\sin\alpha$对应着一个可旋转$\alpha$角度的二阶矩阵：$\begin{bmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{bmatrix}$

好了，虚数 i 表示旋转 90°，复数$\cos\alpha+\mathrm{i}\sin\alpha$表示旋转$\alpha$度角，它们都有对应的矩阵。那么一个一般的复数$a+\mathrm{i}b$呢？

这个好理解，因为可以把$a+\mathrm{i}b$表示成$r(\cos\alpha+\mathrm{i}\sin\alpha)$形式，因此$a+\mathrm{i}b$可以把一个任意复数（向量）既旋转一个$\alpha$角度又同时改变其模的长度。

我们也刚好知道一个矩阵对向量的变换既有旋转也有伸缩作用。实际上，复数$a+\mathrm{i}b$对应着如矩阵$\begin{bmatrix}a & b\\ -b & a\end{bmatrix}$的形式。我们给出一个对应的表格，见表 5-8。

表 5-8　复数和矩阵运算的对应

顺序	复数运算	矩阵的运算	备注
1	$a+\mathrm{i}b$	$\begin{bmatrix}a & -b\\ b & a\end{bmatrix}$	定义的对应
2	$(a_1+\mathrm{i}b_1)\pm(a_2+\mathrm{i}b_2)$ $=(a_1\pm a_2)+\mathrm{i}(b_1\pm b_2)$	$\begin{bmatrix}a_1 & -b_1\\ b_1 & a_1\end{bmatrix}\pm\begin{bmatrix}a_2 & -b_2\\ b_2 & a_2\end{bmatrix}=\begin{bmatrix}a_1\pm a_2 & -(b_1\pm b_2)\\ b_1\pm b_2 & a_1\pm a_2\end{bmatrix}$	加减法的对应
3	$(a_1+\mathrm{i}b_1)(a_2+\mathrm{i}b_2)$ $=(a_1a_2-b_1b_2)+\mathrm{i}(a_1b_2+a_2b_1)$	$\begin{bmatrix}a_1 & -b_1\\ b_1 & a_1\end{bmatrix}\begin{bmatrix}a_2 & -b_2\\ b_2 & a_2\end{bmatrix}=\begin{bmatrix}a_1a_2-b_1b_2 & -(a_1b_2+a_2b_1)\\ a_1b_2+a_2b_1 & a_1a_2-b_1b_2\end{bmatrix}$	乘法的对应
4	共轭运算：$\overline{(a+\mathrm{i}b)}=(a-\mathrm{i}b)$	转置运算：$\begin{bmatrix}a & -b\\ b & a\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}a & b\\ -b & a\end{bmatrix}=\begin{bmatrix}a & -(-b)\\ -b & a\end{bmatrix}$	共轭对应转置
5	分解为实数和虚数的和：	分解为对称和反对称矩阵的和：	分解的对应

	$a+\mathrm{i}b=(a)+(\mathrm{i}b)$	$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}=\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}+\begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix}$	
6	模的平方：$\lvert a+\mathrm{i}b\rvert^2=a^2+b^2$	矩阵的行列式：$\begin{vmatrix} a & -b \\ b & a \end{vmatrix}=a^2+b^2$	模对应行列式
7	倒数： $(a+\mathrm{i}b)^{-1}=\dfrac{1}{a^2+b^2}(a-\mathrm{i}b)$	逆阵： $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}^{-1}=\dfrac{1}{a^2+b^2}\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ $=\dfrac{1}{a^2+b^2}\begin{bmatrix} a & -(-b) \\ -b & a \end{bmatrix}$	倒数对应逆阵

第6章 线性方程组的几何意义

线性方程组是线性代数的核心。因为在各个工程及管理领域中，问题总是被表述为一个或多个线性方程组。因此，弄清线性方程组的几何意义，深刻理解其几何本质，对于我们分析解决实践中的线性问题甚有裨益。

我们仍然从向量的角度来分析线性方程组的几何意义。比如，将二阶线性方程组用向量作两种不同的考察：

一种是常规的，即写成$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} m \\ n \end{pmatrix}$形式，解方程就是找矩阵$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$的一个逆矩阵，恰将向量$\begin{pmatrix} m \\ n \end{pmatrix}$变为$\begin{pmatrix} x \\ y \end{pmatrix}$。

另一种就是写成$x\begin{pmatrix} a \\ c \end{pmatrix}+y\begin{pmatrix} b \\ d \end{pmatrix}=\begin{pmatrix} m \\ n \end{pmatrix}$形式，解方程就是设法将$\begin{pmatrix} m \\ n \end{pmatrix}$表示成$\begin{pmatrix} a \\ c \end{pmatrix}$和$\begin{pmatrix} b \\ d \end{pmatrix}$的线性组合。

这两种看法都有明晰的几何意义，下面一一讨论。

6.1 两种线性方程组表示形式的几何意义

线性方程组就是多元一次方程组。一般形式的线性方程组如下：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{6-1}$$

这个方程组是由 m 个 n 元一次方程组成的。在解析几何的意义上，一个方程可以表示为一个直线、一个平面或一个超平面。每个方程的解都在其对应的直线、平面或超平面上，那么几个方程同时满足的解就在直线的交点、平面或超平面的交线上。

为了从向量/矩阵这个更有效率的工具解决线性方程组的问题，需要引入向量方程和矩阵方程的两种表述方式。

向量方程：从向量的角度研究线性方程组。如果对于线性方程组引入列向量：

$$\boldsymbol{a}_1=\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix},\boldsymbol{a}_2=\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix},\cdots,\boldsymbol{a}_n=\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix},\boldsymbol{b}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

那么原线性方程组可以改写为如下的向量方程组：

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$$

从向量组的线性相关性出发，向量方程求解的意义就是求出向量 $\boldsymbol{b}$ 被向量组$\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$线性表示的系数，向量方程是否有解就是向量 $\boldsymbol{b}$ 能否被向量组$\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots\boldsymbol{a}_n\}$所线性表示，线性表示系数唯一就是方程组有唯一解，不唯一则表明方程组有无穷多解。

用线性表示的几何意义来解释就是，向量$\boldsymbol{b}$如果在系数矩阵的列空间里，那么表示方程组有解。

矩阵方程：进一步从矩阵和向量的角度研究线性方程组。对于向量方程 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{b}$，分别定义矩阵$\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$，定义向量$\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$，那么线性方程组可以改写为如下的矩阵方程：

$$\boldsymbol{A}x=\boldsymbol{b}$$

由矩阵的线性变换的意义出发，解线性方程组就是寻找哪些向量可以被一个已知的矩阵A变换成一个已知的向量$\boldsymbol{b}$。方程组有唯一解就是只有一个向量被变换为向量$\boldsymbol{b}$，方程组有无穷多解就是可以有一个一根直线上（一维空间）的向量族或一个平面上（二维空间）的向量族被变换成一个向量$\boldsymbol{b}$。

我们举一个二元的例子：

对于二元一次方程组$\begin{cases}ax+by=e\\cx+dy=f\end{cases}$，可以用矩阵表示为$\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}e\\f\end{pmatrix}$。设$\boldsymbol{A}=\begin{bmatrix}a&b\\c&d\end{bmatrix}$，则前式可表示为$\boldsymbol{A}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}e\\f\end{pmatrix}$。从变换的观点看，解二元一次方程组的问题可以看成是，已知变换矩阵及变换的结果（向量（点）），求该结果是由哪一个向量（点）变过来的。如果变换$\boldsymbol{A}$可逆，只要把结果$\begin{pmatrix}e\\f\end{pmatrix}$反变回去就得到$\begin{pmatrix}x\\y\end{pmatrix}$。因此，若矩阵$\boldsymbol{A}$有逆矩阵$\boldsymbol{A}^{-1}$，则此二元一次方程组的解为$\begin{pmatrix}x\\y\end{pmatrix}=\boldsymbol{A}^{-1}\begin{pmatrix}e\\f\end{pmatrix}$。设$\boldsymbol{A}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$，则$\boldsymbol{A}$为向$x$轴作投影变换，它把直线$x=1$上的所有点变为$\begin{pmatrix}1\\0\end{pmatrix}$，因此，对于二元一次方程组$\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$，直线$x=1$上的所有点（向量）都是它的解，即有无穷多组解。而对于y轴上的点（向量）$\begin{pmatrix}0\\1\end{pmatrix}$，平面上任何一个点在投影变换$\boldsymbol{A}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$下都无法变成$\begin{pmatrix}0\\1\end{pmatrix}$，即二元一次方程组$\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0\\1\end{pmatrix}$无解。这样，我们可以直观地认识线性方程组的解的特征：有些线性方程组有唯一解，有些线性方程组有无穷多组解，有些线性方程组无解。

6.2 高斯消元法的几何解释

高斯消元法也叫矩阵消元法，是对一个线性方程组解的过程，这个过程是对方程组进行以下的三种变换：

(1) 互换两个方程的位置；

(2) 把一个的两边同乘以一个非零的常数c；

(3) 把一个方程加上另一个方程的k倍。

这些变换是可逆的和等价的，是通过消元变换把方程组化为容易求解的同解方程组。消元法的基本思想是：反复利用上述的三个初等变换，最后把方程组转换为一个阶梯形的同解方程组。

如何用几何图形来解释消元的过程呢？下面我们就从二元方程组的消元过程来看看。

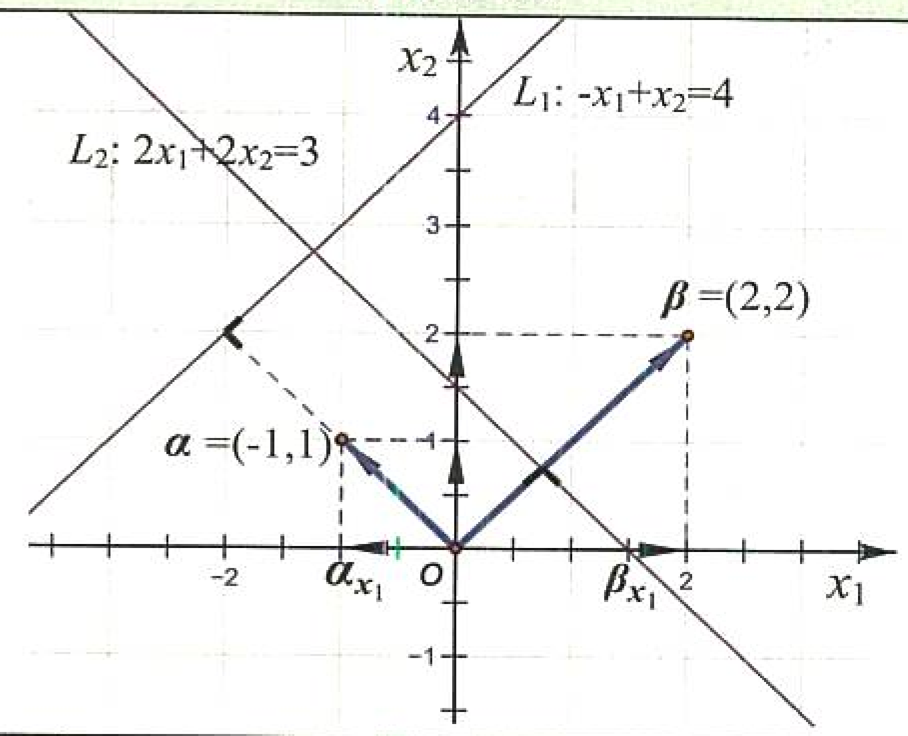
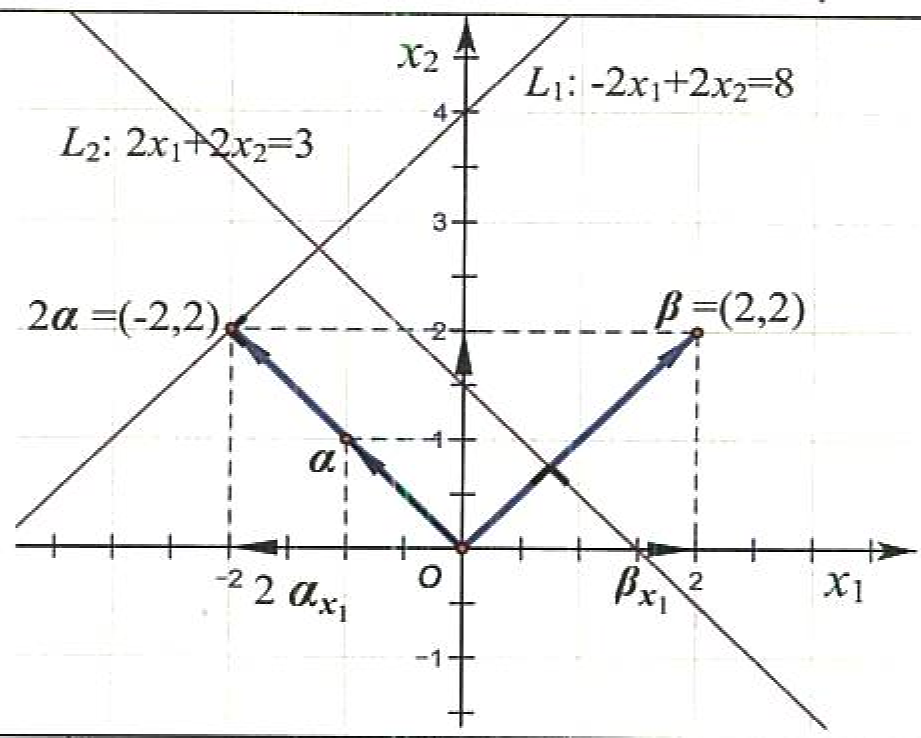
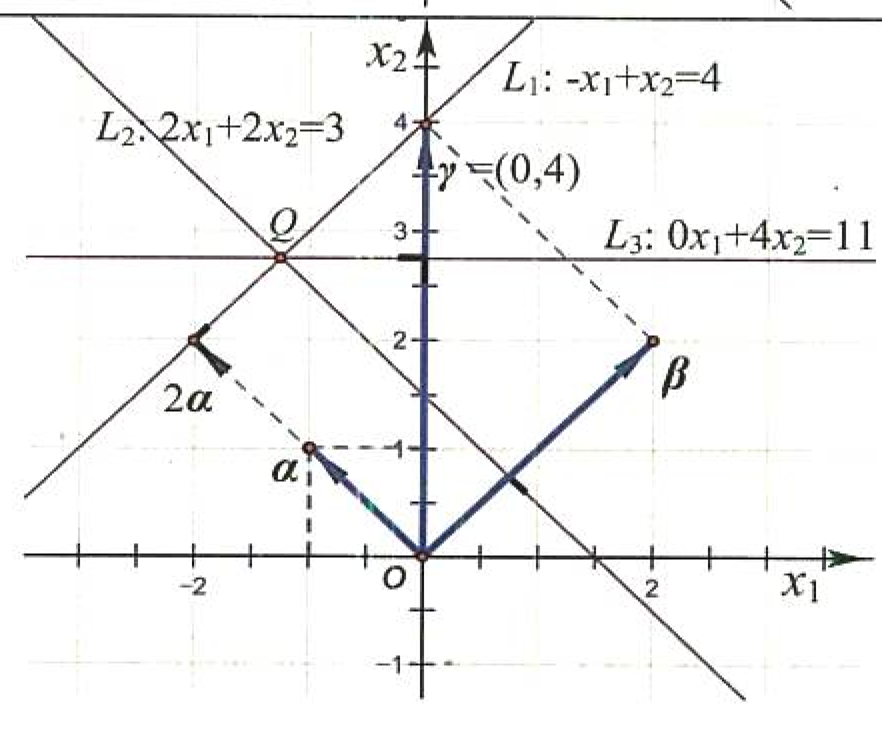
对于方程组：

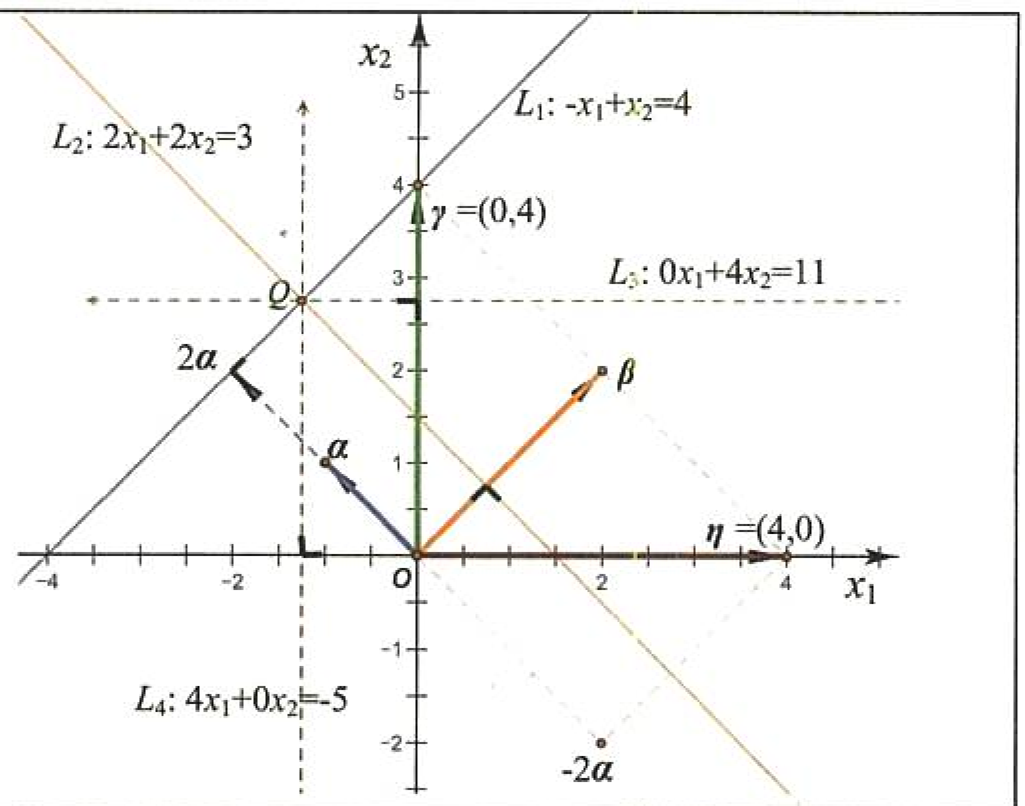
$$\begin{cases} a_1x_1 + b_1x_2 = c_1 \\ a_2x_1 + b_2x_2 = c_2 \end{cases}$$

从解析几何的知识我们知道， $a_1x_1 + b_1x_2 = c_1$ （定义为 L_1 ） $a_2x_1 + b_2x_2 = c_2$ （定义为 L_2 ）各表示一根直线，方程组的解是这两根直线的交点，这是方程组有解的几何意义。向量 (a_1, b_1) （定义为 $\boldsymbol{\alpha}$ ）和 (a_2, b_2) （定义为 $\boldsymbol{\beta}$ ）分别是 L_1 和 L_2 的法向量（法向量与其对应的直线垂直，其实法向量也是方程组系数矩阵的行向量）。在法向量的帮助下，我们讨论消元过程的一个例子：

对方程组 $\begin{cases} -x_1 + x_2 = 4 & (L_1) \\ 2x_1 + 2x_2 = 3 & (L_2) \end{cases}$ 进行消元法求解，过程详见表 6-1。

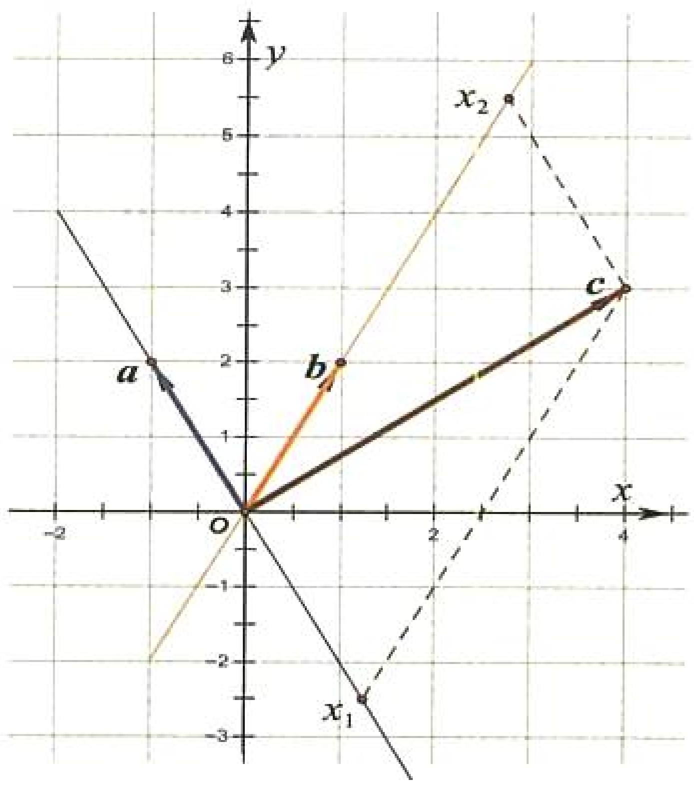
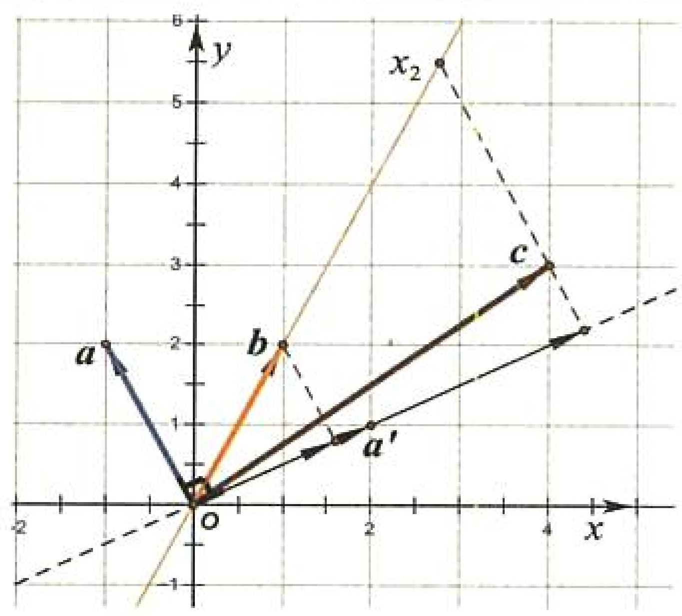
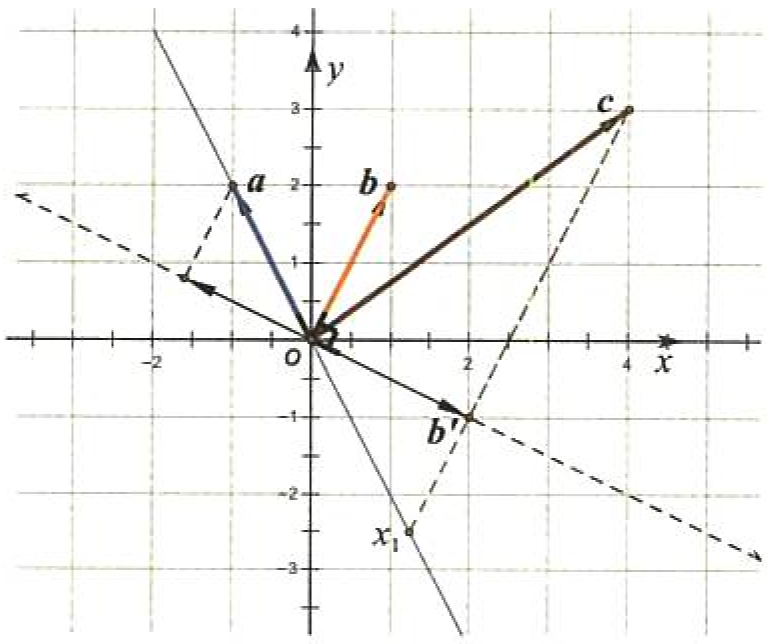
表 6-1 从行向量角度看二元方程组消元法的图解

顺序	方程组的消元过程及行向量解释	几何图解
1	$\begin{cases} -x_1 + x_2 = 4 & (L_1) \\ 2x_1 + 2x_2 = 3 & (L_2) \end{cases}$ 右图画出了系数矩阵的行向量的图形，即直线的法向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$	
2	L_1 乘以 2： $\begin{cases} -2x_1 + 2x_2 = 8 & (2L_1) \\ 2x_1 + 2x_2 = 3 & (L_2) \end{cases}$ 右图显示了 L_1 的法向量 $\boldsymbol{\alpha}$ 沿着原方向长度扩大到 2 倍，但直线图形没有变化，为了区别，我们用 2 L_1 表示。 注意：2 $\boldsymbol{\alpha}$ 的 x_1 轴的分量 2 $\boldsymbol{\alpha}_{x1}$ 与 $\boldsymbol{\beta}$ 的 x_1 轴的分量 $\boldsymbol{\beta}_{x1}$ 方向相反，长度变得一样	
3	L_1 +2 L_1 得到 L_3 ： $0x_1 + 4x_2 = 11 \quad (L_3)$ L_3 是消去了 x_1 变量后的直线，因此与 x_1 轴平行。 L_3 的法向量 $\boldsymbol{\gamma}$ 是 $2\boldsymbol{\alpha} + \boldsymbol{\beta}$ 的结果，**向量相加对消了 x_1 方向上的分量**	

4	L_1 -2 L_1 得到 L_4： $4x_1+0x_2=-5\quad(L_4)$ L_4 是消去了 x_2 变量后的直线，因此与 x_2 轴平行。 L_4 的法向量 $\boldsymbol{\eta}$ 是 $\boldsymbol{\beta}-2\boldsymbol{\alpha}$ 的结果，**向量相减 对消了 x_2 方向上的分量**	

同样，对方程组 $\begin{cases}-x_1+x_2=4 & (L_1)\\ 2x_1+2x_2=3 & (L_2)\end{cases}$ 进行消元法求解时，我们从列向量的角度考察消元过程，重写方程组为向量方程：$\begin{pmatrix}-1\\2\end{pmatrix}x_1+\begin{pmatrix}1\\2\end{pmatrix}x_2=\begin{pmatrix}4\\3\end{pmatrix}$，为方便，定义向量 $\boldsymbol{a}=\begin{pmatrix}-1\\2\end{pmatrix}$，$\boldsymbol{b}=\begin{pmatrix}1\\2\end{pmatrix}$，$\boldsymbol{c}=\begin{pmatrix}4\\3\end{pmatrix}$，变换过程详见表 6-2。

表 6-2 从列向量角度看二元方程组消元法的图解

顺序	方程组的消元过程及列向量解释	几何图解
1	$\begin{pmatrix}-1\\2\end{pmatrix}x_1+\begin{pmatrix}1\\2\end{pmatrix}x_2=\begin{pmatrix}4\\3\end{pmatrix}$ 等同于向量方程： $\boldsymbol{a}x_1+\boldsymbol{b}x_2=\boldsymbol{c}$ 右图画出了系数矩阵的列向量 $\boldsymbol{a}$、$\boldsymbol{b}$、$\boldsymbol{c}$ 的图形。如何找到由向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 所线性表示的向量 $\boldsymbol{c}$ 的系数 x_1 和 x_2 呢？ 注意：图中的 x_1、x_2 分别是以向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 的长度和方向为基本单位得到的倍数	
2	对于方程 $\boldsymbol{a}x_1+\boldsymbol{b}x_2=\boldsymbol{c}$ 两边，如果同时点乘向量 $\boldsymbol{a}$ 的正交向量 $\boldsymbol{a}'$（构造出来），就可以消掉 $\boldsymbol{a}x_1$ 项，x_1 的未知元也就消去了，即得到 $\boldsymbol{a}'\cdot\boldsymbol{b}x_2=\boldsymbol{a}'\cdot\boldsymbol{c}$，进而有 $x_2=\dfrac{\boldsymbol{a}'\cdot\boldsymbol{c}}{\boldsymbol{a}'\cdot\boldsymbol{b}}$，其中 $\boldsymbol{a}'\cdot\boldsymbol{b}\neq 0$	
3	类似地，对于方程 $\boldsymbol{a}x_1+\boldsymbol{b}x_2=\boldsymbol{c}$ 两边，如果同时点乘向量 $\boldsymbol{b}$ 的正交向量 $\boldsymbol{b}'$（构造出来），就可以消掉 $\boldsymbol{b}x_2$ 项，x_2 的未知元也就消去了，即有 $\boldsymbol{b}'\cdot\boldsymbol{a}x_1=\boldsymbol{b}'\cdot\boldsymbol{c}$，进而有 $x_1=\dfrac{\boldsymbol{b}'\cdot\boldsymbol{c}}{\boldsymbol{b}'\cdot\boldsymbol{a}}$，其中 $\boldsymbol{b}'\cdot\boldsymbol{a}\neq 0$	

为了更清晰地看到向量消元法和经典的系数高斯消元法的联系，我们给出一般的二元一次实系数线性方程组的形式：

$$\begin{cases} a_1x_1+b_1x_2=c_1 \\ a_2x_1+b_2x_2=c_2 \end{cases}$$

将其写做如下等价的向量方程形式：

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}x_1+\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}x_2=\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

为了方便，定义 $\boldsymbol{a}=\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$，$\boldsymbol{b}=\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$，$\boldsymbol{c}=\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$，则向量方程简写为 $\boldsymbol{a}x_1+\boldsymbol{b}x_2=\boldsymbol{c}$。

我们也用表 6-3 对比一下方程组解法和向量解法的差异，进而看出其几何意义。

表 6-3 方程组经典消元法和矩阵消元法的比对

顺序	方程组的消元过程	向量方程的消元过程
1	$\begin{cases} a_1x_1+b_1x_2=c_1 & (1) \\ a_2x_1+b_2x_2=c_2 & (2) \end{cases}$ 对方程组进行以下运算： $b_2\times(1)-b_1\times(2)$，消掉 x_2，得方程（3）	$\underbrace{\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}_{\boldsymbol{a}}x_1+\underbrace{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}_{\boldsymbol{b}}x_2=\underbrace{\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\boldsymbol{c}}$　(1) 构造 $\boldsymbol{b}'=(b_2,-b_1)$（$\boldsymbol{b}'$ 垂直于 $\boldsymbol{b}$），等式(1)两边右点乘 $\boldsymbol{b}'$，消掉 $\boldsymbol{b}x_2$ 项，得式（2）
2	$(b_2a_1-b_1a_2)x_1=b_2c_1-b_1c_2$　（3） 如果 $b_2a_1-b_1a_2\neq 0$，得方程(4)	$\underbrace{(b_2,-b_1)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}_{\boldsymbol{b}'\cdot\boldsymbol{a}}x_1=\underbrace{(b_2,-b_1)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}}_{\boldsymbol{b}'\cdot\boldsymbol{c}}$　（2） 如果 $\boldsymbol{b}'\cdot\boldsymbol{a}\neq\boldsymbol{0}$，得式（3）
3	$x_1=\dfrac{b_2c_1-b_1c_2}{b_2a_1-b_1a_2}$　（4）	$x_1=\dfrac{\boldsymbol{b}'\cdot\boldsymbol{c}}{\boldsymbol{b}'\cdot\boldsymbol{a}}$　（3）
4	同理，对方程组进行以下运算： $(1)\times a_2-(2)\times a_1$，消掉 x_2，得方程(5) $x_2=\dfrac{a_2c_1-a_1c_1}{a_2b_1-a_1b_2}$　(5)	同理，构造 $\boldsymbol{a}'=(a_2,-a_1)$（$\boldsymbol{a}'$ 垂直于 $\boldsymbol{a}$），等式 (1) 两边右点乘 $\boldsymbol{a}'$，消掉 $\boldsymbol{a}x_1$ 项，得式(4)： $x_2=\dfrac{\boldsymbol{a}'\cdot\boldsymbol{c}}{\boldsymbol{a}'\cdot\boldsymbol{b}}$　(4)

我们可以从上面的例图可以看出：消元法的本质是构造一个个正交向量逐步消去变元的过程。三元乃至多元的线性方程组也是完全同样的情况。

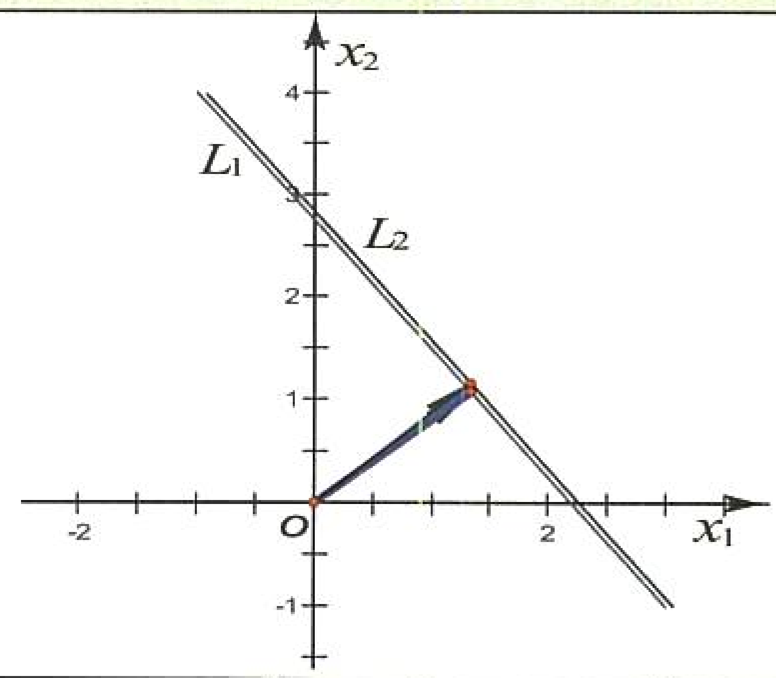
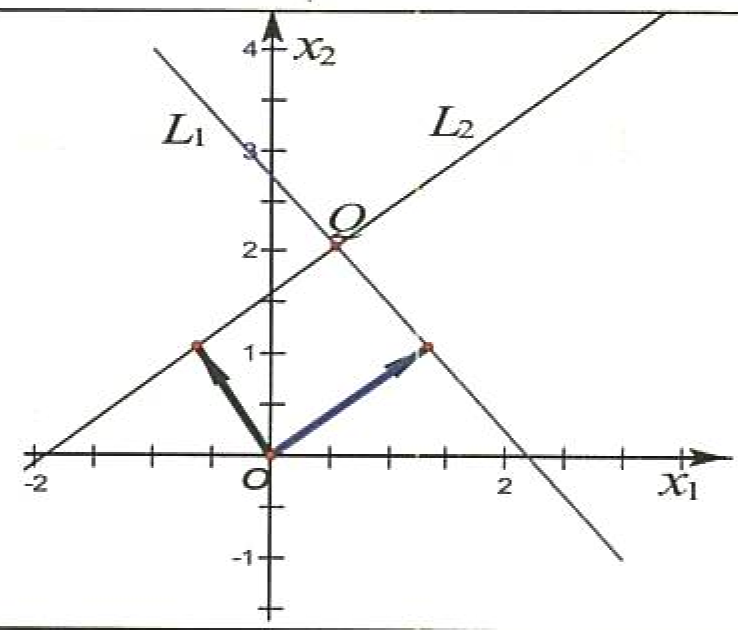
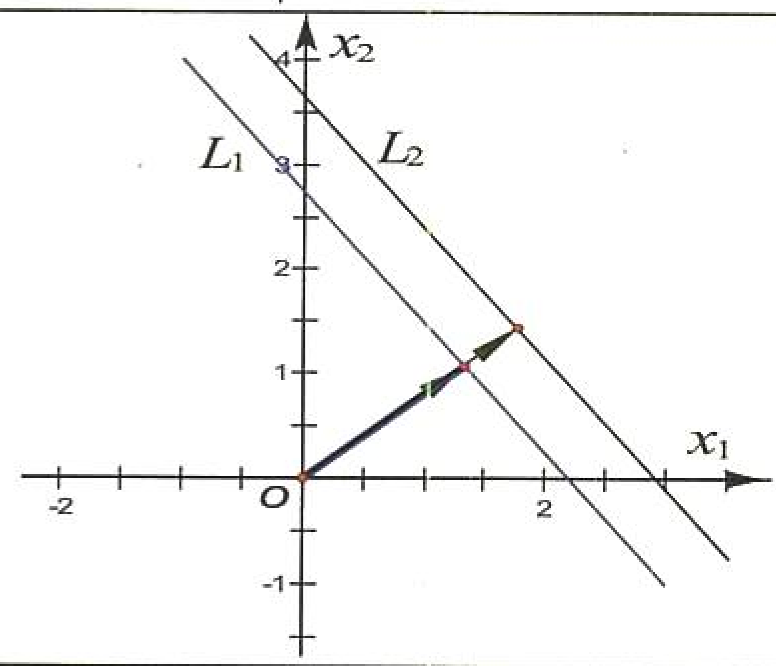
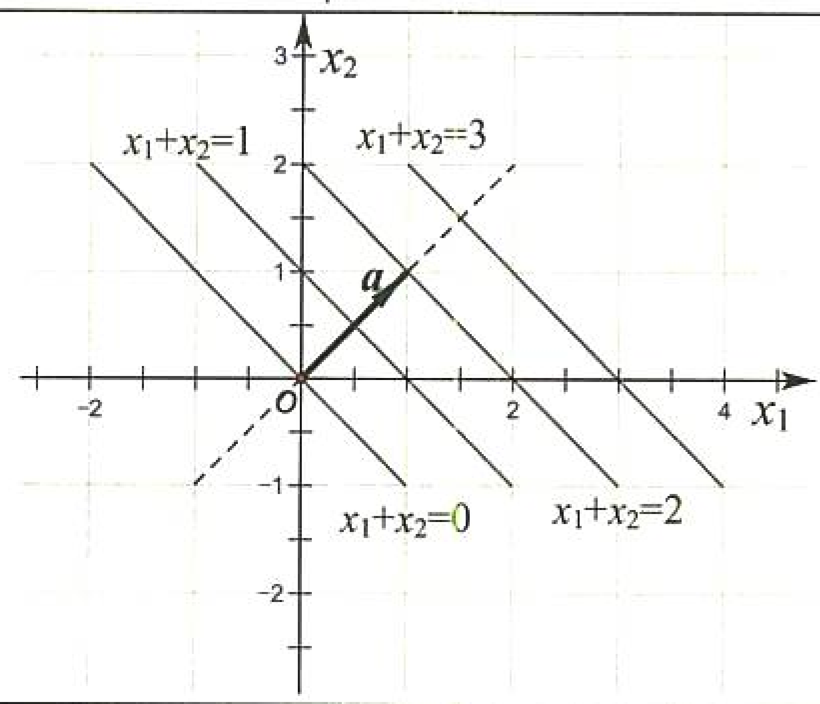
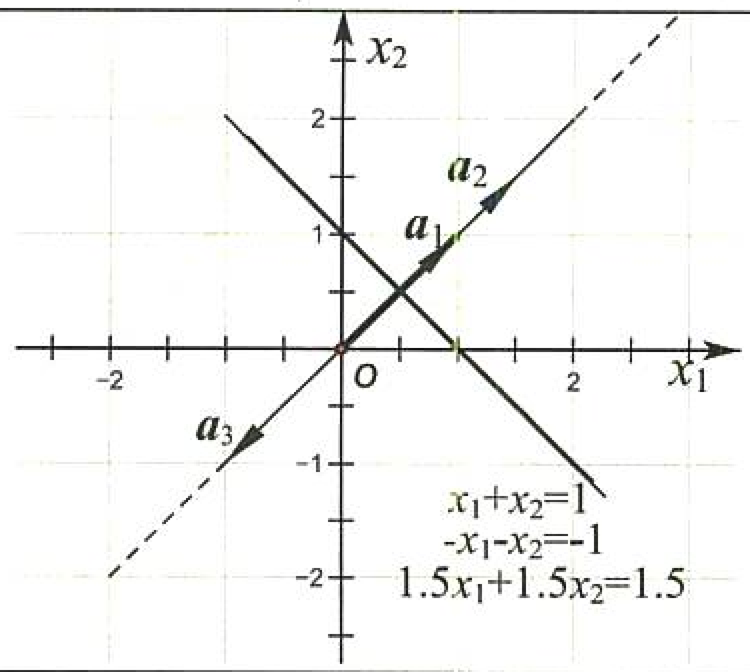
6.3 线性方程组的秩及解的关系的几何意义

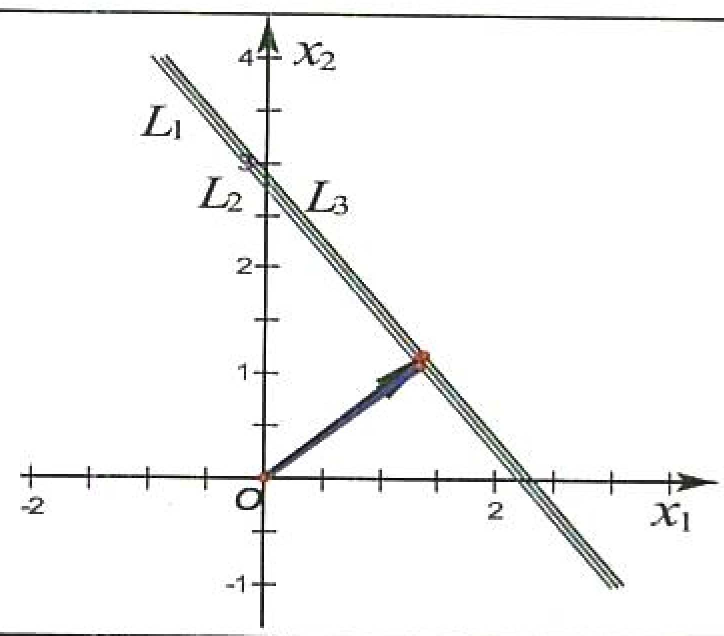
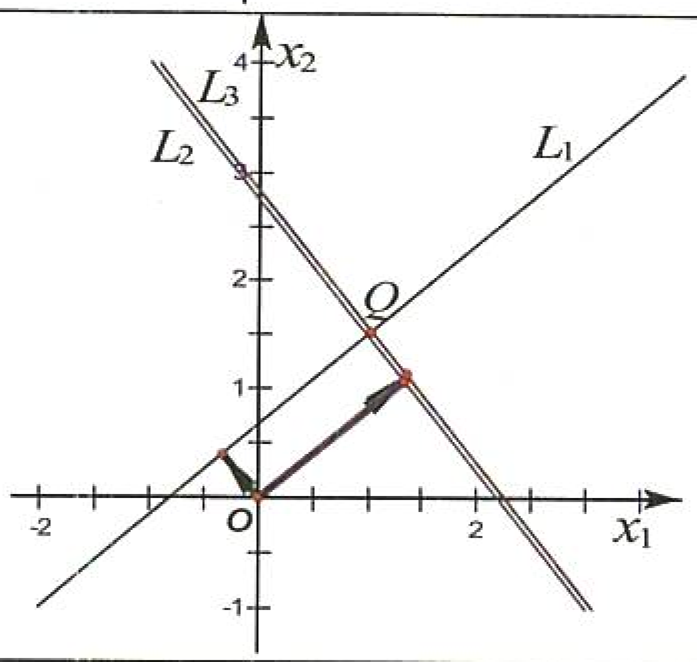
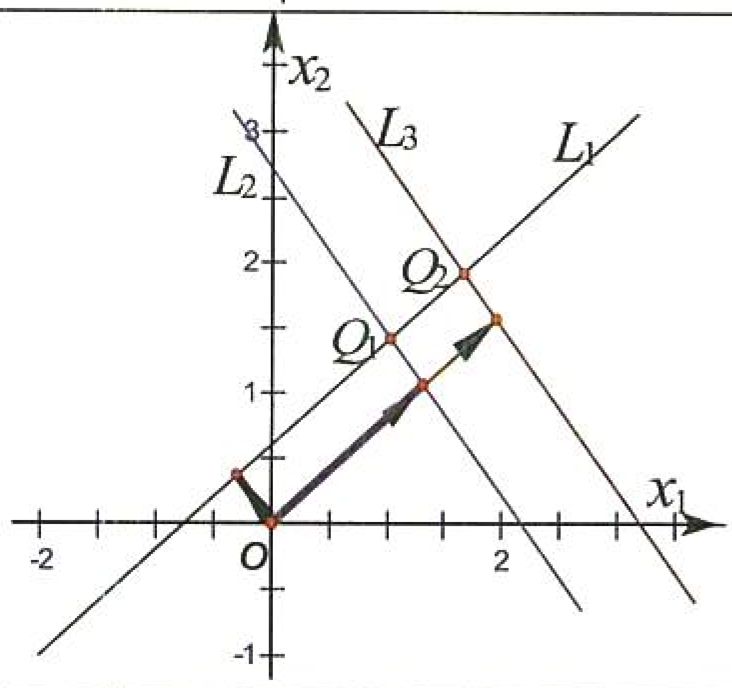
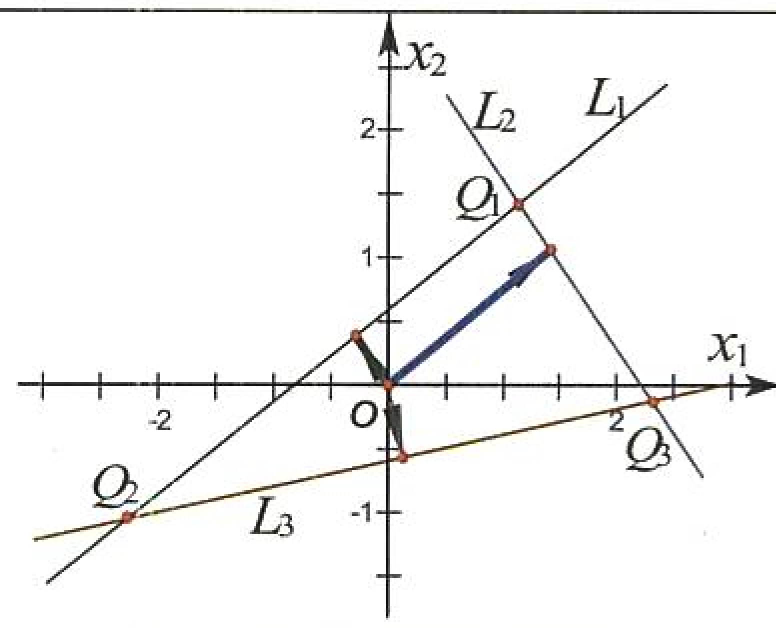
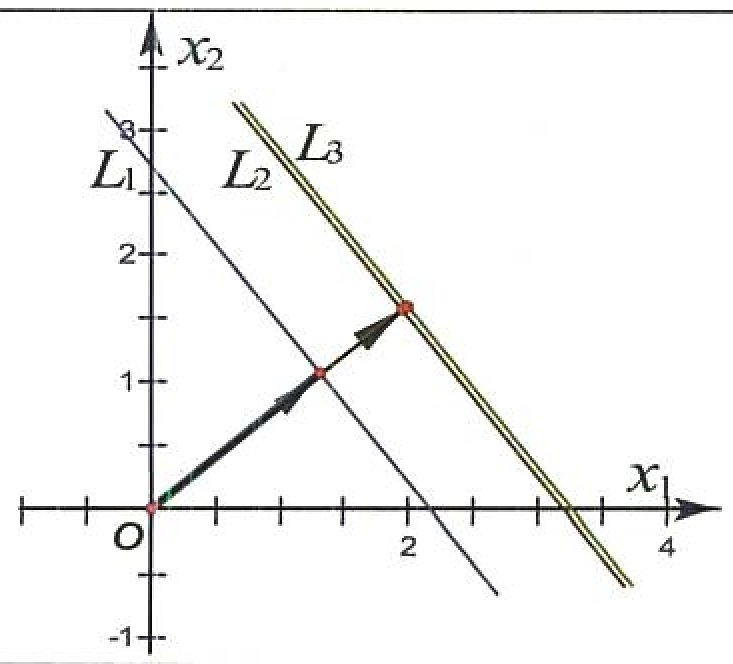
讨论有关线性方程组解的不同情况是线性代数中的重要内容，直线的位置关系和平面的位置关系为线性方程组解的不同情况提供了直观背景，而线性方程组解的相关理论促使我们更深刻领会直线的位置关系和平面的位置关系。秩的理论完整地刻画了线性方程组解的情况。

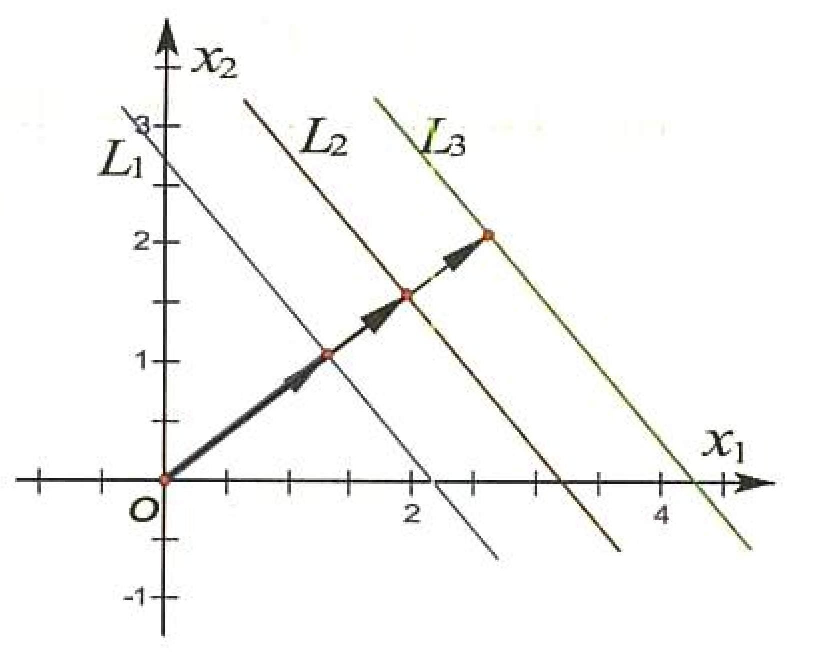
6.3.1 二元线性方程组的秩及解的图形

我们知道，如果 a、b 不全为零，一个有两个变量的线性方程 $ax_1+bx_2=c$ 的解 (x_1,x_2) 在平面 $\mathbf{R}^2$ 上构成一条直线。那么不同个数的二元线性方程组的秩和解的意义可以有以下的解析几何的解释。表 6-4 汇总了所有的类型。

表 6-4 二元方程组的秩与法向量的关系

顺序	二元线性方程组的类型	秩的不同情况及说明	对应的几何图形
1.1	$\begin{cases} a_1x_1+b_1x_2=c_1 & (L_1) \\ a_2x_1+b_2x_2=c_2 & (L_2) \end{cases}$ 定义：系数矩阵 $A=\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$，增广矩阵 $\overline{A}=\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$， L_1 的法向量取 (a_1,b_1)， L_2 的法向量取 (a_2,b_2)	$r(A)=r(\overline{A})=1$。化简后，法向量有 1 个（原法向量重合或相关），方程有 1 个。方程组有无穷多解；L_1 和 L_2 重合，即 L_1 和 L_2 有无穷多个交点，这些交点都是方程组的解	
1.2		$r(A)=r(\overline{A})=2$。化简后，法向量有 2 个，方程有 2 个。方程组有唯一解；L_1 和 L_2 相交，有一个交点 Q，Q 点就是方程组的解	
1.3		$r(A)=1,\ r(\overline{A})=2$。化简后，法向量只有 1 个（原法向量重合或相关），方程有 2 个。因为无法化简（如果化简会出现矛盾等式），方程组没有解；L_1 和 L_2 平行，没有交点	
1.4	**总结：** （1）一个法向量可以有无穷多的平行直线与之对应，如图所示，图中所有的直线只有一个法向量 $\boldsymbol{a}=(1,\ 1)$，这些直线都是平行的。如果由这些直线方程组成方程组，化简后法向量只有一个（原法向量重合），整个方程化简后出现矛盾等式，可以认为无法化简，此方程组无解		
	（2）反过来，一根直线图形也有无数个法向量（与之垂直的向量都是法向量），这些向量线性两两线性相关，分布在一条垂直线上，如图所示。因为一个直线图形可以表示为等比例的无穷多方程式，因而有无穷多法向量。**但是，如果由这些直线方程组成方程组，方程组化简后只剩下一个方程，因而只有（剩下的那个方程）一个法向量。这些方程的图形重合为一根直线**		

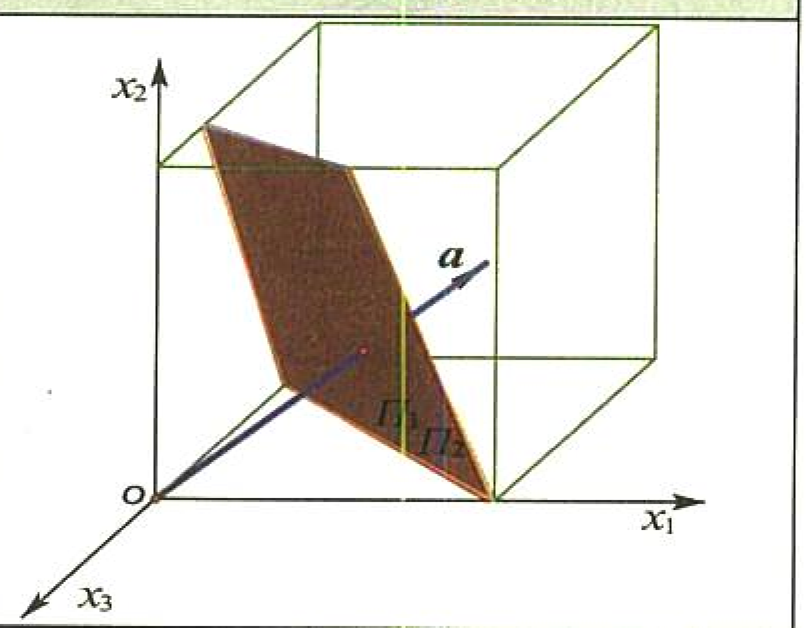
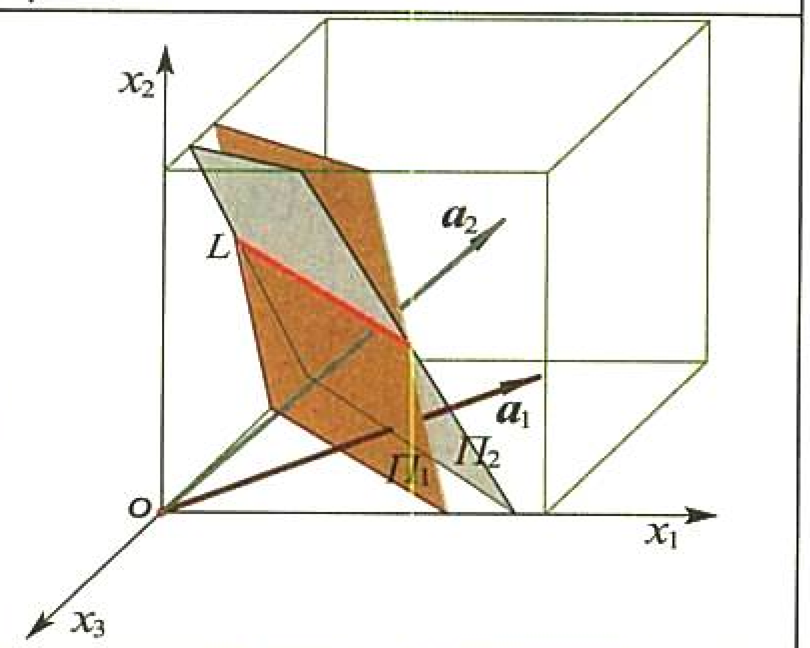
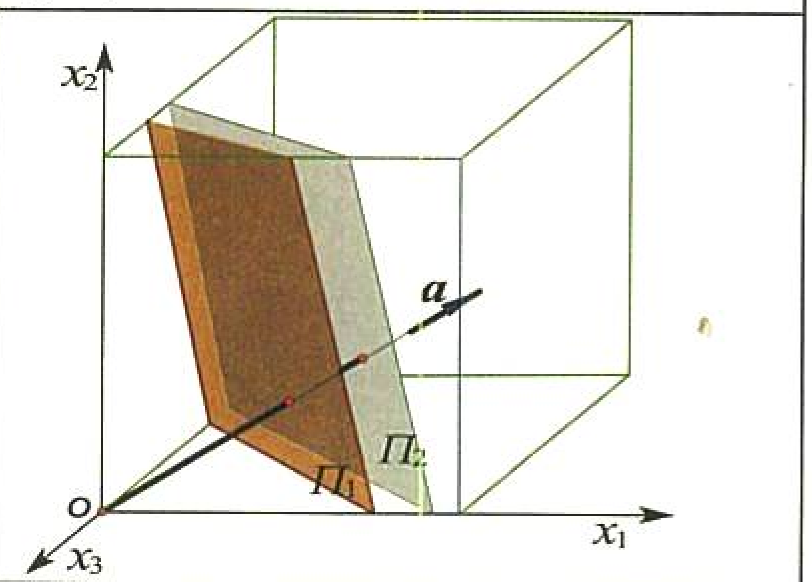
	注意：**为了方便说事，咱约定一根直线对应一个法向量；最省事的做法是取由这根直线方程式确定的法向量为这根直线的唯一对应法向量，如** $x_1+x_2=1$ 可指定唯一法向量是 $\boldsymbol{a}=(1,\ 1)$		
2.1	$\begin{cases} a_1x_1+b_1x_2=c_1 & (L_1) \\ a_2x_1+b_2x_2=c_2 & (L_2) \\ a_3x_1+b_3x_2=c_3 & (L_3) \end{cases}$ 定义： 系数矩阵 $\boldsymbol{A}=\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}$， 增广矩阵 $\overline{\boldsymbol{A}}=\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$， L_1 的法向量是 (a_1,b_1)， L_2 的法向量是 (a_2,b_2)， L_3 的法向量是 (a_3,b_3)	$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=1$。化简后，法向量有 1 个（重合），方程有 1 个。方程组无穷多解；L_1、L_2 和 L_3 重合，即 L_1、L_2 和 L_3 有无穷多个交点，这些交点都是方程组的解	
2.2		$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=2$。化简后，法向量仍有 2 个(其中 2 个重合为一个)，方程变为 2 个。方程组有唯一解；L_2 和 L_3 重合后与 L_1 相交（其中的一个可能），有一个公共交点 Q，Q 点就是方程组的解	
2.3		$r(\boldsymbol{A})=2,r(\overline{\boldsymbol{A}})=3$。化简后，法向量有 2 个，方程有 3 个。因为整个方程无法化简或化简会出现矛盾等式，故方程组无解；三根直线没有一个公共交点。直线间关系有两个交点 Q_1、Q_2 和三个交点 Q_1、Q_2 和 Q_3 两种情况	
		注：右图中，直线有三根，法向量也有 3 个，但其中一个向量可以被其他两个线性表示，因此在化简时被消减掉一个	
2.4		$r(\boldsymbol{A})=1,r(\overline{\boldsymbol{A}})=2$。化简后，法向量有 1 个（原法向量线性相关或者重合），方程有 2 个。因为方程化简会出现矛盾等式，故方程组无解；两根直线重合并与第三根直线平行，没有一个公共交点	

2.5		$r(A)=1, r(\bar{A})=3$。化简后，法向量有 1 个（原法向量彼此线性相关），方程有 3 个。因为整个方程无法化简或化简会出现矛盾等式，故方程组无解；三根直线平行，没有一个公共交点 **注**：$r(A)=1, r(\bar{A})=3$ 还可以进一步化简为 $r(A)=1, r(\bar{A})=2$，这两种情况可以归为一类	
2.6	前面的分类很是繁杂，不过有个对应关系可以帮助我们化繁就简： $r(A)=m, r(\bar{A})=n$：m 就是化简后法向量的个数，n 就是化简后的方程的个数（注意：出现矛盾的等式表示此方程不能化简）		

6.3.2 三元线性方程组的秩及解的图形

同二元线性方程组的情况类似，如果 a_1、a_2、a_3 不全为零，一个三元线性方程 $a_1x_1+a_2x_2+a_3x_3=b$ 的全部解 (x_1,x_2,x_3) 可以构成为空间 $\mathbf{R}^3$ 中的一个平面。方程组的解决定于平面的相交情况及其法向量的关系。表 6-5 给出了主要方程组类型及其图解。

表 6-5 三元方程组的秩与法向量的关系

顺序	三元线性方程组类型	秩的不同情况及说明	对应的几何图形
1.1	$\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3=b_1\ (\Pi_1) \\ a_{21}x_1+a_{22}x_2+a_{23}x_3=b_2\ (\Pi_2) \end{cases}$ 定义：系数矩阵 $\boldsymbol{A}$ $=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$， 增广矩阵 $\bar{\boldsymbol{A}}$ $=\begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \end{pmatrix}$， Π_1 的法向量是 $\boldsymbol{a}_1=(a_{11}, a_{12}, a_{13})$， Π_2 的法向量是 $\boldsymbol{a}_2=(a_{21}, a_{22}, a_{23})$	$r(A)=r(\bar{A})=1$。化简后，法向量只有 1 个（原法向量重合或相关）；方程有 1 个。原平面 Π_1 和 Π_2 重合，即 Π_1 和 Π_2 有无穷多个交点，这些交点构成一个平面，方程组有无穷多解。平面上的点都是方程组的解	
1.2		$r(A)=r(\bar{A})=2$。化简后，法向量有 2 个，方程有 2 个。平面 Π_1 和 Π_2 相交成一条直线 L，即 Π_1 和 Π_2 有无穷多个交点，这些交点都是方程组的解，因此方程组有无穷多解	
1.3		$r(A)=1, r(\bar{A})=2$。化简后，法向量有 1 个（原两个法向量线性相关），方程有 2 个。平面 Π_1 和 Π_2 平行，没有交点，因此方程组无解	

<table>
<tr>
<td>2.1</td>
<td rowspan="4">

$$\begin{cases}a_{11}x_1+a_{12}x_2+a_{13}x_3=b_1\ (\Pi_1)\\ a_{21}x_1+a_{22}x_2+a_{23}x_3=b_2\ (\Pi_2)\\ a_{31}x_1+a_{32}x_2+a_{33}x_3=b_3\ (\Pi_3)\end{cases}$$

定义：系数矩阵 $\boldsymbol{A}$

$$=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix},$$

增广矩阵 $\overline{\boldsymbol{A}}$

$$=\begin{pmatrix}a_{11}&a_{12}&a_{13}&b_1\\a_{21}&a_{22}&a_{23}&b_2\\a_{31}&a_{32}&a_{33}&b_3\end{pmatrix},$$

Π_1 的法向量是 $\boldsymbol{a}_1$

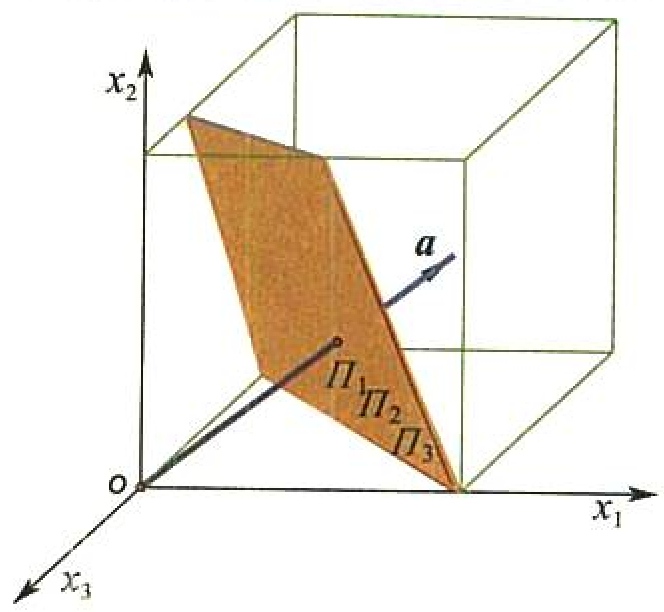
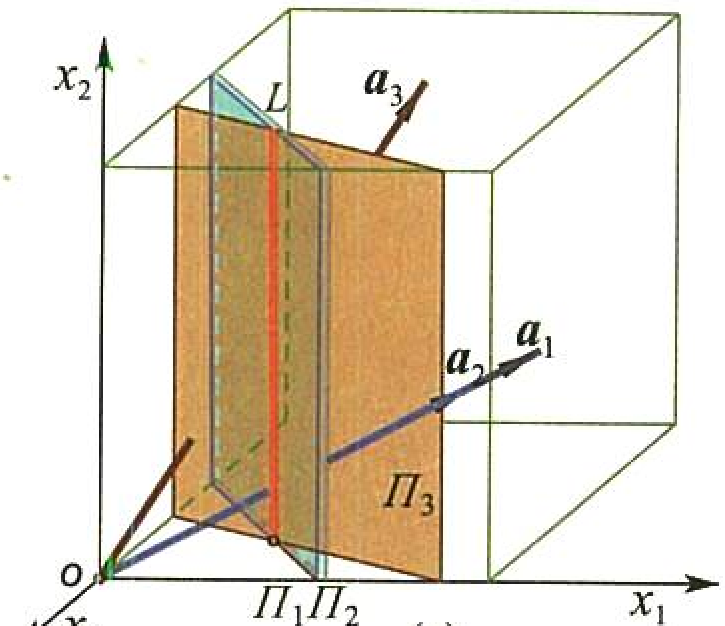
$=(a_{11},a_{12},a_{13})$，

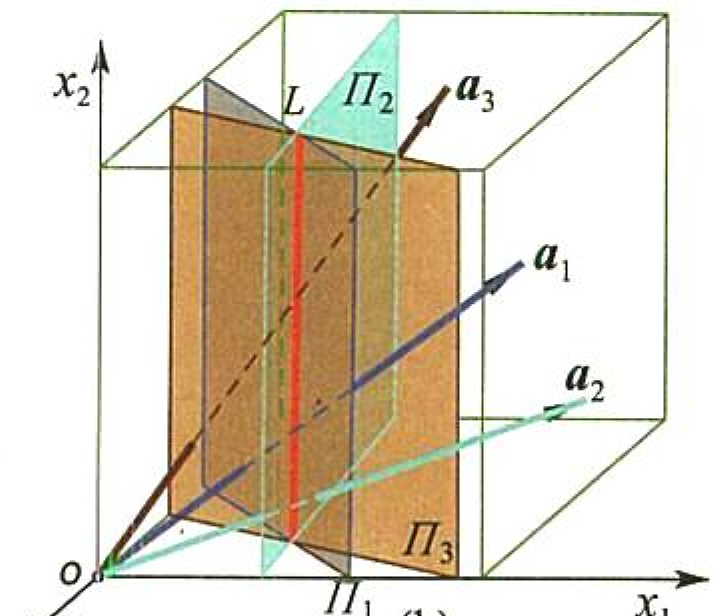
Π_2 的法向量是 $\boldsymbol{a}_2$

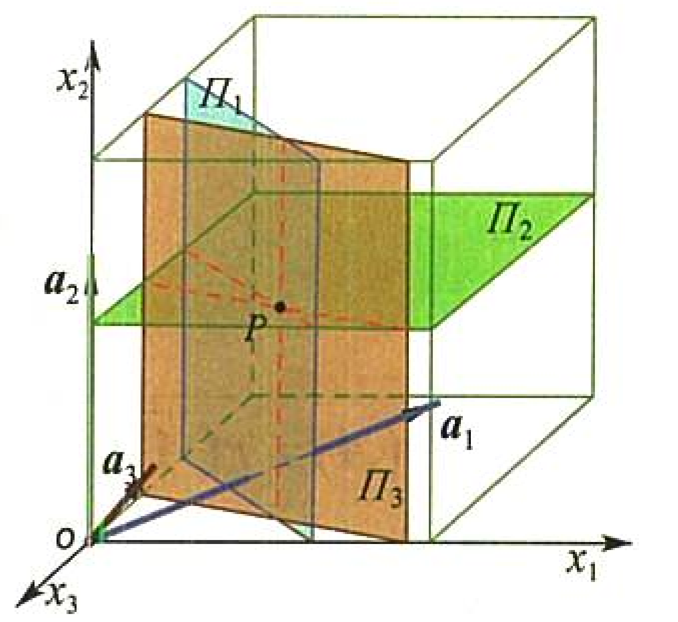
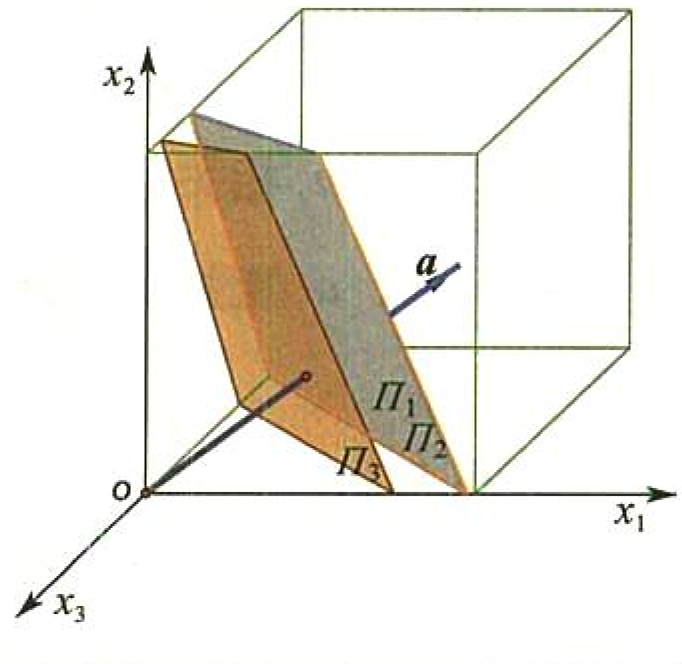
$=(a_{21},a_{22},a_{23})$，

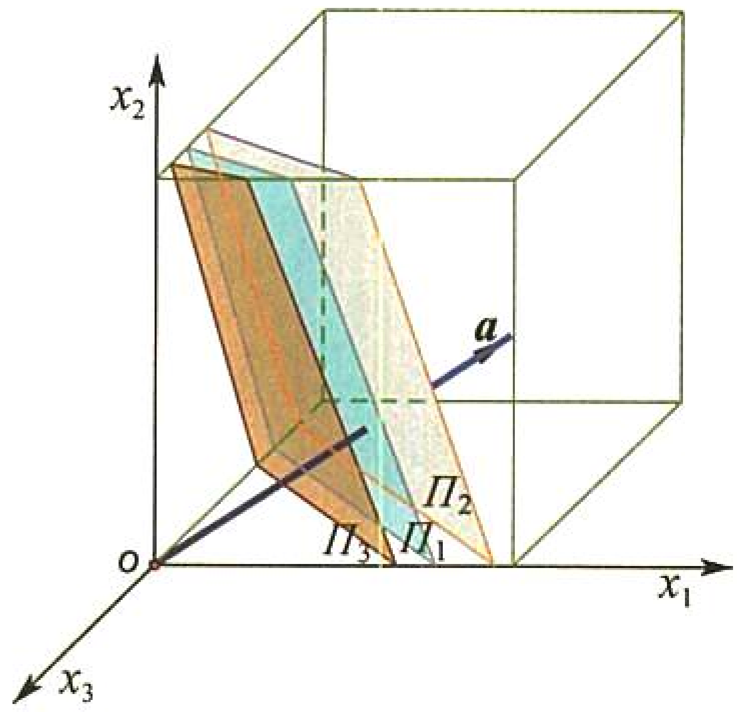
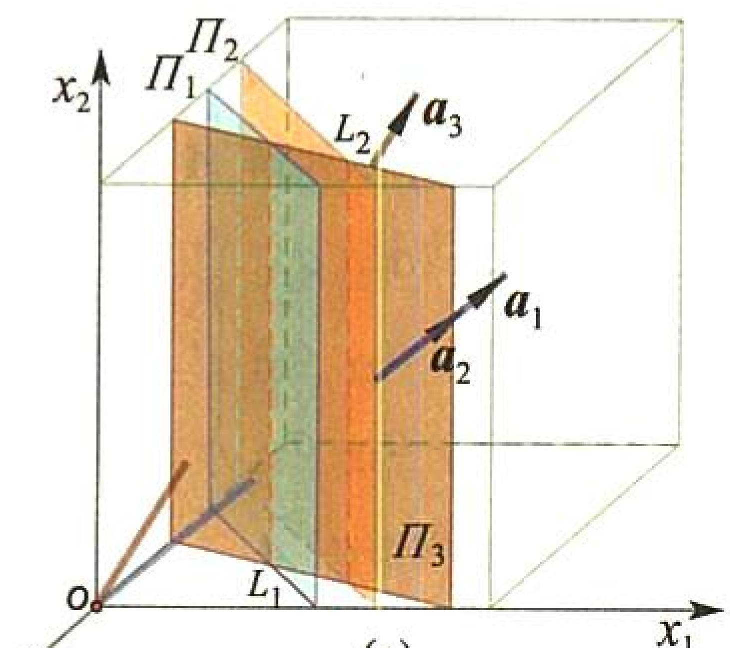
Π_3 的法向量是 $\boldsymbol{a}_3$

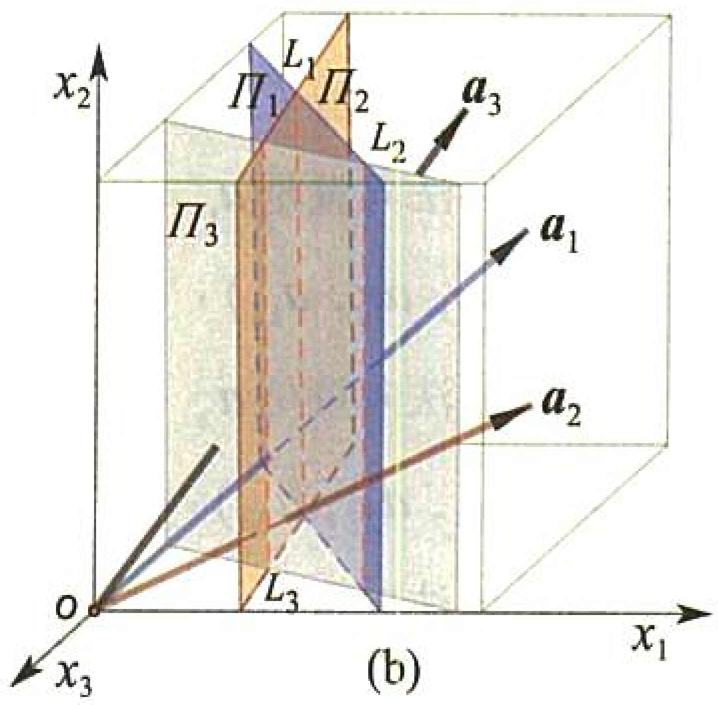
$=(a_{31},a_{32},a_{33})$

</td>
<td>$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=1$。化简后，法向量只有 1 个（原法向量重合或相关），方程有 1 个。原平面 Π_1、Π_2 和 Π_3 重合，即三个平面有无穷多个交点，这些交点构成一个平面，方程组有无穷多解。平面上的点都是方程组的解。</td>
<td>
</td>
</tr>
<tr>
<td>2.2</td>
<td>$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=2$。化简后，法向量有 2 个，方程有 2 个。原平面 Π_1、Π_2 和 Π_3 相交成一条直线 L，有无穷多个交点，这些交点都是方程组的解，因此方程组有无穷多解。
相交情况有两种：
一种是如图（a）所示，Π_1 和 Π_2 平面重合并与 Π_3 相交于 L。
第二种是如图（b）所示，三个平面互不重合或平行地相交于一条直线 L 上。
注：图（b）中绘出了三个法向量，其实，这三个法向量线性相关，第三个法向量可被另外两个法向量线性表示。或者 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$、$\boldsymbol{a}_3$ 在一个平面上，或者有两个向量线性相关，如图（a）中的 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 线性相关</td>
<td>

(a)


(b)</td>
</tr>
<tr>
<td>2.3</td>
<td>$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=3$。化简后，法向量有 3 个，方程有 3 个。原平面 Π_1、Π_2 和 Π_3 相交成空间一点 P，这个交点就是方程组的解，因此方程组有唯一解。
三个法向量不相关</td>
<td>
</td>
</tr>
<tr>
<td>2.4</td>
<td>$r(\boldsymbol{A})=1,r(\overline{\boldsymbol{A}})=2$。化简后，法向量有 1 个 $\boldsymbol{a}$，方程有 2 个；方程无解。法向量是 1，说明三个平面有平行或重合的关系；$r(\overline{\boldsymbol{A}})$ 为 2，说明化简后只有两个有效的平面，因而有两个平面是重合的。
图中设平面 Π_1、Π_2 相重合。
三个原平面无公共的交点、交线或交面，因而无解</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>2.5</td>
<td></td>
<td>$r(\boldsymbol{A})=1, r(\overline{\boldsymbol{A}})=3$。化简后，法向量有 1 个 $\boldsymbol{a}$，方程有 3 个；方程无解。法向量是 1，说明三个平面有平行或重合的关系；$r(\overline{\boldsymbol{A}})$ 为 3，说明化简后只有三个有效的平面，因而三个平面是平行的。
三个原平面无公共的交点、交线或交面，因而无解。
特别注意：$r(\boldsymbol{A})=1, r(\overline{\boldsymbol{A}})=3$ 的情形一般被化简成了 $r(\boldsymbol{A})=1$，$r(\overline{\boldsymbol{A}})=2$。因此我们可以把三平面互相平行的情形归入 $r(\boldsymbol{A})=1, r(\overline{\boldsymbol{A}})=2$ 类中。</td>
<td>
</td>
</tr>
<tr>
<td>2.6</td>
<td></td>
<td>$r(\boldsymbol{A})=2, r(\overline{\boldsymbol{A}})=3$。化简后，法向量有 2 个，方程有 3 个；方程无解。法向量是 2，说明有一个法向量可以被其他两个法向量线性表示，有两种情况：一是两个平面有平行或重合的关系；二是三个平面两两相交，三交线互相平行。$r(\overline{\boldsymbol{A}})$ 为 3，说明化简后有三个有效的平面，因而不存在两平面重合的情况。总之，相交情况有两种：
一种是如图（a）所示，Π_1 和 Π_2 平面平行并与 Π_3 相交于 L_1 和 L_2。
第二种是如图（b）所示，三个平面两两相交，三交线互相平行。
注：图（b）中绘出了三个法向量，其实，这三个法向量线性相关，第三个法向量可被另外两个法向量线性表示。或者 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$、$\boldsymbol{a}_3$ 在一个平面上，或者有两个向量线性相关。如图（a）中 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 线性相关</td>
<td>

(a)


(b)</td>
</tr>
<tr>
<td>2.7</td>
<td colspan="3">当 $r(\boldsymbol{A})\neq r(\overline{\boldsymbol{A}})$ 时，注意何时平面是重合的，何时平面是平行的。有个要点可以帮助你区分：设 $\begin{pmatrix}\boldsymbol{a}_1\\\boldsymbol{a}_2\\\boldsymbol{a}_3\end{pmatrix}$ 是系数矩阵 $\boldsymbol{A}$ 的行向量组，$\begin{pmatrix}\boldsymbol{\beta}_1\\\boldsymbol{\beta}_2\\\boldsymbol{\beta}_3\end{pmatrix}$ 是增广矩阵 $\overline{\boldsymbol{A}}$ 的行向量组。显然 $\begin{pmatrix}\boldsymbol{\beta}_1\\\boldsymbol{\beta}_2\\\boldsymbol{\beta}_3\end{pmatrix}$ 是 $\begin{pmatrix}\boldsymbol{a}_1\\\boldsymbol{a}_2\\\boldsymbol{a}_3\end{pmatrix}$ 的加长向量组，$\boldsymbol{a}_1$、$\boldsymbol{a}_2$、$\boldsymbol{a}_3$ 分别是 Π_1、Π_2 和 Π_3 的法向量，则有
● $\boldsymbol{a}_i$、$\boldsymbol{a}_j$ 相关，当且仅当平面 Π_i 与 Π_j 平行或重合；$\boldsymbol{\beta}_i$、$\boldsymbol{\beta}_j$ 相关，当且仅当 Π_i 与 Π_j 重合；
● $\boldsymbol{a}_i$、$\boldsymbol{a}_j$ 相关，化简后，一个向量 $\boldsymbol{a}_i$ 或 $\boldsymbol{a}_j$ 化为零向量，$r(\boldsymbol{A})$ 将会减 1；$\boldsymbol{\beta}_i$、$\boldsymbol{\beta}_j$ 相关， $r(\overline{\boldsymbol{A}})$ 将会减 1。
由此，如果化简前的原方程数是 m 个，化简后的方程数是 $r(\overline{\boldsymbol{A}})$，那么就有 $m-r(\overline{\boldsymbol{A}})$ 个平面重合，有 $r(\overline{\boldsymbol{A}})-r(\boldsymbol{A})$ 个平面平行</td>
</tr>
</table>

实际上，由立体几何的知识知道：三平面的相互关系有八种。因为两平面的相互位置关系

有三种：相交、平行、重合，而对于每一种情况增加第三个平面，除了重复的一种外，共有八种，见图 6-1。

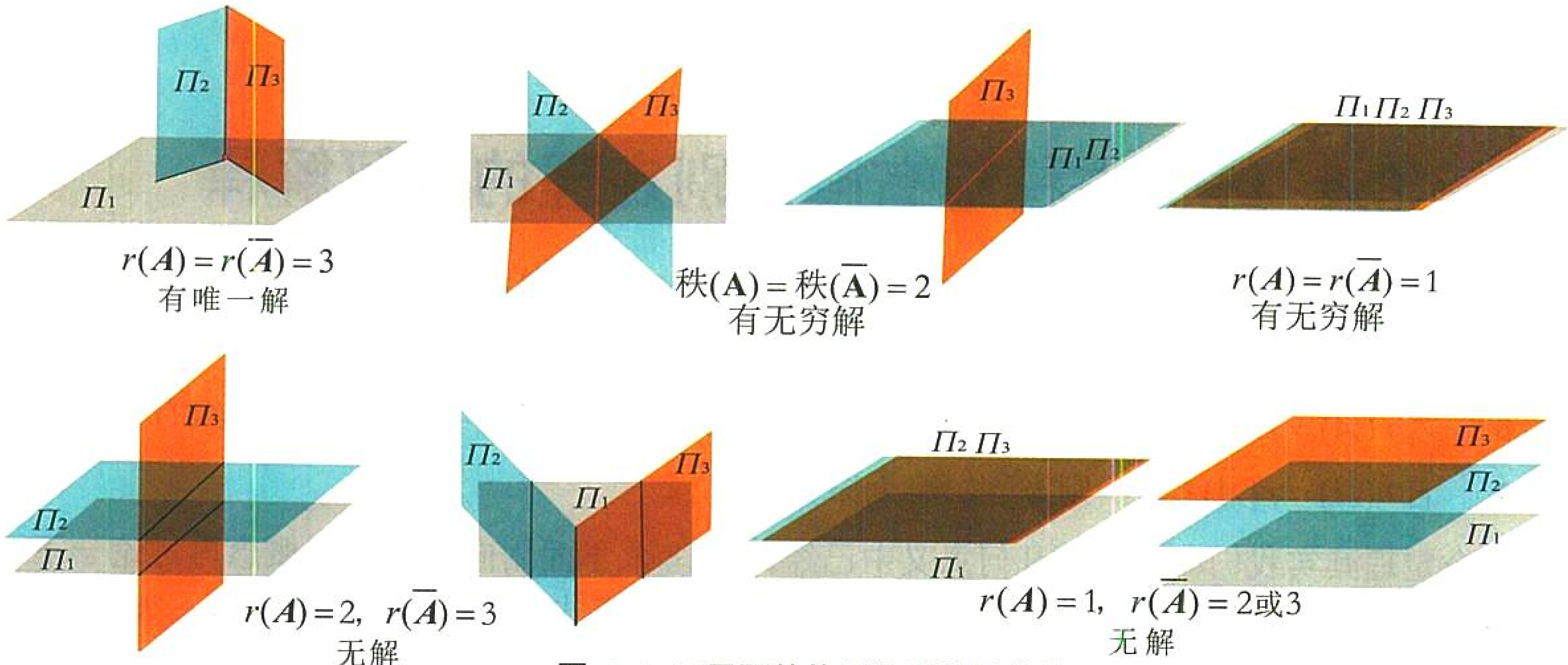


图 6-1 三平面的位置关系类型总结

根据线性方程组的秩及其解的不同情况，我们可以这样总结并定义线性方程组的分类：

设矩阵方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 的系数矩阵 $\boldsymbol{A}$ 是秩为 r 的 $m\times n$ 矩阵，增广矩阵 $\overline{A}=[A,b]$ 的秩为 r_c。在这里，我们要记住 m、n、r、r_c 代表的含义才能迅速理解下面的说明。m、n、r、r_c 代表的含义如下：

- m **：表示原方程组中方程的个数或者系数矩阵 A 的行数，化简后 m 一般会变小。**
- n **：表示原方程组中变元 x_i 的个数或者系数矩阵 A 的列数，化简后 n 一般也会变小。**
- r **：表示化简后变元 x_i 的个数，化简后 r 可能会变小，即 $r\leqslant n$。**
- r_c **：表示化简后方程的个数，化简后 r_c 可能会变小，即 $r\leqslant m$。**

那么，方程组的类型可以分为如下三类：

(1) **适定方程**：$r=r_c=n$。这里 $r=r_c$ 是说系数矩阵的秩 r 等于增广矩阵的秩 r_c，或者是说化简后变元个数等于方程个数，那么方程组有解；同时又有 $r=n$，说明化简后（或打假后）变元的个数没有变。总的说来，化简后原方程的变元个数没有变化，但化简后“真正”的方程个数就等于变元的个数，不多不少，几个变元就有几个方程，几个方程就可以解出几个变元，正合适确定一个向量（或点）解，称为适定方程；有唯一解。适定方程的系数矩阵 $\boldsymbol{A}$ 是 $n\times n$ 方阵。

(2) **欠定方程**：$r=r_c<n$。这里系数矩阵的秩 r 等于增广矩阵的秩 r_c，说明方程组有解；同时 $r_c<n$ 说明化简后（或打假后）变元的个数和方程的个数都减少了，个数都比原方程的变元个数 n 要小，“打假”后方程“真正”的个数不够，n 个变元只有小于 n 个的方程与之对应。小于 n 个方程只能解出小于 n 个变元，不能完全确定 n 个变元。化简后方程组如要确定一个向量解有欠缺（欠缺方程），称为欠定方程；不能确定的变元在数域内可以取无穷多，因此欠定方程组有无穷解。

(3) **超定方程**：$r<r_c$。系数矩阵的秩 r 小于增广矩阵的秩 r_c，方程组没有解。这里是说，

经过化简后的方程组中变元 x_i 的个数 r 小于方程的个数 r_c（原来方程组变元 x_i 的个数有 n 个，现在变小了，只有 r 个）；反过来讲，方程组的个数多于变元的个数，超过了变元的个数，因此称为超定方程。

6.4 线性方程组有解判别定理的几何解释

线性方程组的相容性定理（或有解判别定理）给出了方程组有解的充要条件：方程组的系数矩阵 $\boldsymbol{A}$ 的秩和增广矩阵 $\overline{\boldsymbol{A}}$ 的秩相等。进一步：

- 若 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=n$（$n$ 为方程组中未知量的个数：$x_1,x_2,\cdots,x_n$），则方程组有唯一解；
- 若 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})<n$，则方程组有无穷多解。

这个定理其实我们在上节的几何意义中已经得到理解,下面用向量空间的概念再讨论一下。

必要性：如果方程组有解，就是说线性方程组的向量等式 $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n=\boldsymbol{b}$ 成立，因此说明向量 $\boldsymbol{b}$ 可由向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 线性表示（或表出），故两个向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 和 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 等价。又因为向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 是向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 的子向量组，这两个向量组张成的向量空间 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 和 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 必然相同，当然秩也相同。而向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 和 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 分别是系数矩阵 $\boldsymbol{A}$ 和增广矩阵 $\overline{\boldsymbol{A}}$ 的列向量，因此矩阵 $\boldsymbol{A}$ 和 $\overline{\boldsymbol{A}}$ 的列秩相等，具有相同的列空间，从而有相同的秩：$r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})$。

充分性：反过来，如果有 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})$，就是说系数矩阵 $\boldsymbol{A}$ 和增广矩阵 $\overline{\boldsymbol{A}}$ 的列向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 和 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 张成的向量空间 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 和 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 的维数相同。又因为（显然）$\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 是 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 的子空间，换句话说，一个向量空间和它的子空间的维数一样，显然向量空间和子空间基相同，完全重合。因此向量 $\boldsymbol{b}$ 存在于向量空间 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 中，也就是存在于向量空间 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 中。故向量 $\boldsymbol{b}$ 可以由向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 线性表示，即 $\boldsymbol{b}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n$ 成立，方程组有解。

如图 6-2 所示，$\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 表示为一个超平面子空间 H（所谓超平面，是指图中所画的平面泛指一个子空间，表示二维或三维或更高维的子空间），方程组有解表示向量 $\boldsymbol{b}$ 在超平面子空间 H 内（见图 6-2（a））；如果无解，向量 $\boldsymbol{b}$ 在超平面子空间 H 外（见图 6-2（b））。

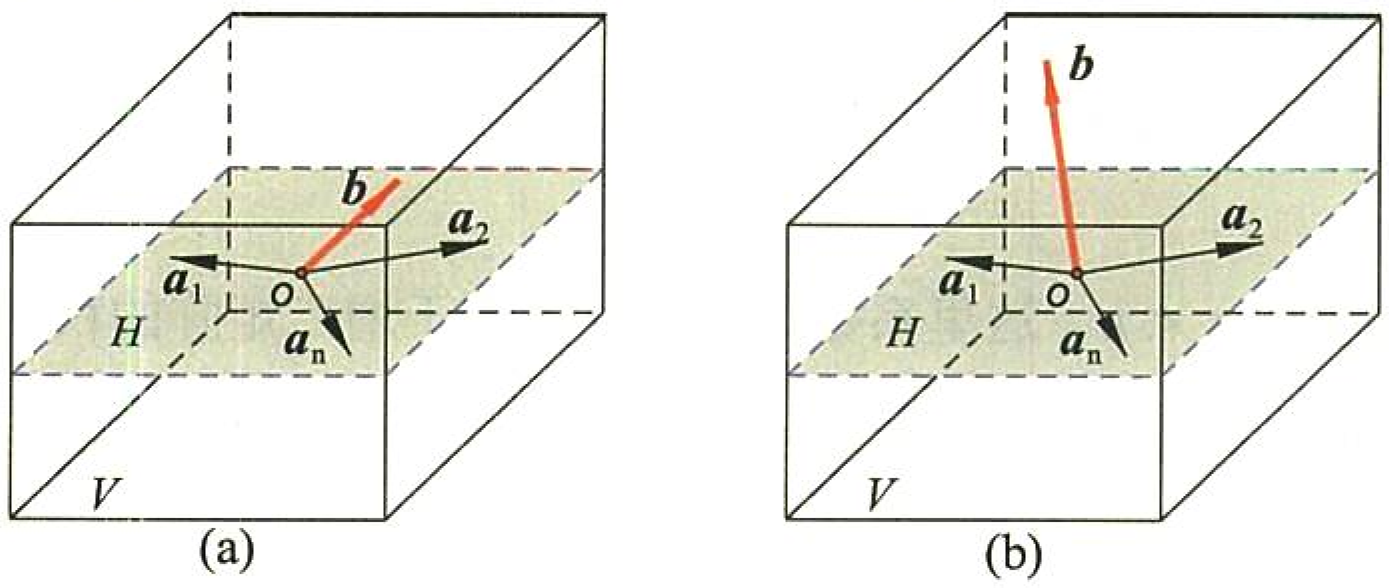


图 6-2 方程组有解和无解的几何意义

进一步的讨论：

- 如果 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})=n$，就是说方程组是一个由 n 个未知数的 n 个方程组成的。系数矩阵 $\boldsymbol{A}$ 的列向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$ 和增广矩阵 $\overline{\boldsymbol{A}}$ 的列向量 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}$ 张成的向量空间 $\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$

和 $\text{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n,\boldsymbol{b}\}$ 都是 n 维空间（前面讲过，两空间完全重合）。显然 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\}$ 是空间的基。由表达式 $\boldsymbol{b}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n$ 知道 $(x_1,x_2,\cdots,x_n)$ 是向量 $\boldsymbol{b}$ 的坐标（坐标是唯一的），因而 $(x_1,x_2,\cdots,x_n)$ 是唯一的，故方程组的解也是唯一的。

● 如果 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})<n$，就是说在一个 n 维的空间中，矩阵所张成的空间是低维的空间，增广矩阵的秩等于矩阵的秩，说明向量 $\boldsymbol{b}$ 在矩阵的空间内，所以可以被变换到，就是有解，有一个解了。那些高于矩阵秩的维数的向量都可以被矩阵变换到自身的空间中去——降维变换——同样可以变成向量 $\boldsymbol{b}$，就是说方程组有无穷多解。

前面高斯消元法一节说过，方程组的消元法体现在矩阵的操作上就是化增广矩阵为阶梯形矩阵（即矩阵三角化）。矩阵增广的意思就是要把所有的向量和变换矩阵的列向量放在一起同时进行基坐标系的更换，这样就可以保证所有的向量是参照同一个基的坐标值，从而保证解的等价性。

证明线性方程组相容的充分必要条件是它的系数矩阵与增广矩阵同秩。一般教材的证明尽管推导的每一步骤无比正确，无懈可击，但就是觉得难于理解。实际上，我们可以用前面矩阵的列初等变换知识这样处理：

设线性方程组相容，要证明 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})$。重写线性方程组的向量形式：

$$k_1\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix}+k_2\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{m2}\end{pmatrix}+\cdots+k_n\begin{pmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{mn}\end{pmatrix}=\begin{pmatrix}b_1\\b_2\\\vdots\\b_m\end{pmatrix}$$

据此化简增广矩阵，见表 6-6。

表 6-6　增广矩阵初等变换的步骤

顺序	初等变换的步骤说明	线性方程组的矩阵形式
1	原方程组的矩阵形式如右所示	$\overline{\boldsymbol{A}}=\left[\begin{array}{cccc:c}a_{11}&a_{12}&\cdots&a_{1n}&b_1\\a_{21}&a_{22}&\cdots&a_{2n}&b_2\\\vdots&\vdots&&\vdots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}&b_m\end{array}\right]$
2	对方程组进行列变换： 把第一列 $\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix}$ 乘以 $-k_1$ 加到最后一列； 把第二列乘以 $-k_2$ 加到最后一列……把第 n 列乘以 $-k_n$ 加到最后一列	$\left[\begin{array}{cccc:c}a_{11}&a_{12}&\cdots&a_{1n}&b_1-k_1a_{11}-k_2a_{12}\cdots-k_na_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}&b_2-k_1a_{21}-k_2a_{22}\cdots-k_na_{2n}\\\vdots&\vdots&&\vdots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}&b_m-k_1a_{m1}-k_2a_{m2}\cdots-k_na_{mn}\end{array}\right]$
3	因为原方程组相容，所以矩阵 $\overline{\boldsymbol{A}}$ 化简为右所示的矩阵。因为是经过初等的列变换得到的，所以 $r(\boldsymbol{A})=r(\overline{\boldsymbol{A}})$	$\overline{\boldsymbol{A}}=\left[\begin{array}{cccc:c}a_{11}&a_{12}&\cdots&a_{1n}&0\\a_{21}&a_{22}&\cdots&a_{2n}&0\\\vdots&\vdots&&\vdots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}&0\end{array}\right]\sim\boldsymbol{A}$

我们对增广矩阵施行初等行变换得到行的阶梯形矩阵后，就可以判断线性方程组的秩及解的情况了。无论如何，增广矩阵阶梯形及其解的个数之间的关系的三种情形可以形象地总结为图 6-3。

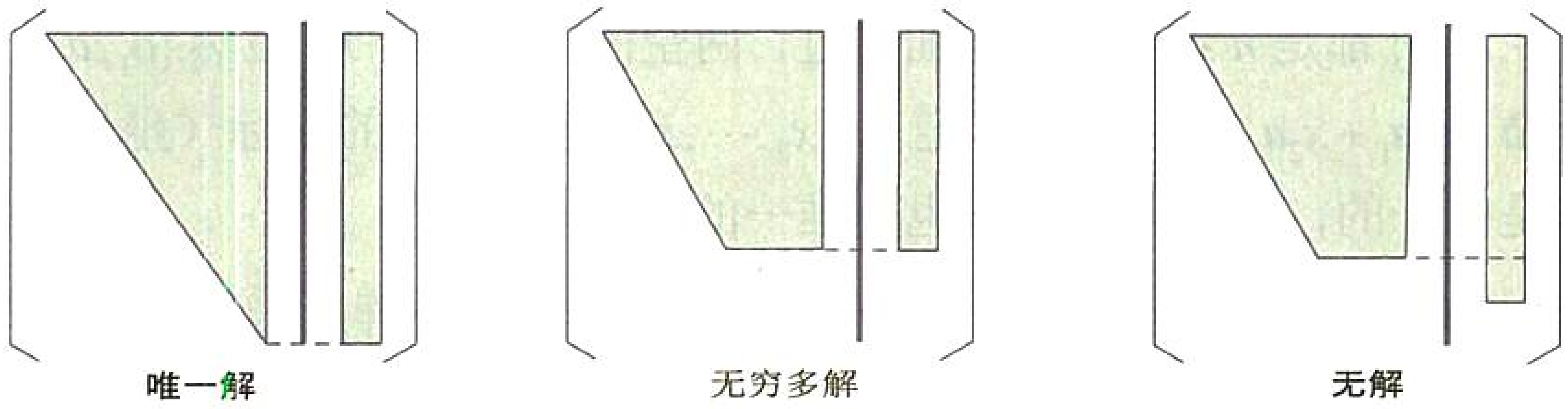


图 6-3 阶梯增广矩阵是否有解的外形

6.5 线性方程组解结构的几何意义

我们在前面看到，方程组没有解、有唯一解和有无穷多解的几何解释中，分别对应着多个超平面是否有唯一交点、交线或交面的情形。有解时的唯一交点当然是一个具体的向量；交线、交平面就是有无穷多的向量集合。用已知向量表示出来的交线或交平面的代数式就是方程组的解，代数式的构成形式就是方程组的解结构。

好，下面我们抄录经典的线性方程组解系通常的描述出来，便于讨论。

设有线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{6-2}$$

如果此方程组有解，则系数矩阵的秩 $r < n$，其通解或全部解为

$$\boldsymbol{Z}=\boldsymbol{\eta}^*+c_1\xi_1+c_2\xi_2+\cdots+c_{n-r}\xi_{n-r} \qquad (c_1,c_2,\cdots,c_{n-r}\text{ 为任意实数}) \tag{6-3}$$

其中：$\boldsymbol{\eta}^*$ 是方程组的特解向量；$\xi_1,\xi_2,\ldots,\xi_{n-r}$ 是方程组中令 $b_i\ (i=1,2,3,\cdots,m)$ 为零而得到的齐次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{6-4}$$

的基础解系。

其实式(6-3)就给出了解系的代数及几何结构，下面咱的任务是理解它。

6. 5. 1 线性方程组解的代数形式

再强调一下，一个方程或方程组表示的几何图形其实就是这个方程或方程组的**解**的几何图形。那么对于四元线性方程或方程组，我们将会看到它们的解的图形有一个规律（假设每个方程组没有“多余”的方程）：

1 个四元线性方程 $a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1$，表示四维空间里的一个**三**维“空间”体；

2 个四元线性方程 $\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=b_2 \end{cases}$，表示四维空间里的一个**二**维平面；

3 个四元线性方程 $\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=b_2 \\ a_{31}x_1+a_{32}x_2+a_{33}x_3+a_{34}x_4=b_3 \end{cases}$，表示四维空间里的一个**一**维直线；

4 个四元线性方程$\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=b_2 \\ a_{31}x_1+a_{32}x_2+a_{33}x_3+a_{34}x_4=b_3 \\ a_{41}x_1+a_{42}x_2+a_{43}x_3+a_{44}x_4=b_4 \end{cases}$，表示四维空间里的一个零维点。

可以看出，方程组解的几何图形的维数和方程组里方程的个数有关系。解的维数加上方程个数等于空间的维数（将会明白，方程组的个数就是系数矩阵的秩）。

这个结论也可以从上面方程组解的函数式看出来：

1 个四元线性方程$a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1$改写为函数式$x_1=\frac{-a_{12}}{a_{11}}x_2+\frac{-a_{13}}{a_{11}}x_3+\frac{-a_{14}}{a_{11}}x_4+\frac{b_1}{a_{11}}$，那么这个四维空间里的**三**维“空间”体是以$x_2$、$x_3$、$x_4$为自变量的坐标轴所张成的空间体，$x_2$、$x_3$、$x_4$可以取任意值，但$x_1$是因变量，不能任意取值，因而有 3 个自由度，所以是三维空间体。

如果要用向量的形式表示解，就要把x_1的因变量扩充为四维向量（四维空间里当然是四维向量），把恒等式编入方程组：

$$\begin{cases} x_1=\frac{-a_{12}}{a_{11}}x_2+\frac{-a_{13}}{a_{11}}x_3+\frac{-a_{14}}{a_{11}}x_4+\frac{b_1}{a_{11}} \\ x_2=\quad x_2 \\ x_3=\qquad\quad x_3 \\ x_4=\qquad\qquad\quad x_4 \end{cases} \tag{6-5}$$

方程组形式改写为向量形式：

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=\begin{pmatrix} \frac{-a_{12}}{a_{11}} \\ 1 \\ 0 \\ 0 \end{pmatrix}x_2+\begin{pmatrix} \frac{-a_{13}}{a_{11}} \\ 0 \\ 1 \\ 0 \end{pmatrix}x_3+\begin{pmatrix} \frac{-a_{14}}{a_{11}} \\ 0 \\ 0 \\ 1 \end{pmatrix}x_4+\begin{pmatrix} \frac{b_1}{a_{11}} \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{6-6}$$

从上式可以知道，解的三维空间体是由四个向量的线性组合而成的。x_2、x_3、x_4可以取任意值，任意常数一般写做k_i，为了与坐标区别，常把式（6-6）改写为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=\begin{pmatrix} \frac{-a_{12}}{a_{11}} \\ 1 \\ 0 \\ 0 \end{pmatrix}k_1+\begin{pmatrix} \frac{-a_{13}}{a_{11}} \\ 0 \\ 1 \\ 0 \end{pmatrix}k_2+\begin{pmatrix} \frac{-a_{14}}{a_{11}} \\ 0 \\ 0 \\ 1 \end{pmatrix}k_3+\begin{pmatrix} \frac{b_1}{a_{11}} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

因为加了一个常向量，但这个空间体不过原点，故不是线性空间。

2 个四元线性方程所组成的方程组$\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=b_1 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=b_2 \end{cases}$通过行变换消元法改写为函数式$\begin{cases} x_1=a'_{13}x_3+a'_{14}x_4+b'_1 \\ x_2=a'_{23}x_3+a'_{24}x_4+b'_2 \end{cases}$，那么这个四维空间里的**二**维平面是由$x_3$、$x_4$为自变量的坐标轴所张成的平面。进一步写成向量的形式为$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} a'_{13} \\ a'_{23} \end{pmatrix}x_3+\begin{pmatrix} a'_{14} \\ a'_{24} \end{pmatrix}x_4+\begin{pmatrix} b'_1 \\ b'_2 \end{pmatrix}$，那么这个四维空间里的二维向量平面是由向量$\begin{pmatrix} a'_{13} \\ a'_{23} \end{pmatrix}$、$\begin{pmatrix} a'_{14} \\ a'_{24} \end{pmatrix}$、$\begin{pmatrix} b'_1 \\ b'_2 \end{pmatrix}$的线性组合而成的。

同样地，在四维空间里讨论向量，当然要把因变向量$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$扩充为四维向量，类似如式(6-5)要把恒等式编入方程组，并把x_3、x_4改为k_1、k_2，得到

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} a'_{13} \\ a'_{23} \\ 1 \\ 0 \end{pmatrix} k_1 + \begin{pmatrix} a'_{14} \\ a'_{24} \\ 0 \\ 1 \end{pmatrix} k_2 + \begin{pmatrix} b'_1 \\ b'_2 \\ 0 \\ 0 \end{pmatrix}$$

3 个四元线性方程所组成的方程组 $\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = b_3 \end{cases}$ 通过行变换消元法改写为函数式 $\begin{cases} x_1 = a'_{14}x_4 + b'_1 \\ x_2 = a'_{24}x_4 + b'_2 \\ x_3 = a'_{34}x_4 + b'_3 \end{cases}$，那么这个四维空间里的**一**维直线是由 x_4 为自变量所张成的图形。进一步写成向量的形式为 $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a'_{14} \\ a'_{24} \\ a'_{34} \end{pmatrix} x_4 + \begin{pmatrix} b'_1 \\ b'_2 \\ b'_3 \end{pmatrix}$，那么这个四维空间里的二维向量平面是由向量 $\begin{pmatrix} a'_{14} \\ a'_{24} \\ a'_{34} \end{pmatrix}$ 和 $\begin{pmatrix} b'_1 \\ b'_2 \\ b'_3 \end{pmatrix}$ 的线性组合而成的，表示四维空间里的一个**一**维直线。把因变向量 $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ 扩充为四维向量，得到

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} a'_{14} \\ a'_{24} \\ a'_{34} \\ 1 \end{pmatrix} k + \begin{pmatrix} b'_1 \\ b'_2 \\ b'_3 \\ 0 \end{pmatrix}$$

4 个四元线性方程所组成的方程组 $\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = b_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 = b_4 \end{cases}$ 通过行变换消元法改写为函数式 $\begin{cases} x_1 = b'_1 \\ x_2 = b'_2 \\ x_3 = b'_3 \\ x_4 = b'_4 \end{cases}$，这个四维空间里的 **0** 维点就已经确定了。直接改写为向量形式就是

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b'_1 \\ b'_2 \\ b'_3 \\ b'_4 \end{pmatrix}$$

到了这里，方程组的解的代数形式就清楚了。

总结下，如果 n 阶方程组的秩为 r，那么其解的一般结构或解系就是 $n\text{-}r\text{+}1$ 个向量的线性组合，即

$$k_1\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + k_2\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} + \cdots + k_{n-r}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \quad （k_i \text{ 为实变量}） \qquad (6\text{-}7)$$

把线性方程组解的代数形式——向量的线性组合用图形绘制出来就是解结构的几何意义。其实式（6-7）由两部分组成，第一部分是 $n\text{-}r$ 个变向量的线性组合部分，第二部分是特解常向量 $\boldsymbol{d}$，即 $(d_1, d_2, \ldots, d_n)^{\mathrm{T}}$。

第一部分是齐次线性方程组的解，其解显然构成了一个线性空间（见 4.2.1 节），因此称为齐次线性方程组的解空间。非齐次方程组的解是齐次线性方程组解空间的平移——加上一个常向量，因为平移后不再过原点了，所以不能称之为解空间，只好称之为解结构或解系了。

6.5.2 齐次线性方程组的解空间

齐次线性方程组解的几何意义就是：

n 元线性齐次方程组的秩是 r 时，它的一切解向量的集合是 $\mathbf{R}^n$ 上的一个 n-r 维的子空间。

另外，我们也常听说，齐次线性方程组的这个 n-r 维的解空间正交于系数矩阵 $\boldsymbol{A}$ 的行向量，还听说，解空间和行空间是一对正交补。结合上节的分析，这个已不难理解了。但如果要朴素地理解的话咱就再分析下。

以三元齐次线性方程组 $\boldsymbol{Ax}=\mathbf{0}$ 作为例子，把系数矩阵 $\boldsymbol{A}$ 的元素改写成行向量后，方程组的形式为

$$\begin{pmatrix} \boldsymbol{a}'_1 \\ \boldsymbol{a}'_2 \\ \boldsymbol{a}'_3 \end{pmatrix} \boldsymbol{x} = \mathbf{0}$$

那么方程组的任意解向量 $\boldsymbol{x}$ 都与每个行向量 $\boldsymbol{a}'_i$ 正交。进一步说，在三维欧氏空间里 $\mathbf{R}^3$，只要与每个行向量都正交的向量都是解向量。

如果三个行向量 $\boldsymbol{a}'_i$ 都线性无关，那么这些行向量会张成一个三维子空间 V_3（见图 6-4(a)），V_3 和 $\mathbf{R}^3$ 同构——两空间重合了，那么 $\mathbf{R}^3$ 里能与 V_3 正交的向量只有 $\mathbf{0}$ 向量，也就是解向量只有 $\mathbf{0}$ 向量。$\mathbf{0}$ 向量张成的解空间只能称为零维空间了。因此，立体的行空间 V_3 和零解空间 O 是正交补。

补充下，正交补的意思是：n 维空间里的两个子空间正交，且两个子空间原点重合起来刚好能张成 n 维空间，互为补充，则这两个子空间互为正交补。好，继续讨论。

如果三个行向量 $\boldsymbol{a}'_i$ 里只有两个线性无关，则这些行向量会张成一个二维子空间 V_2，这是一个平面（见图 6-4（b）），那么 $\mathbf{R}^3$ 里面与平面 V_2 正交的向量很多，它们张成一个直线，也就是解向量空间是一根过原点的直线，解空间是一维空间了。因此，平面的行空间 V_2 和直线解空间是正交补。

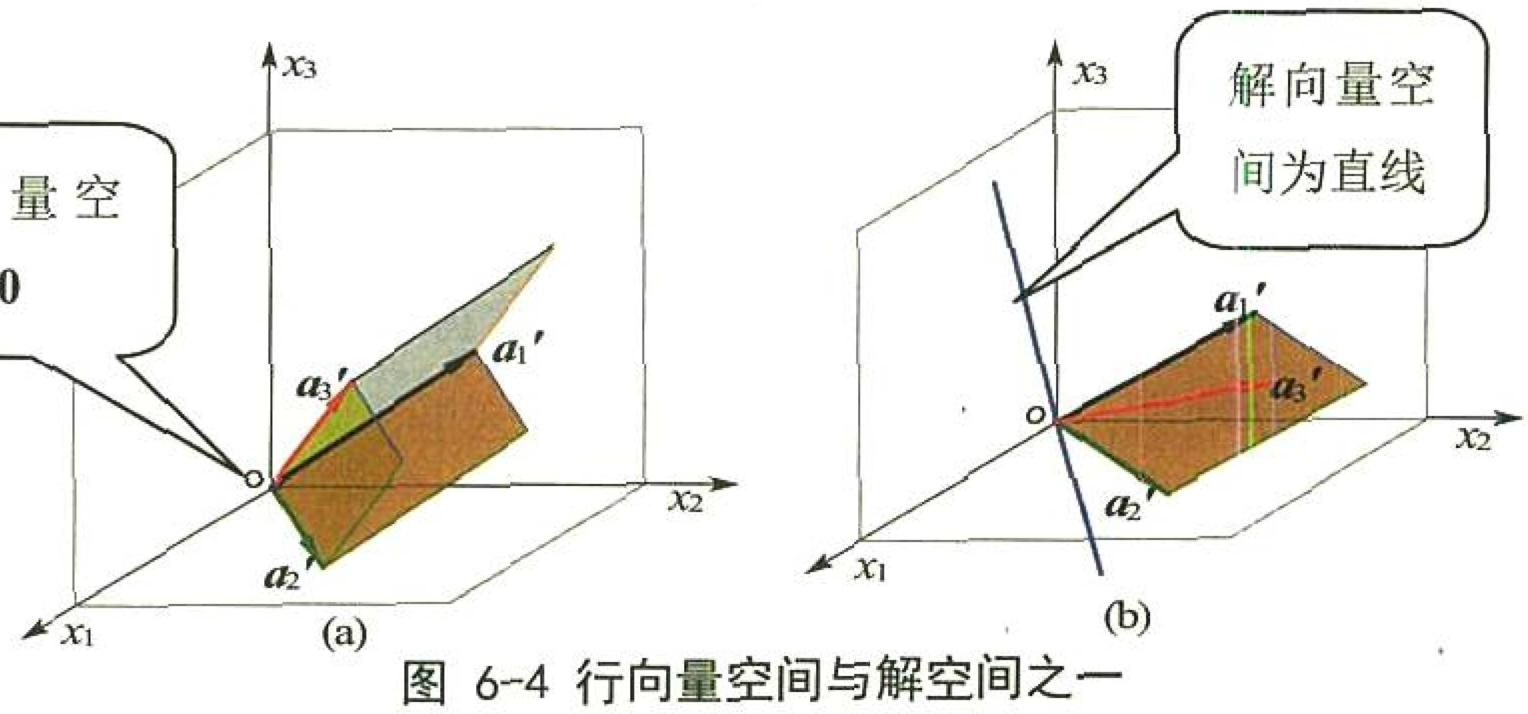


图 6-4 行向量空间与解空间之一

如果三个行向量 $\boldsymbol{a}'_i$ 里两两都线性相关，则这些行向量会张成一个一维子空间 V_1，这是一个直线（见图 6-5（a）），那么 $\mathbf{R}^3$ 里面与一直线 V_1 正交的向量是啥？显然是一个平面，也就是解向量空间是一个过原点的平面，解空间是二维空间了。因此，直线的行空间 V_1 和平面的解空间是正交补。

如果，呃，如果三个行向量 $\boldsymbol{a}'_i$ 等于 $\mathbf{0}$，系数矩阵是 $\mathbf{0}$（呃，方程组不存在啊，不用讨论了吧），与原点正交的向量就是整个三维空间了（见图 6-5（b））。形式上也满足原点的行空间和

立体的解空间是正交补。

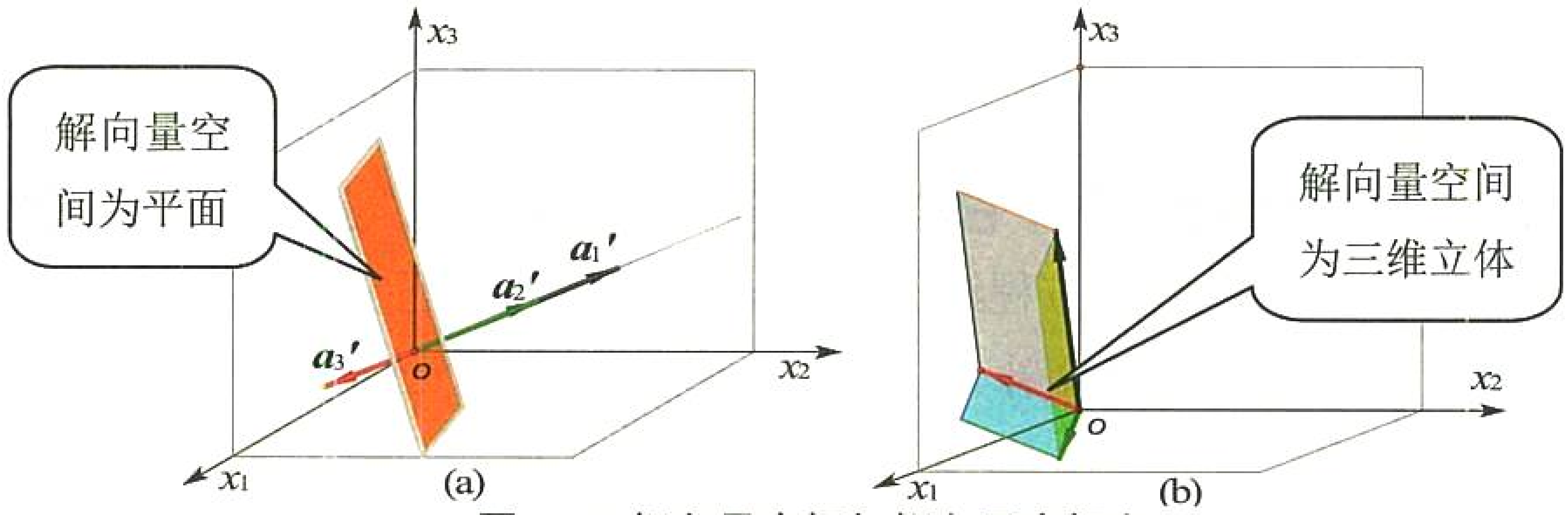


图 6-5 行向量空间与解向量空间之二

6.5.3 非齐次线性方程组的解结构

根据式(6-7)，非齐次方程组解的几何图形是齐次方程组解空间与一个特定常向量$\boldsymbol{d}$的和，这实际上就是把齐次方程组解空间的几何图形沿着特定向量方向平移了一个距离，这个距离等于常向量 $\boldsymbol{d}$ 的长度。

其实在变向量一节（见 2.8 节）中，我们已经用向量的形式表示出来了空间中一个过原点的线、面和不过原点的线和面的图形，如图 6-6 所示。

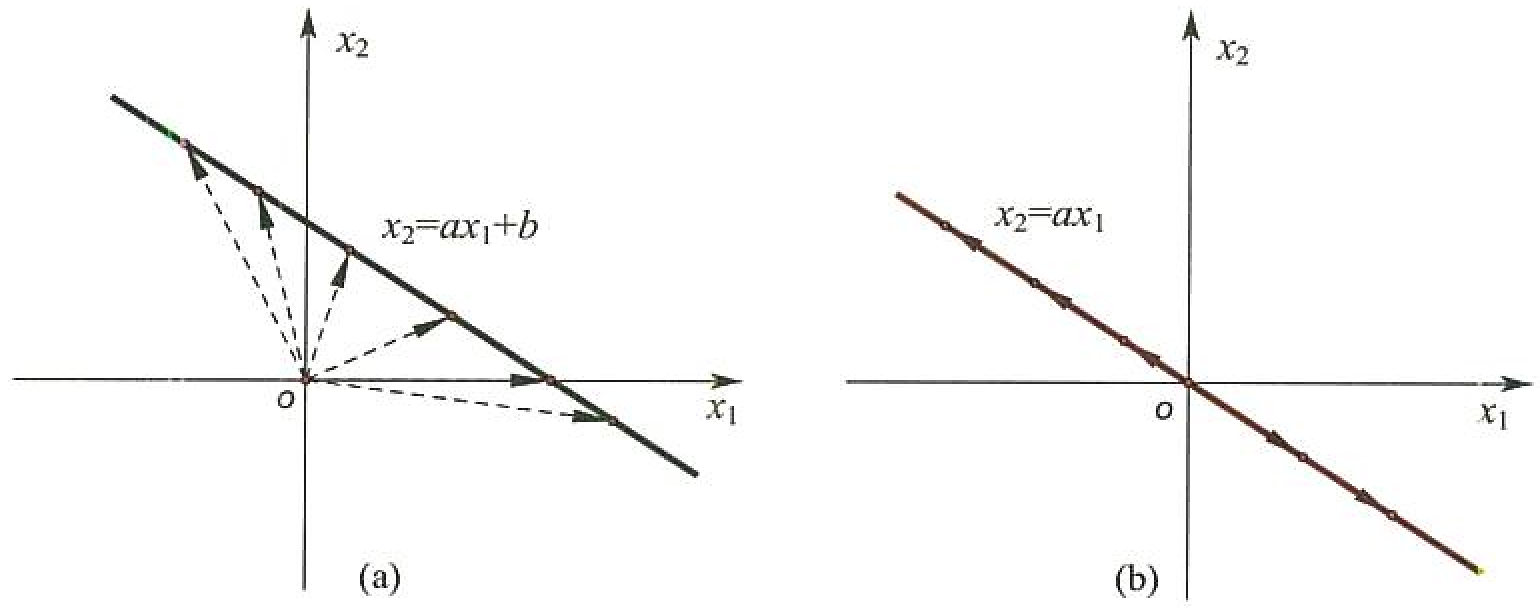


图 6-6 不过原点直线和过原点直线上的向量图形

图中图形是不过原点的直线 $x_2 = ax_1 + b$ 和过原点的直线 $x_2 = ax_1$，用变向量来表示这对平行线分别就是 $(x_1, ax_1 + b)$ 和 (x_1, ax_1)。不过原点的直线可以用过原点的直线和一个常向量的和表示：

$$(x_1, ax_1 + b) = (x_1, ax_1) + (0, b)$$

上述的变向量表述为行向量的合并形式，大家可能看得有点陌生。OK，用大家常见的列向量形式改写这对平行线，分别就是 $x_1\begin{pmatrix}1\\a\end{pmatrix}+\begin{pmatrix}0\\b\end{pmatrix}$ 和 $x_1\begin{pmatrix}1\\a\end{pmatrix}$。这里，不过原点的直线是过原点的直线和一个常向量的和的关系一目了然（见图 6-7）。

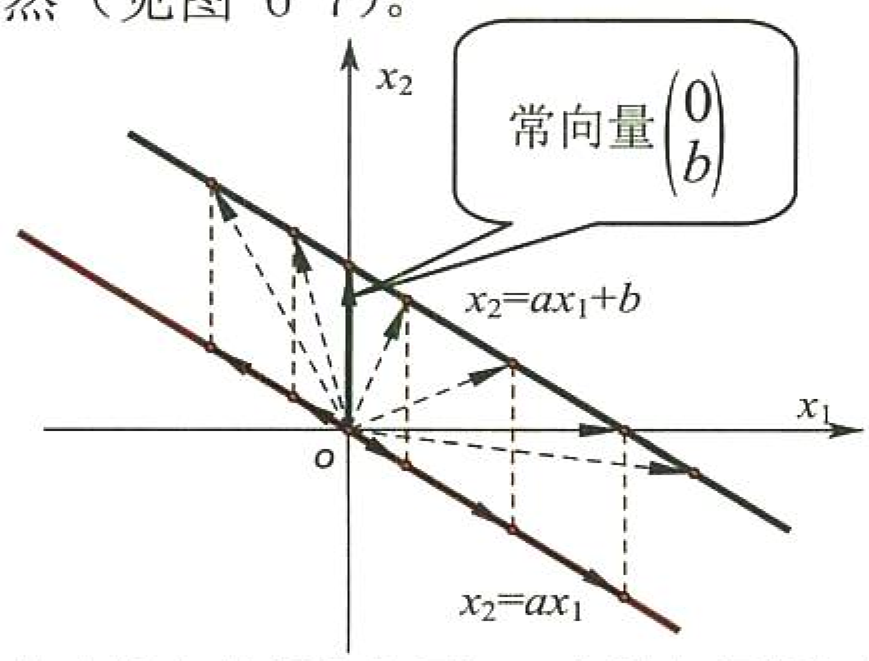


图 6-7 过原点的直线上的所有向量与一个常向量相加得到平移直线

好。这样咱就可以用图形表示非齐次线性方程组解的几何图形了。

比如上节中三元齐次线性方程组的例子里，如果解空间是一个直线或者平面，那么同一个系数矩阵的非齐次线性方程组的解系就是把这根直线或平面平移一个特解向量的距离（见图 6-8）。

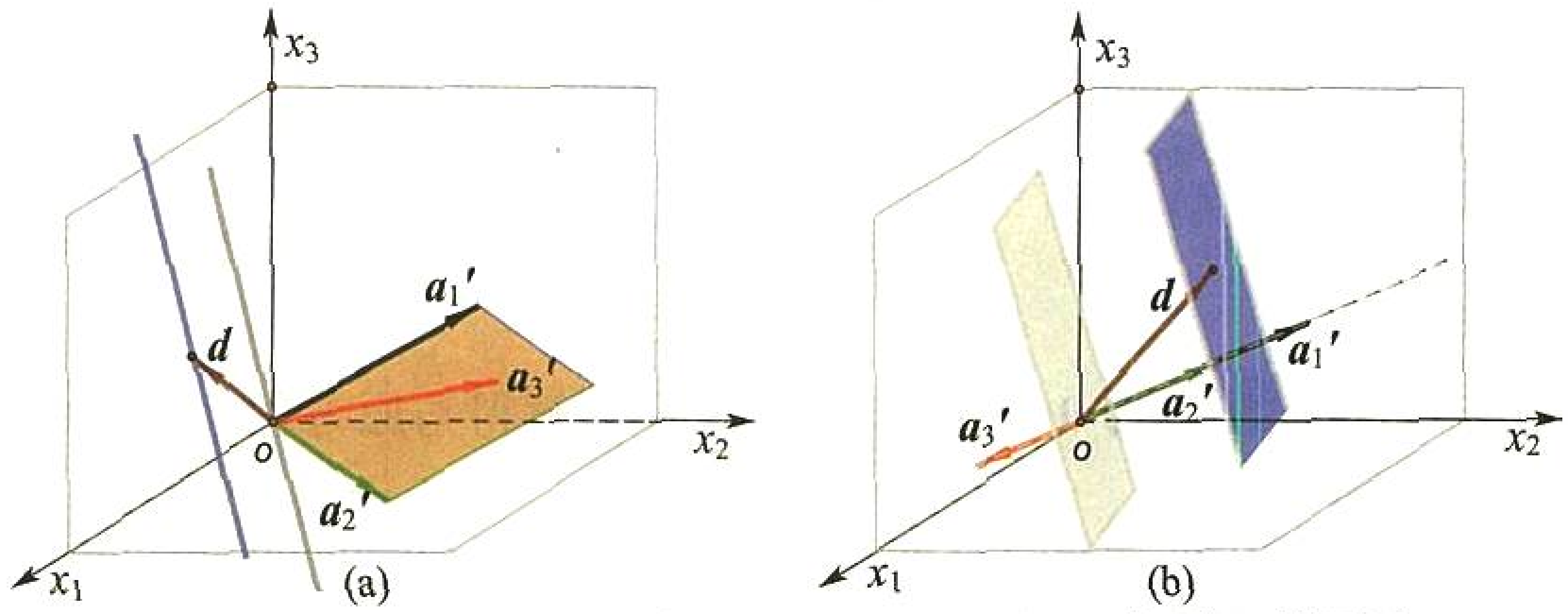


图 6-8 齐次和非齐次方程组解的向量图形是平行的平移图形

所以，齐次方程组和对应的非齐次方程组的解析几何图形是平行的，或者说齐次方程组的解的几何图形和对应的非齐次方程组的解的几何图形是平行的。具体来说，齐次方程组的所谓基础解系 $c_1\xi_1+c_2\xi_2+\cdots+c_{n-r}\xi_{n-r}$ 的几何图形和非齐次方程组的通解 $\boldsymbol{\eta}^*+c_1\xi_1+c_2\xi_2+\cdots+c_{n-r}\xi_{n-r}$ 的几何图形是平行的，基础解系是过原点的超平面，通解是不过原点（移开原点）的平行平面。

值得注意的是，齐次解的向量图形只要加上任意一个顶点在非齐次图形上的常向量，就可以得到所有的非齐次解向量，非齐次解向量的顶点构成了非齐次解的几何图形。

还有一个值得注意的是，虽然非齐次解向量图形与行向量空间不再正交，但是非齐次解的几何图形仍然与行空间的几何图形保持**垂直**的解析性质（见图 6-8）。

6.5.4 非齐次线性方程组的例解

例 6.1 解下列线性方程组，并说明其通解的几何意义：

$$\begin{cases} x_1 - x_2 - x_3 = -1 \\ x_1 + x_2 + 3x_3 = 3 \\ 3x_1 - x_2 + x_3 = 1 \end{cases}$$

解　对方程组的增广矩阵 $\boldsymbol{B}$ 作初等变换：

$$\boldsymbol{B} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 3 & 3 \\ 3 & -1 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -1 & -1 & -1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

因为 $r(\boldsymbol{A})=r(\boldsymbol{B})=2$，所以方程组有解。由此得同解方程组

$$\begin{cases} x_1 \quad + x_3 = 1 \\ \quad x_2 + 2x_3 = 2 \end{cases}$$

把上式变量 x_3 列右移，并补齐恒等式 $x_3=x_3$ ，得到

$$\begin{cases} x_1 = -x_3 + 1 \\ x_2 = -2x_3 + 2 \\ x_3 = x_3 \end{cases}$$

将其改写为向量的形式：

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \tag{6-8}$$

令式（6-8）中的自由未知量$x_3 = 0$，得到方程组的一个特解：

$$\boldsymbol{\eta} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

令式（6-8）中的常向量项为**0**，并取$x_3 = 1$，便可解得与原方程组对应的齐次线性方程组的基础解系：$\boldsymbol{\xi} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$。因此所给的方程组的通解或全部解为

$$\boldsymbol{Z} = C\boldsymbol{\xi} + \boldsymbol{\eta} = C\begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \qquad （C\text{为任意实数}）$$

如图 6-9（a）所示，过向量$\boldsymbol{\xi}$作直线L，则以原点为起点、以L上任意一点为终点的向量$C\boldsymbol{\xi}$都是基础解，它们全体构成的直线L就是所给方程组的基础解空间。

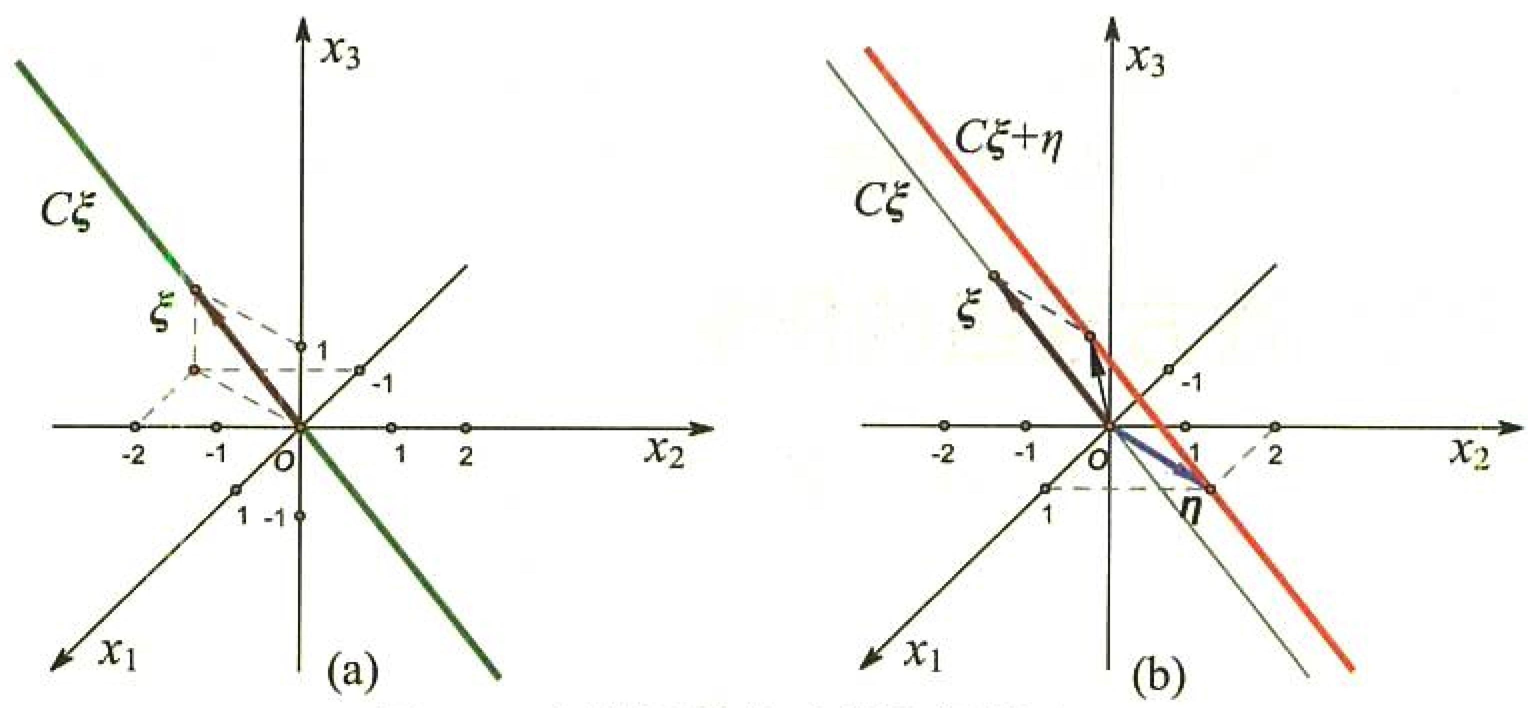


图 6-9 方程组的基础解和通解之一

如图 6-9（b）所示，过特解向量$\boldsymbol{\eta}$的终点作平行于直线$C\boldsymbol{\xi}$的直线$C\boldsymbol{\xi}+\boldsymbol{\eta}$，则以原点为起点，以直线上任意点为终点的向量$C\boldsymbol{\xi}+\boldsymbol{\eta}$，都是方程组的解，它们全体构成了方程组的解系。当通解中的C取遍所有实数时，可以得到解系中的任意解向量。

例 6.2 解下列方程组，并说明其通解的几何意义：

$$\begin{cases} 2x_1 + x_2 - 4x_3 = 4 \\ -x_1 - 0.5x_2 + 2x_3 = -2 \end{cases}$$

解 对方程组的增广矩阵$\boldsymbol{B}$作初等变换：

$$\boldsymbol{B} = \begin{bmatrix} 2 & 1 & -4 & 4 \\ -1 & -0.5 & 2 & -2 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 1 & -4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0.5 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

因为$r(\boldsymbol{A}) = r(\boldsymbol{B}) = 1$，所以方程有解。由此得同解方程组

$$x_1 + 0.5x_2 - 2x_3 = 2$$

把上式 x_2、x_3 变量右移，并补齐缺省的恒等式 $x_2=x_2$，$x_3=x_3$，有同解方程组：

$$\begin{cases} x_1=-0.5x_2+2x_3+2 \\ x_2= \quad x_2 \\ x_3= \qquad x_3 \end{cases}$$

将其写成向量的形式：

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=x_2\begin{pmatrix} -0.5 \\ 1 \\ 0 \end{pmatrix}+x_3\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}+\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

把上式中的 x_2、x_3 改写成 C_1、C_2 的形式，得到方程组的通解或全部解为

$$\boldsymbol{Z}=C_1\xi_1+C_2\xi_2+\boldsymbol{\eta}=C_1\begin{pmatrix} -0.5 \\ 1 \\ 0 \end{pmatrix}+C_2\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}+\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \qquad （c_1, c_2 \text{ 为任意实数}）$$

如图 6-10（a）所示，过向量 ξ_1、ξ_2 作平面 Π_1，则以原点为起点、以 Π_1 上任意点为终点的向量 $C_1\xi_1+C_2\xi_2$ 均为基础解，它们全体构成的平面 Π_1 即为所给方程组的基础解空间。

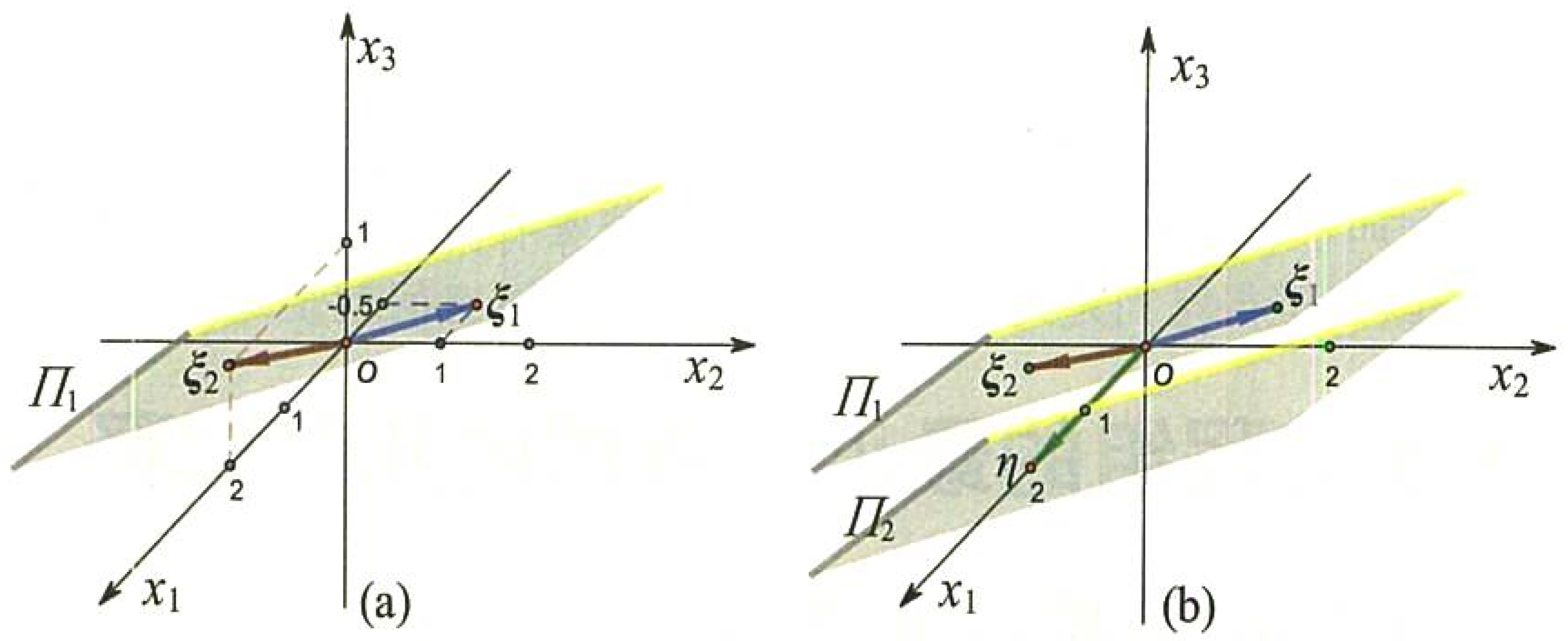


图 6-10 方程组的基础解和通解之二

如图 6-10（b）所示，过特解向量 $\boldsymbol{\eta}$ 的终点作与 Π_1 平行的平面 Π_2，则以原点为起点，以平面 Π_2 上任意点为终点的向量 $C_1\xi_1+C_2\xi_2+\boldsymbol{\eta}$ 都是原方程组的解向量，它们全体构成了方程组的通解解系。当通解中的 C_1、C_2 取遍所有实数时，可以得到所有解向量。

6.6　数域上的线性方程组（或向量空间）的意义

在使用高斯消元法解方程组时，大家有没有注意一个有意思的细节：在整个解的过程中，只进行加减乘除四则运算，没有涉及开方运算，更没有涉及解二次以至更高次方程的计算。由此看来，我们可以得到以下的推论：

- 假设线性方程组的增广矩阵的元素全部是有理数（意思方程组的系数和常数项全部是有理数），那么加、减、乘、除的消元计算结果只能是有理数。如果我们定义方程组的基础解系中的任意数为有理数，则方程组的解集也是有理数，从而我们可以完全在有理数的范围内讨论线性方程组。
- 类似地，我们也可以把对方程组的讨论限制在实数域，因为加、减、乘、除不会出现虚数。

对方程组的讨论可以限定在某个数域，数域的定义是：

一个集合 K 是由一些数组成的，如果这些数中的任意两个数相加、减、乘、除的结果仍然属于 K；换句话讲，集合里的全部数对加、减、乘、除四种运算是封闭的，则我们称 K 为一个数域。

数域的例子有：有理数域 $\mathbf{Q}$、实数域 $\mathbf{R}$、复数域 $\mathbf{C}$ 等。

还有一个数域的例子是集合 $\{a+b\mathrm{i}$，（a，b 是有理数）$\}$，可以验证全体 $a+b\mathrm{i}$ 对于加、减、乘、除封闭。

全体的整数集合不是数域，因为对除法不封闭，两整数相除会出现分数。

有了数域的概念后，前面的讨论都可以限制在某一个数域上了。如以数域 K 的元素为系数和常数项的线性方程组，就可以称之为**数域 K 上的线性方程组**。进一步，对基础解系的任意数也可以限制在数域 K 内来。

如果以数域 K 的元素为分量的全体 n 维向量所组成的集合，集合里的向量关于加法和数乘（数乘的数也属于数域 K）封闭，因此，对这个向量集合的讨论就可以限制在数域 K 内进行，故称之为**数域 K 上的 n 维向量空间**。例如实数域 $\mathbf{R}$ 的 n 维向量空间 $\mathbf{R}^n$，复数域 $\mathbf{C}$ 上的 n 维向量空间 $\mathbf{C}^n$。

注：
原来俺在学习函数或映射时，不太在意声明元素属于（∈）某个数域、定义域、值域之类，现在看来这个域不是随便叫的，它和所讨论的具体问题的运算有关，它和线性方程组甚至向量空间的运算有关！比如，欧几里得空间是定义在实数域上的，而酉空间就是定义在复数域上的。

6.7 超定方程组的最小二乘解的几何解释

6.7.1 最小二乘法的向量解的几何意义

到现在为止，一个线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 可能有解也可能无解，如果 $\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的列空间中，则方程组是不相容的，高斯消元法将无法使用。另外，超定方程组也是不相容的。所谓的超定方程组是方程个数多于未知数个数的方程组，而且每个方程都不能通过行变换化为其他的方程。在实践中由于测量误差的存在和多次测量，超定方程组是很常见的。

尽管不相容的方程组在数学严格的意义上不可解，但我们还要求必须解出来它们，即使结果是个非精确的解。最小二乘法就是这样的一个求解方法。

对于线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 无解，向量 $\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的列空间 H 中（见图 6-11（a）），那我们可以在列空间中找个替代向量来求解，这个替代向量在误差要求的范围内可以接受。有一个寻找替代向量的方法就是把向量 $\boldsymbol{b}$ 向列空间投影，得到一个投影向量 $\boldsymbol{b}'$ 作为替代向量。可以证明或看出，这个投影 $\boldsymbol{b}'$ 是所有列空间向量里最接近原向量 $\boldsymbol{b}$ 的，是误差最小的，因为在点 $\boldsymbol{b}$ 到空间平面上任意一点的连线中只有到投影点是最近的，投影线正交（或垂直）于平面。比如，在空间平面上任找一个向量 $\boldsymbol{b}''$ 或 $\boldsymbol{b}'''$（见图 6-11（b）），在直角三角形 $\Delta\boldsymbol{bb}'\boldsymbol{b}''$ 或 $\Delta\boldsymbol{bb}'\boldsymbol{b}'''$ 中，投影线 $\boldsymbol{bb}'$ 作为直角边永远小于任意斜角边。也就是说，作为误差的投影线，其长度是最小的，亦即投影向量替代原向量的误差最小。

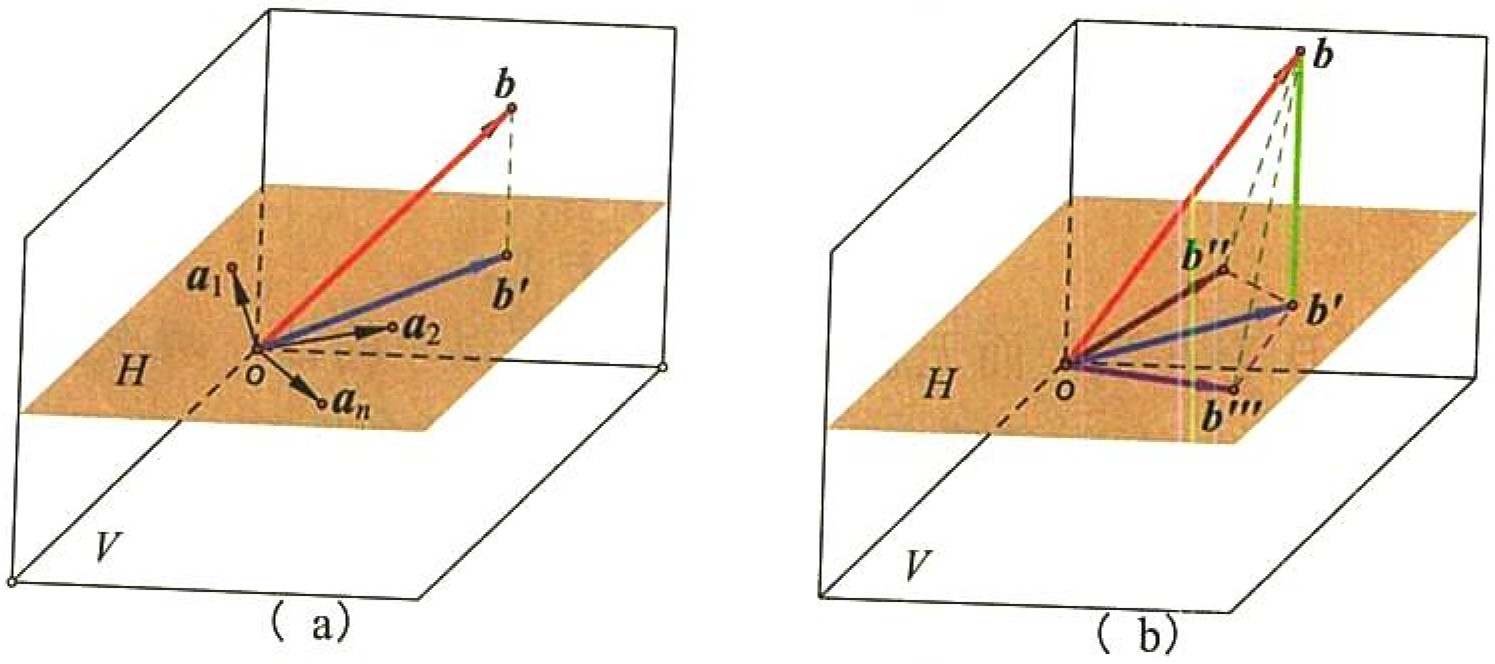


图 6-11 无解方程组的最小误差解法图示

设向量的坐标式为 $\boldsymbol{b}=(x_{b1},x_{b2},\ldots,x_{bn})$，$\boldsymbol{b}'=(x'_{b1},x'_{b2},\cdots,x'_{bn})$，那么投影线向量是 $\boldsymbol{b}$ 和 $\boldsymbol{b}'$ 相减的差（就是误差向量）：

$$\overrightarrow{\boldsymbol{bb}'}=\boldsymbol{b}-\boldsymbol{b}'=(x_{b1}-x'_{b1},x_{b2}-x'_{b2},\cdots,x_{bn}-x'_{bn})$$

其长度的平方就是

$$\left|\overrightarrow{\boldsymbol{bb}'}\right|^2=(x_{b1}-x'_{b1})^2+(x_{b2}-x'_{b2})^2+\cdots+(x_{bn}-x'_{bn})^2$$

长度最小等价于长度的平方最小。现在看来这个最小误差是以平方和最小的形式来评估的，因此称为最小二乘法。

总结，方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解（近似解）就是方程组 $\boldsymbol{Ax}=\boldsymbol{b}'$ 的解。

6.7.2 一般最小二乘解的公式推导

一般地，方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解有一个公式，我们给出推导的解释。

我们知道，误差向量 $\boldsymbol{b}-\boldsymbol{b}'$ 与矩阵 $\boldsymbol{A}$ 的列空间正交，也就是与 $\boldsymbol{A}$ 的每个列向量正交，那么向量 $\boldsymbol{b}-\boldsymbol{b}'$ 与每个列向量的点积等于零；$\boldsymbol{A}$ 转置一下，列向量变成行向量，仍然有内积等于零：

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{b}')=\boldsymbol{0}$$

把 $\boldsymbol{Ax}=\boldsymbol{b}'$ 代入上式，有 $\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{Ax})=\boldsymbol{0}$。化简得到所谓的正规方程或法方程：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

这个方程组的解就是原方程组的最小二乘解。解出这个方程组，问题就搞定了。

进一步，如果 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 可逆（注意，不可逆上述正规方程也有解），那么有解的公式为

$$\boldsymbol{x}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{6-9}$$

6.7.3 最小二乘解的例析

下面我们看一个例子，由此了解最小二乘解的解析上的几何意义。

例 6.3 在科学或工程中的一项任务就是分析或理解一些数量变化之间的联系，其中一些数量的点列形成的图形近似接近于直线。比如在向火星发射探测器时，我们每隔一段时间测量一下探测器到地球的距离。在这种情况下，t 是时间，y 是距离，假设探测器发动机不再工作且地球对其的引力可忽略，那么这个探测器应该是以近乎常数的速度 V 在运动：$y=Y_0+Vt$。经过 m 次测量后得到一组方程组：

$$\begin{cases} Y_0 + Vt_1 = y_1 \\ Y_0 + Vt_2 = y_2 \\ \quad\vdots \\ Y_0 + Vt_m = y_m \end{cases}$$

这是一个超定方程组，因存在误差而无解（如果不存在误差的话，两次测量就可以搞定直线方程了）。因有两个未知数Y_0、V需要确定，根据最小二乘法的图解，这是在m维空间里，一个m维向量向一个二维子空间投影的问题。把方程组写成向量形式：

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}\begin{pmatrix} Y_0 \\ V \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

假设测量了4组时间和距离的数据（t,y）为（0，0）、（1，1）、（3，2）、（4，5），则相应的方程组就是

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}\begin{pmatrix} Y_0 \\ V \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix}$$

对应于$\boldsymbol{A}x = \boldsymbol{b}$，如使用公式计算，首先要计算$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$及其逆：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}, \quad (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1} = \frac{1}{20}\begin{bmatrix} 13 & -4 \\ -4 & 2 \end{bmatrix}$$

于是由最小二乘解公式（6-9）得

$$\begin{pmatrix} Y_0 \\ V \end{pmatrix} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \frac{1}{20}\begin{bmatrix} 13 & -4 \\ -4 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix}\begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} -0.2 \\ 1.1 \end{pmatrix}$$

因此最佳拟合直线为$y = -0.2 + 1.1t$，见图 6-12。其中，所测量的数据y_i到直线的竖直距离的平方和最小。

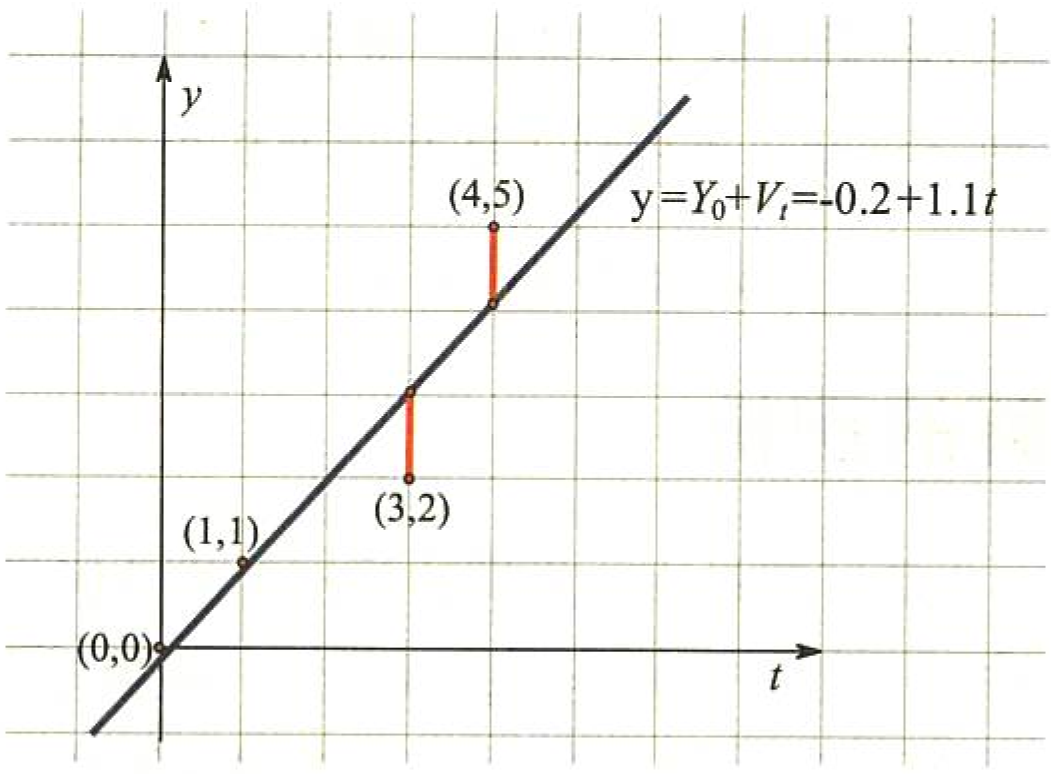


图 6-12 最小二乘法的拟合

6.8 方程组和矩阵、向量组的关系

到此方程组的东东基本搞定，本节简要地总结一下方程组和矩阵、向量组的关系。

6.8.1 线性方程组与矩阵乘法的运算关系

我们可以从线性方程组之间的运算关系来看矩阵之间的运算关系。这种对应关系可以看做矩阵运算的代数意义。

设给出了两组线性方程组

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{cases} \tag{6-10}$$

和

$$\begin{cases} z_1 = b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n \\ z_2 = b_{21}x_1 + b_{22}x_2 + \cdots + b_{2n}x_n \\ \vdots \\ z_m = b_{m1}x_1 + b_{m2}x_2 + \cdots + b_{mn}x_n \end{cases} \tag{6-11}$$

把这两组方程组对应加起来，就是

$$\begin{cases} y_1 + z_1 = (a_{11} + b_{11})x_1 + (a_{12} + b_{12})x_2 + \cdots + (a_{1n} + b_{1n})x_n \\ y_2 + z_2 = (a_{21} + b_{21})x_1 + (a_{22} + b_{22})x_2 + \cdots + (a_{2n} + b_{2n})x_n \\ \vdots \\ y_m + z_m = (a_{m1} + b_{m1})x_1 + (a_{m2} + b_{m2})x_2 + \cdots + (a_{mn} + b_{mn})x_n \end{cases}$$

因此由这组方程组得到的系数矩阵就是方程组（6-10）和方程组（6-11）的系数矩阵的和：

$$\begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

意思是说，矩阵的加法可以认为是来之两组等阶方程组的对应相加。

同样，对方程组（6-10）两边同乘以一个常数 K，就可以得到矩阵数乘的关系式：

$$k\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}$$

同样，矩阵与矩阵相乘的关系就可做看作两个方程组的带入关系，即把方程组（6-11）带入方程组

$$\begin{cases} y_1 = a_{11}z_1 + a_{12}z_2 + \cdots + a_{1m}z_m \\ y_2 = a_{21}z_1 + a_{22}z_2 + \cdots + a_{2m}z_m \\ \vdots \\ y_k = a_{k1}z_1 + a_{k2}z_2 + \cdots + a_{km}z_m \end{cases} \tag{6-12}$$

得到

$$\begin{cases} y_1 = (a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1m}b_{m1})x_1 + \cdots + (a_{11}b_{1n} + a_{12}b_{2n} + \cdots + a_{1m}b_{mn})x_n \\ y_2 = (a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2m}b_{m1})x_1 + \cdots + (a_{21}b_{1n} + a_{22}b_{2n} + \cdots + a_{2m}b_{mn})x_n \\ \vdots \\ y_k = (a_{k1}b_{11} + a_{k2}b_{21} + \cdots + a_{km}b_{m1})x_1 + \cdots + (a_{k1}b_{1n} + a_{k2}b_{2n} + \cdots + a_{km}b_{mn})x_n \end{cases} \tag{6-13}$$

也就是两个矩阵的积的关系是

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{km} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1m}b_{m1} & \cdots & a_{11}b_{1n} + a_{12}b_{2n} + \cdots + a_{1m}b_{mn} \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2m}b_{m1} & \cdots & a_{21}b_{1n} + a_{22}b_{2n} + \cdots + a_{2m}b_{mn} \\ \vdots & & \vdots \\ a_{k1}b_{11} + a_{k2}b_{21} + \cdots + a_{km}b_{m1} & \cdots & a_{k1}b_{1n} + a_{k2}b_{2n} + \cdots + a_{km}b_{mn} \end{bmatrix} \tag{6-14}$$

注意矩阵的乘积的左右顺序，令方程组（6-12）的系数矩阵为 $\boldsymbol{A}$，方程组（6-11）的系数矩阵为 $\boldsymbol{B}$。把方程组（6-11）代入方程组（6-12）得到方程组（6-13），其系数矩阵就是 $\boldsymbol{AB}$。

原来，矩阵乘法实际上就是为了解方程组而发明的算法。

6.8.2 线性方程组、矩阵、向量组的关系

从前面的章节我们看到，对线性方程组的研究，包含了对向量的线性组合和矩阵方程的研究。方程组可以写成矩阵方程和向量组合形式等不同的形式，也反映了我们研究内容的不同。即线性方程组、向量的线性组合和矩阵及其矩阵方程可作等价研究。这样，就把线性方程组的内容、向量的线性组合的内容和矩阵及其矩阵方程的内容统一起来了。下面举一例。

例 6.4 问向量组 $\boldsymbol{\alpha}_1=(1,3,-1)$，$\boldsymbol{\alpha}_2=(2,2,0)$，$\boldsymbol{\alpha}_3=(1,-1,1)$ 是否线性相关。

求解分析：要解决一个具体的向量组是否线性相关，常用的方法是定义法。即是要考察，是否存在一组不全为 0 的数 k_1、k_2、k_3，使得

$$k_1\begin{pmatrix}1\\3\\-1\end{pmatrix}+k_2\begin{pmatrix}2\\2\\0\end{pmatrix}+k_3\begin{pmatrix}1\\-1\\1\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$$

这是一个向量的线性组合式，会感到有点不适应，但如果改写成线性方程组的形式就习惯了：

$$\begin{cases}k_1+2k_2+k_3=0\\3k_1+2k_2-k_3=0\\-k_1+k_3=0\end{cases}$$

即是要考察该齐次线性方程组是否有非零解。如果改写成矩阵方程的形式

$$\begin{bmatrix}1&2&1\\3&2&-1\\-1&0&1\end{bmatrix}\begin{pmatrix}k_1\\k_2\\k_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$$

问题就变成考察系数矩阵的秩是否小于未知量的个数。这样，从线性方程组的解和矩阵的秩都很容易得到向量组 $\boldsymbol{\alpha}_1$、$\boldsymbol{\alpha}_2$、$\boldsymbol{\alpha}_3$ 是线性相关的。从这个分析过程可以看出，我们可以把线性方程组、线性组合、矩阵及其矩阵方程，从形式到内容都统一起来作思考。

6.8.3 秩的关系

在线性代数课程中，秩的概念很重要，下面以线性方程组为基础，总结下矩阵的秩和向量组的秩。

对于一般的线性方程组：

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2\\\qquad\vdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m\end{cases}$$

通常采用消元法求解。在消元过程中，会出现方程组中的若干个方程被消去的情况，剩下的方

程数记为 $r(r \leqslant m)$，这剩下的方程数是固定的，即不管用何种方式消元，数 r 都是不变的，称此为线性方程组的秩。

我们再看矩阵的秩：

矩阵的秩一般是用矩阵中不等于 0 的子式的最大阶数来定义的，如果令矩阵的每行对应一个方程，那么一个矩阵就可对应一个方程组，由前面的讨论可知，矩阵的秩实际上就是它所对应的线性方程组的秩；同时，可由矩阵的初等行变换求矩阵的秩。

矩阵的秩还可以通过其他角度来理解。可逆矩阵的秩与矩阵的阶数相同，且其行列式不为 0。因此，对于方阵，其行列式的值是否为 0 是它的一个重要数量指标。那么，对于一般的矩阵和行列式的值为 0 的方阵有没有一个不为 0 的数量指标呢？矩阵的秩就是此数量指标，一个矩阵不一定有行列式，但是它有子式，我们用不为 0 子式的最高阶数作为矩阵的一个数量指标(即秩)是非常恰当的，它不仅与方程组的秩相容，也和向量组的秩相一致。

矩阵的行向量组的秩称为矩阵的行秩，列向量组的秩称为矩阵的列秩，而矩阵中不等于零的子式的最高阶数可称为矩阵的行列式秩，“三秩合一”，故可统一称为矩阵的秩。另一方面，如果把矩阵行(列)向量组生成的空间称为矩阵的行(列)空间，把矩阵行(列)向量组的最大无关组称为矩阵的基行(基列)，那么矩阵的秩就是行空间、列空间的维数，基行、基列分别是行空间、列空间的基。

而向量组的秩定义是：一个 n 维向量组，它的极大线性无关组所含有的向量个数相同，这个数即为此向量组的秩。

一个向量组即对应一个方程组，这个方程组中通过消元所剩下的方程的最少个数(这个数是固定的)，即为方程组的秩(记为 r)，即在方程组中存在 r 个方程，这 r 个方程可以表示此方程组中所有的方程，并且这 r 个方程不能再少，对应到向量组，这 r 个方程所对应的部分向量组就是该向量组的极大线性无关组。由此，方程组的秩和向量组的秩本质上是一样的，可以用求方程组的秩的方法求向量组的秩，即把向量组看成矩阵的形式，通过初等变换求秩。

综上所述，秩、维数等，在某种意义上说都是实质相同的概念，都可看做向量组极大线性无关组的向量个数的引申。

第7章 二次型的几何意义

二次型是变量的二次乘积项的和，类似于$ax^2+bxy+cy^2$的多项式子，一般包括两个变量乘积项和一个变量平方两种形式。二次型是个函数，如上式可以写成$f(x,y)=ax^2+bxy+cy^2$的等式形式。

有人疑惑：这个多项式不是线性函数啊，这二次型和线性代数有什么关系？

我们知道，一个一元n次多项式可以用一个向量来表示，次数不超过n的多项式全体可以构成一个向量空间。那么一个二次型是一个二元二次多项式，是不是也可以用向量表示出来？基本答对了！这个二元二次型可以用几个向量的联合——矩阵来表示！

如果我们把函数$f(x,y)=ax^2+bxy+cy^2$的自变量换一下，把单变量x、y合成一个变向量$\boldsymbol{x}$，即$\boldsymbol{x}=(x,y)$，那么再引入矩阵，原二元二次函数就可以改写成一元二次向量函数：

$$f(\boldsymbol{x})=(x\quad y)\begin{bmatrix} a & \dfrac{b}{2} \\ \dfrac{b}{2} & c \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

这个样子的向量函数就是传说中的二次型（其中的矩阵要写成对称矩阵）。这个对称矩阵：

$$\begin{bmatrix} a & \dfrac{b}{2} \\ \dfrac{b}{2} & c \end{bmatrix}$$

就可以表示二次型$f(x,y)=ax^2+bxy+cy^2$。

嘿，改写成这个样子后，又有向量又有矩阵，果然是线性代数的内容。一旦改写成这个样子就可以利用矩阵的强大功能了。其实，前面讲的矩阵的各种变换还真不能直接利用，因此人们又发明了矩阵的合同变换应用到二次型的研究上。

其实，二次型的内容就是研究线性空间里的一个几何图形如何在不同的坐标基下的不同的矩阵表示，合同的矩阵表示的是同一个二次函数的几何图形（这可能是“合同”名称的由来）。

二次型的应用广泛，比如信号处理里面的噪声功率、物理学里面的势能和动能、力学里面的惯性张量矩阵、微分几何中曲面的法曲率和统计学里的置信椭圆体等，这些函数都是二次的，都可以转化为二次型对称矩阵的研究。

我们不应觉得奇怪，比如功率是电压电流的乘积二次项，动能的公式$mv^2/2$中有平方项出现。据说二次型起源于解析几何中三维坐标系下二次曲线、二次曲面方程的研究。然而一般的n元二次型的一些概念和理论，比起解析几何中二次曲线及曲面的概念和理论，已变得有点抽象。但由于它们之间的内在联系，使得我们可以将二次曲线及曲面的几何图形用于理解二次型的几何意义。比如，二次型的一个典型的应用就是对诸如椭圆曲面、双曲抛物面之类的二次曲面进行清晰地分类。因此，在正式引入二次型的定义前，俺想从二次曲面图形的绘制开始讨论。

7.1　二次曲线及曲面的图形

7.1.1　二次函数的哪些系数对图形是重要的

在中学的解析几何中，对于一元二次函数 $f(x)=ax^2$ 在二维平面上看（把函数值看做 y 轴：$y=ax^2$）我们有非常清晰的几何图形，就是过原点的抛物线：$a>0$ 开口向上，整个曲线在横坐标之上；$a<0$ 开口向下，整个曲线在横坐标之下；$a=0$，抛物线蜕化成直线 x 轴。

那么对于函数 $f(x)=ax^2+c$ 图形呢？我们依然清晰地知道，这是一条在纵轴方向平移后的图形，仍是一条抛物线或直线，偏离原点了，但图形仍然相同。

那么对于函数 $f(x)=ax^2+bx+c$ 图形呢？虽然不是那么清晰了，但我们依然知道这也是一条偏离原点的抛物线或直线，只不过在横轴和纵轴方向上同时进行了平移，但图形仍然相同。因为可以把它经过配方得到上面的形式：

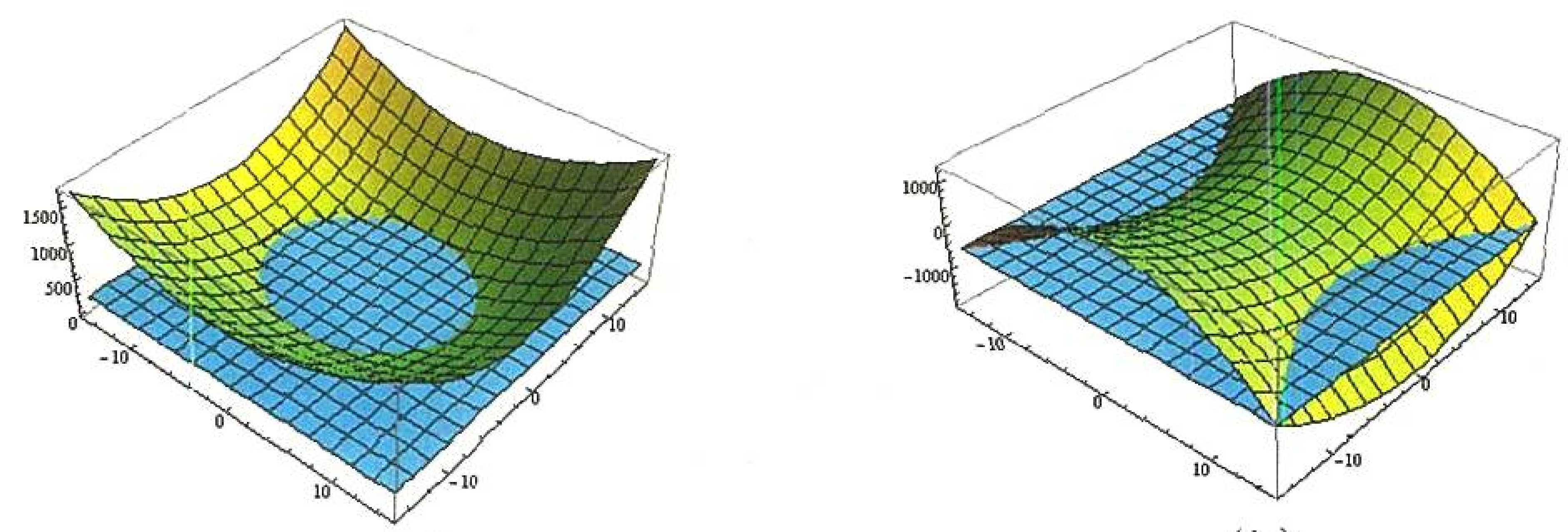
$$f(x)=ax^2+bx+c=a(x+\frac{b}{2a})^2+(c-\frac{b^2}{4a})$$

到这里我们有一个小小的结论，就是三个函数的几何图形是合同的，合同的意思是它们经过平移后图形会完全重合。

既然是这样，如果是要考察几何图形的话，对于 $f(x)=ax^2+bx+c$ 类的函数我们只要考察 $f(x)=ax^2$ 的形式就可以了。换句话说，我们只需研究这个二次函数的二次项。

对于二元二次函数，同样有类似的情况。

二元函数 $f(x,y)=ax^2+cy^2$ 的图形在三维空间中观看（即 $z=ax^2+cy^2$），其典型的图形是椭圆抛物面或双曲抛物面（见图 7-1，至于其退化图形如抛物柱面等我们不再提及），其中的双曲抛物面俗称马鞍面，如图 7-1（b）所示。



（a）　　　　（b）

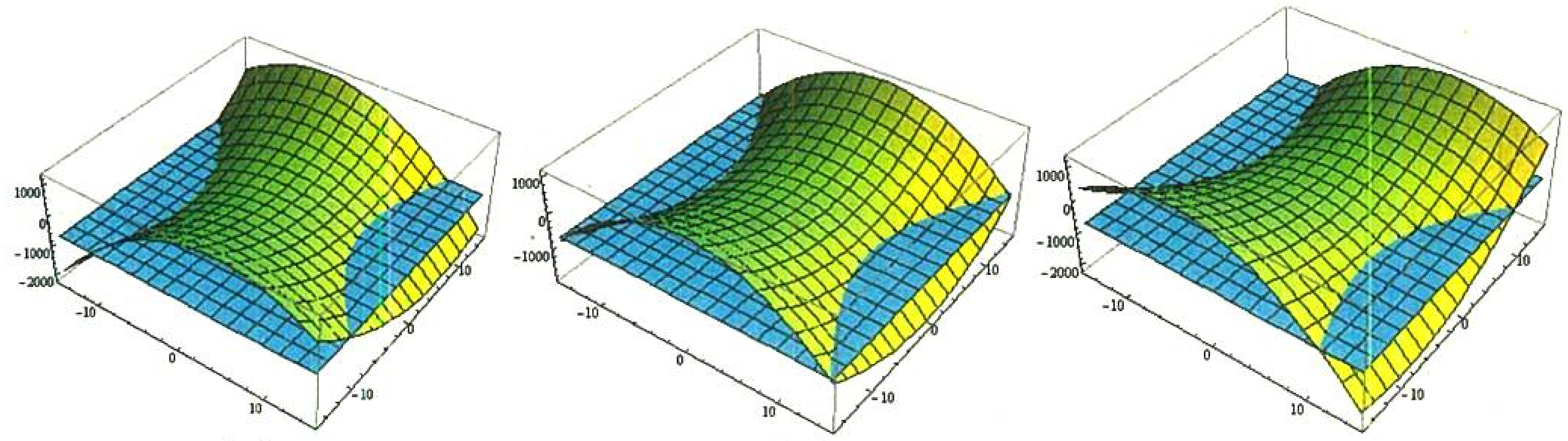
图 7-1　椭圆抛物面或双曲抛物面

在图中，添加了一个平行于 xoy 平面的平面。平面与曲面的相交线让我们清晰地看出，椭圆抛物面的交线是椭圆，双曲抛物面的交线是双曲线（之所以称之为抛物面，是因为曲面与 xoz、yoz 平行平面的交线是抛物线）。

对于添加了交叉项的二元函数 $f(x,y)=ax^2+bxy+cy^2$ 的图形，在三维空间中观看（即令 $z=ax^2+bxy+cy^2$），其典型的图形同样是椭圆抛物面或双曲抛物面，只是它们的图形在坐标

系中的位置发生了旋转。从上面（z 轴反方向）看下去，曲面因系数 b 的正负值的不同发生了左右旋转。

比如上面的双曲抛物面图形，如果我们改变系数 b 的值，通过与平面的截线的变化看出曲面在旋转，图 7-2(a)到(c)的图形变化是改变交叉项 bxy 的系数 b 由-5→0→5 变化得到的。可以看出，双曲抛物面向右进行了旋转，同时曲面的形状发生了一些不明显的伸缩变化（再伸缩也是双曲抛物面）。另外可以看出，当 b=0 即 $f(x,y)=ax^2+cy^2$ 时，曲面的图形相对于 x 和 y 轴是对称的（见图 7-2(b)），旋转后图形明显与坐标轴不再对称了（见图 7-2(c)）。

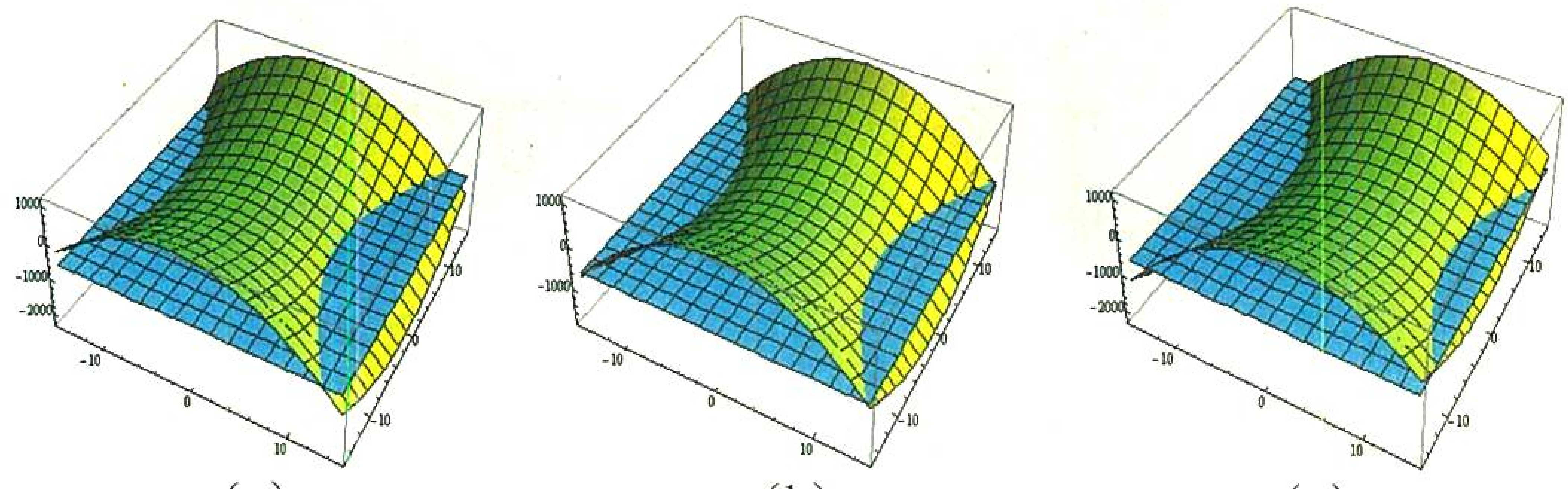


（a）　　　　（b）　　　　（c）

图 7-2 双曲抛物面随着函数项 bxy 的系数 b 由-5→0→5 变化而旋转

我们再来看看一般的二元二次函数 $f(x,y)=ax^2+bxy+cy^2+dx+ey+f$ 的图形，函数式与前者比较看来，式中增添了一次项和常数项。根据一元二次函数的经验，增添了一次项和常数项的函数图形只是把原来的图形在各个坐标轴 (x,y,z) 方向上分别平移了一个距离。对于二元函数来说同样如此。

比如双曲抛物面图形，如果我们改变函数系数 d 的值，通过与平面的截线的变化看出曲面在沿着 x 轴的方向（图 7-3 中左右方向为 x 轴方向）上移动，图 7-3(a)到(c)的图形变化是改变一次项 dx 的系数 d 由-30→0→30 变化得到的。另外，也可以看出，当 d=0 时，曲面的图形相对于 y 轴是对称的（见图 7-3(b)），平移后图形明显与坐标轴不再对称了（见图 7-3(c)）。



（a）　　　　（b）　　　　（c）

图 7-3 双曲抛物面随着改变系数 d 的值其图形沿着 x 轴的方向左右移动

通过以上的图形可以看出，由二次函数的二次项的系数就可以完全看出图形的类型。一次项和常数项并不能影响图形类型的本身。至此我们可以有个总结就是，如果考察几何图形，对于 $f(x,y)=ax^2+bxy+cy^2+dx+ey+f$ 类的函数，只需考察 $f(x,y)=ax^2+bxy+cy^2$ 的形式就可以了。换句话说，我们只研究这个二次函数的二次项就可以了。

函数图形的合同

二元二次函数的几何图形是否合同呢？包含变量交叉项的函数的图形较不包含交叉项的函数图形发生了旋转和伸缩变化，如果只是旋转，图形是一样合同的；如果图形再伸缩变化，就不一定了，比如原来同一平面的截线是一个圆，伸缩变化后变成了一个大圆或一个小圆甚至变成了椭圆形，这还叫合同吗？

看起来几何图形不是合同的：大圆与小圆硬说是合同还勉强些，但椭圆和圆无论如何不能重合啊。看来合同的含义有问题啊。

这个问题可以解决，解决的方法是先假定几何图形是不变的，类同一个刚体，函数式变了的原因是因为坐标系变了……嘿嘿，这样就可以自圆其说了——可以把椭圆说成是圆——都是同一个刚体的在不同坐标系下的像……这其实正是二次型的内涵，我们后面慢慢谈起。

7.1.2 二次函数与二次方程的关系

二次型是一个函数，给定这个函数一个值，就成为二次曲面方程了。

当然，函数有几何图形，方程也有几何图形。两者图形之间的关系是：方程的几何图形是函数几何图形与直线、平面或超平面的截线。

比如二元二次型函数：

$$f(x,y)=\frac{1}{4}x^2+y^2 \tag{7-1}$$

其几何图形绘制在三维坐标系 $\{o,x,y,z\}$ 中是一个椭圆抛物面。令 $f(x,y)=1$，则截痕方程 $\frac{1}{4}x^2+y^2=1$ 在二维坐标系 $\{o,x,y\}$ 中的图形是椭圆。这个椭圆就是椭圆抛物面与平面 $z=1$ 的截线投影到到 xoy（或 z=0）平面上的图形，它与在平面 $z=1$ 上的椭圆完全相同，即

$$\begin{cases} z=f(x,y)=\frac{1}{4}x^2+y^2 \\ z=1 \end{cases} \Rightarrow \frac{1}{4}x^2+y^2=1$$

其图形见图 7-4。

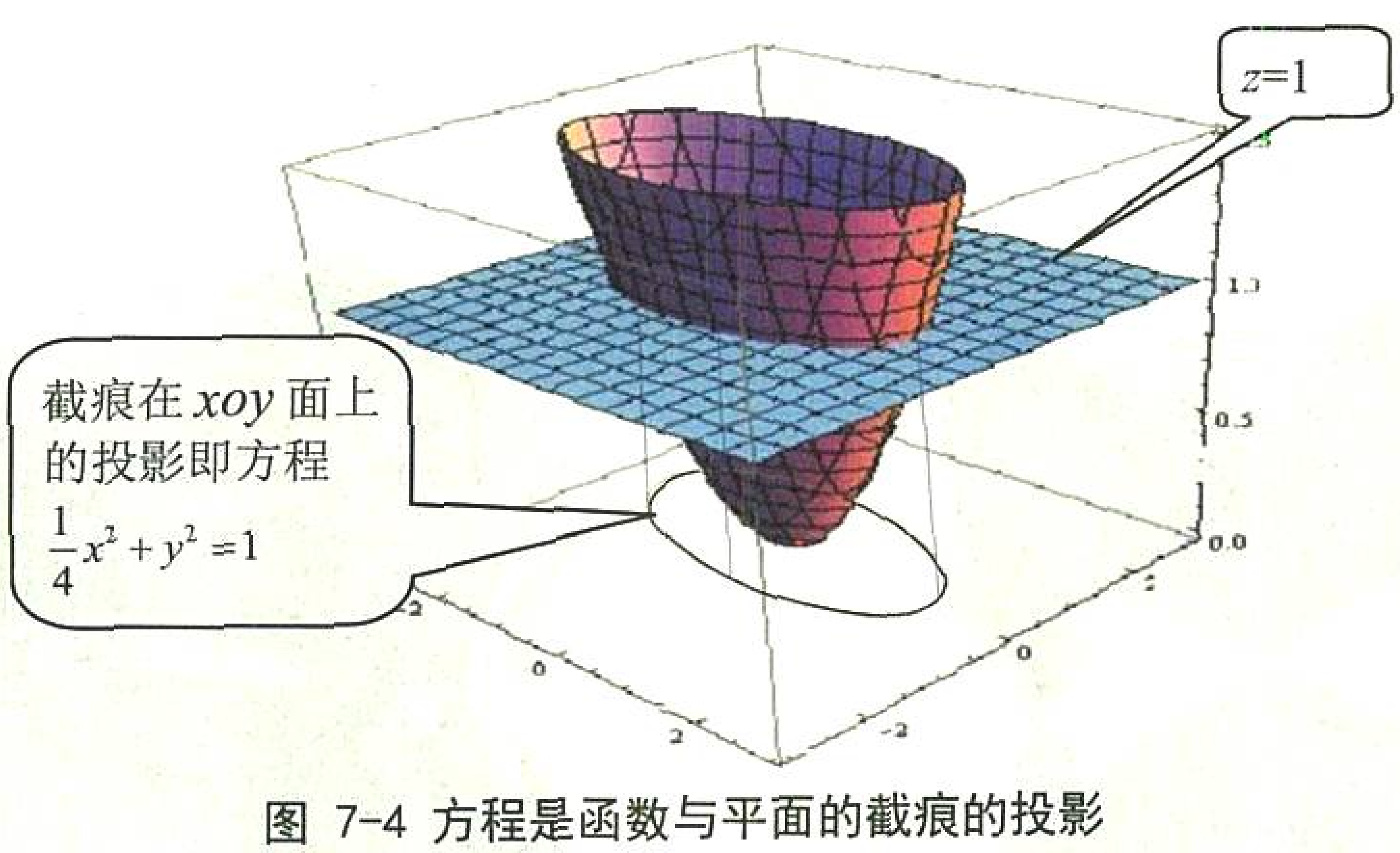


图 7-4 方程是函数与平面的截痕的投影

如果我们是二维的平面动物，将无法看到二元二次型函数的 3D 全貌，但我们可以通过用平行平面多次截图，并把截线投影到二维平面上来想象三维的图形。

比如函数式(7-1)的几何图形四次均匀截线在 $\{o,x,y\}$ 平面上的投影是同心椭圆（见图 7-5），由此我们可以在二维空间近似地想象三维空间里的椭圆抛物曲面的庐山真面目。

由此可以推广为：

如果数域 F 上一个 n 元二次型是 $f(x_1,x_2,\ldots,x_n)$，则方程 $f(x_1,x_2,\ldots,x_n)=d$ 的图形就是：定义在 F^n 的一个二次函数 $z=f(x_1,x_2,\ldots,x_n)$，它在 n+1 维空间 F^{n+1} 的图形（二次超曲面）与平面 $z=d$ 相截交，此截交线在 F^{n+1} 的 n 维子空间——“坐标平面” $\{o,x_1,x_2,\cdots,x_n\}$ 上的正投影（超曲线）的图形。

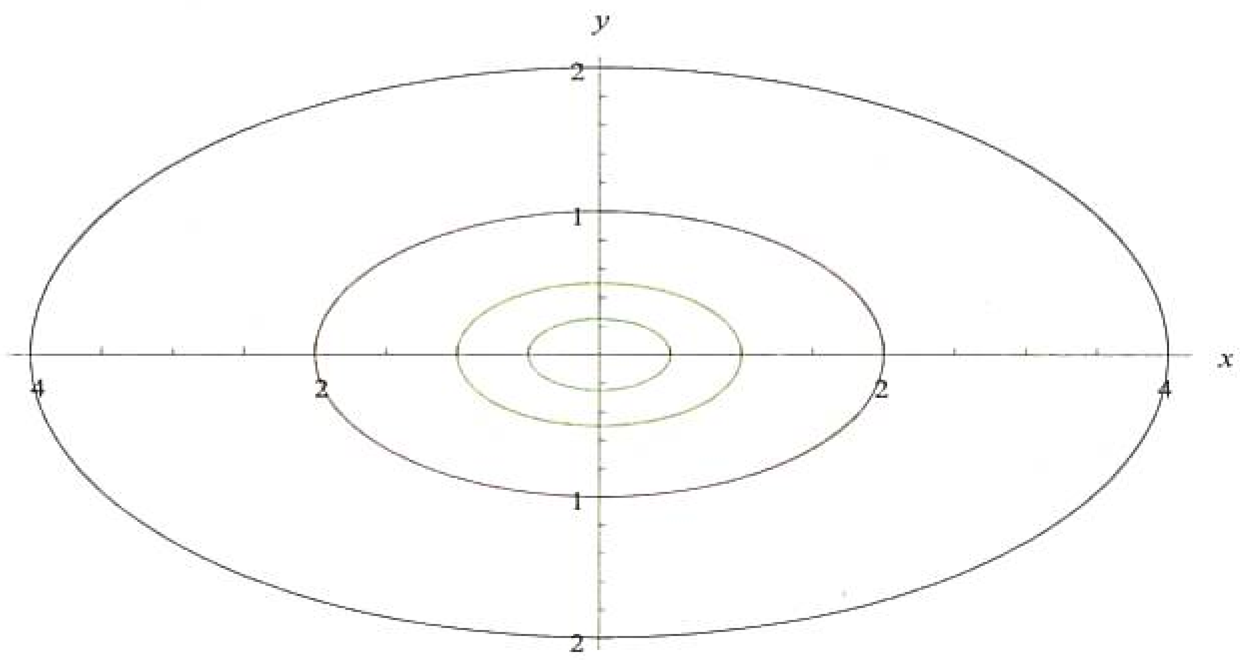


图 7-5 三维图形的平面截线在二维平面上的投影

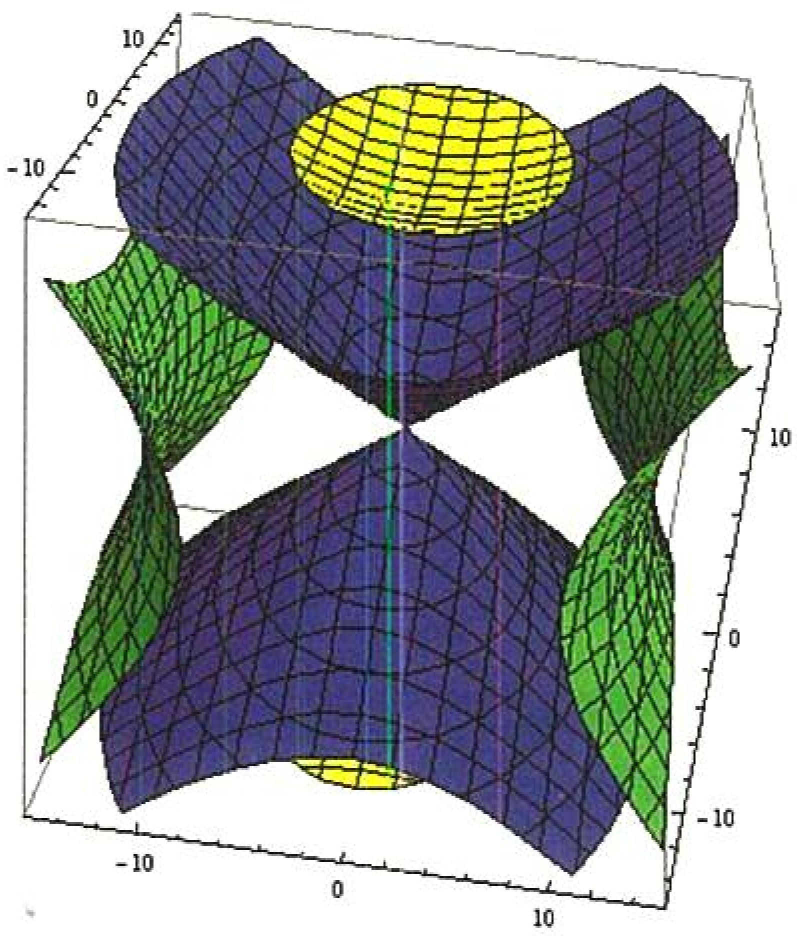
用截线投影的方法可以绘出更高维的几何图形，比如四维图形。

如上节二元函数值的图形刚好在三维空间里观察。对于三元二次函数，它的函数值只能描绘在第四维坐标方向的空间上。但我们是三维动物，应如何看四维的图形？

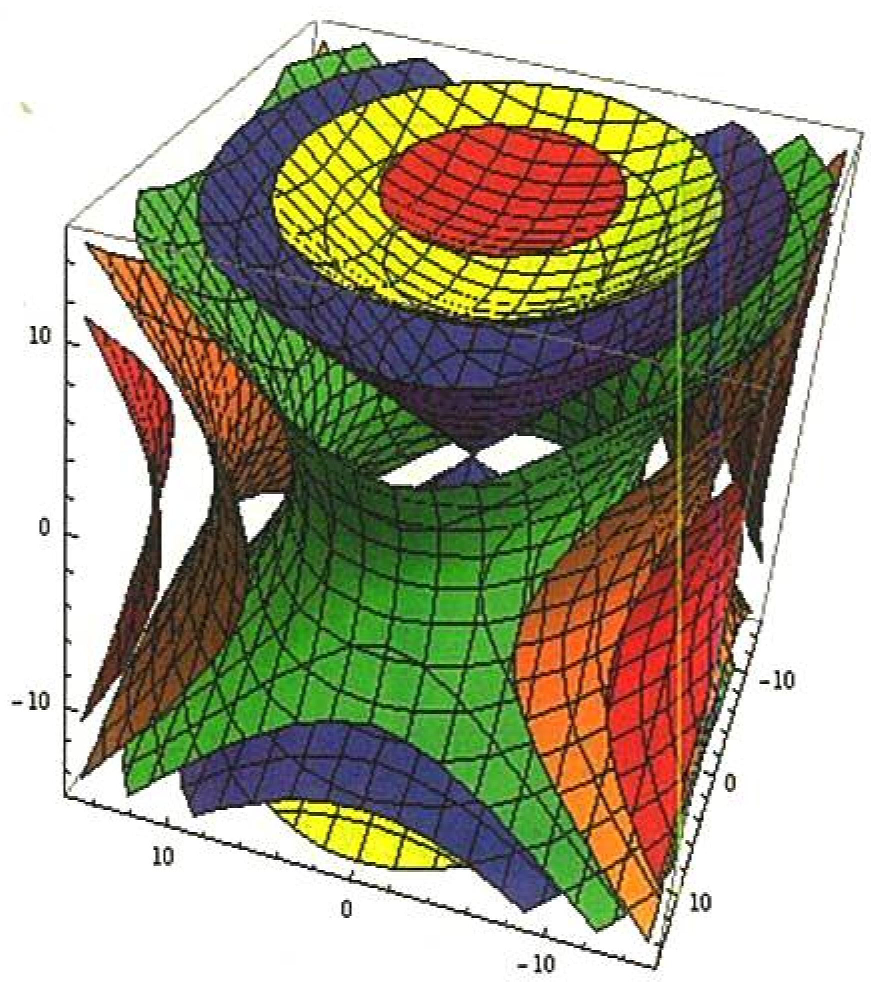
类似地，方法是，对于三元二次函数 $\varphi=f(x,y,z)$，我们在一个三维坐标空间里同时绘制出 φ 为不同数值的方程图形，比如令 φ 为负数、零和正数时的三维图形，这些三维图形正是四维截痕图形的三维投影，把这些投影大范围地组合起来，就可以想象三元二次函数的四维图形了。或者干脆就叫这些投影组合为类四维几何图形。

稍具体来说，如果我们假设三元二次函数 $\varphi=ax^2+by^2+cz^2$ (a、$b>0$, $c<0$)，当 $\varphi<0$ 时是两个椭圆抛物面，当 $\varphi=0$ 时是两个圆锥面，当 $\varphi>0$ 时是两个双曲抛物面，则这三个部分曲面构成了一个典型的类四维曲面图形（见图 7-6（a））。

如果 φ 取更多的值，则我们可得到更多的图形，进而取得三元二次函数非常丰满的类四维图形的概念（见图 7-6（b））。



（a）



（b）

图 7-6 某种三元二次函数的类四维图形

遐想

从图中我们体会到这三个典型的曲面之间存在着许多过渡曲面。这些曲面与某一平面相截将得到各类的所谓“圆锥曲线”，与其中的圆锥相截当然得到圆锥曲线，与其他的曲面相截也得到圆锥曲线。在这里我们扩展了圆锥曲线的几何内涵。

7.1.3 圆锥曲线的向量方程

这一节咱回顾一下高中时就接触到的著名的圆锥曲线，讨论它们的向量方程和图形是如何的。

二维平面空间中的几何图形比如圆、椭圆、抛物线及双曲线等也属于二次曲线图形，这些曲线又称为圆锥曲线，因为它们可以看成是由平面和双圆锥面相交而得到的截线，见图 7-7。

双圆锥面和平面相截交的方程组是

$$\begin{cases} f(x,y)=ax^2+2bxy+cy^2+dx+ey+f \\ f(x,y)=g \end{cases}$$

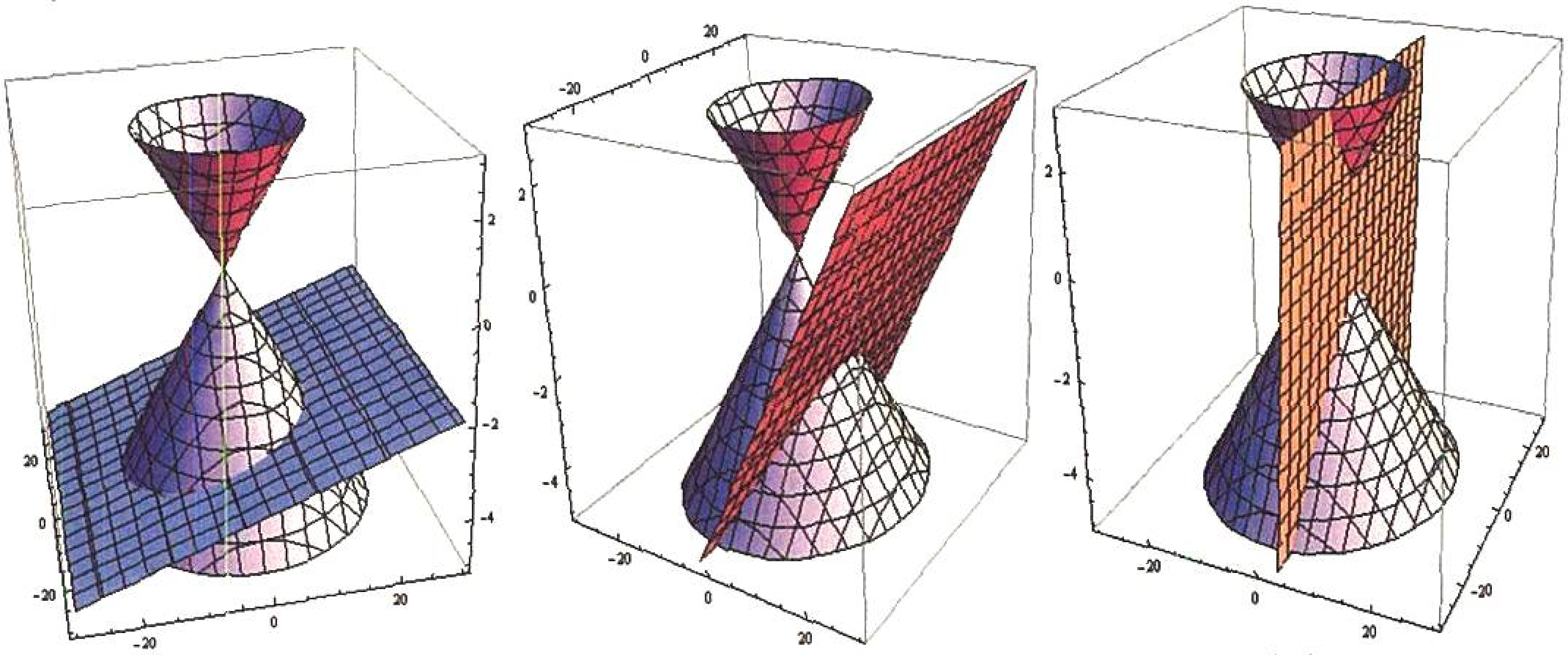
其中，$f(x,y)=g$ 就是那个截平面方程。由此，平面上的二次曲线的一般形式为

$$ax^2+2bxy+cy^2+dx+ey+f=0 \tag{7-2}$$

那么此方程的向量及矩阵的形式可以是

$$(x,y)\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+(d,e)\begin{pmatrix} x \\ y \end{pmatrix}+(f)=(0)$$

（注：xy 的系数写为 $2b$ 是为了得到矩阵方程的一般形式。）



（a）　　（b）　　（c）

图 7-7 椭圆、抛物线、双曲线是圆锥曲面的平面截线

根据我们中学的经验，如果把方程（7-2）化简为 $a'x'^2+c'y'^2=f'$ 的形式，我们就比较容易看出来是椭圆还是双曲线等。下面我们对方程进行分类分析，看如何化简成想要的形式。

对于不包含交叉项的方程（$b=0$时）$ax^2+cy^2+dx+ey+f=0$，我们可以对它进行配方得到 $a'(x+g)^2+c'(y+h)^2=f'$ 的形式，并对它进一步进行平移变换：

$$\begin{cases} x'=x+g \\ y'=y+h \end{cases} \text{或} \begin{pmatrix} x' \\ y' \end{pmatrix}=\begin{pmatrix} x \\ y \end{pmatrix}+\begin{pmatrix} g \\ h \end{pmatrix}$$

平移变换使曲线的对称中心平移到坐标原点，进而得到想要的形式 $a'x'^2+c'y'^2=f'$。

对于只剩下二次项和常数项的方程形式，也就是二次曲线的对称中心与坐标原点重合时，

其方程形式为

$$ax^2+2bxy+cy^2=f$$

为了判别该曲线的类型，需要作坐标旋转，消去乘积项$2bxy$，只含平方项。引入旋转线性变换：

$$\begin{cases}x=x'\cos\theta-y'\sin\theta\\ y=x'\sin\theta-y'\cos\theta\end{cases} \text{或} \begin{pmatrix}x\\y\end{pmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & -\cos\theta\end{bmatrix}\begin{pmatrix}x'\\y'\end{pmatrix}$$

代入二次曲线方程，简化得到想要的方程形式$a'x'^2+c'y'^2=f'$。

实际上，对于完整的方程（7-2）的形式，应该先进行旋转变换，消掉交叉项，然后再进行平移变换，消掉一次项，才能顺利地得到平方项形式的方程。

以上是中学的处理方法，适用于低维的空间。随着二次函数的变量的增加，函数变得更加复杂起来。为了方便研究二次函数，需要使用向量和矩阵工具来分析。比如，方程是高维空间的函数如下：

$$a_1x_1^2+a_2x_2^2+\cdots+a_nx_n^2+2b_{12}x_1x_2+2b_{13}x_1x_3+\cdots+2b_{n-1,n}x_{n-1}x_n+c_1x_1+c_2x_2+\cdots+c_nx_n+f=0$$

一次线性部分仍然可以配方法处理，但首先二次交叉项如何处理呢，如何确定这个高维空间里的旋转变换呢？好复杂，类似前面的写法，把它改写成向量方程就简洁多了：

$$(x_1,x_2,\cdots,x_n)\begin{bmatrix}a_1 & b_{1,2} & \cdots & b_{1,n-1}\\ b_{1,2} & a_2 & \cdots & b_{2,n-1}\\ \vdots & \vdots & & \vdots\\ b_{1,n-1} & b_{2,n-1} & \cdots & a_n\end{bmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}+(c_1,c_2,\cdots,c_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}+(f)=(0)$$

等式左边这三部分当中，最左边的部分是最复杂、最重要、最需要研究的，我们把它拿出来专门研究。这么复杂的式子是否仍然可以用旋转变换消掉交叉项呢？答案是肯定的。这就是我们要进一步讨论的二次型的主要内容之一。

7.2 二次型及其几何意义

二次型就是不含一次项的二次函数。先来看 n 个变量的二次型的定义。

7.2.1 二次型的定义

含有 n 个变量 $x_1,x_2,\ldots,x_n$ 的二次齐次函数：

$$f(x_1,x_2,\cdots,x_n)=a_{11}x_1^2+a_{22}x_2^2+\cdots+a_{nn}x_n^2+2a_{12}x_1x_2+2a_{13}x_1x_3+\cdots+2a_{n-1,n}x_{n-1}x_n$$

称为二次型。只含有平方项的二次型称为二次型的标准形：

$$f(x_1,x_2,\cdots,x_n)=a_{11}x_1^2+a_{22}x_2^2+\cdots+a_{nn}x_n^2$$

二次型实际上是n元二次齐次多项式。为了使用矩阵作为分析二次型的工具，我们找到了二次型和对称矩阵一一对应的关系。实际上，二次型可以这样表示为变元向量和矩阵的乘积：

$$f(x_1,x_2,\cdots,x_n)=(x_1,x_2,\cdots,x_n)\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{12} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{1n} & a_{2n} & \cdots & a_{nn}\end{bmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$$

二次型中二次项的系数构成了矩阵，这个构成如下对应关系：

$$a_{11}x_1^2+a_{22}x_2^2+\cdots+a_{nn}x_n^2+2a_{12}x_1x_2+2a_{13}x_1x_3+\cdots+2a_{1n}x_1x_n+2a_{23}x_2x_3+\cdots+2a_{n-1,n}x_{n-1}x_n$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{nn} \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

$$(x_1,\ x_2,\ x_3\ \cdots,\ x_n)$$

上述对应关系显示，在写二次型的矩阵时，要将交叉项 x_ix_j 的系数 a_{ij} 一分为二，分别放在矩阵的第 (i,j) 和 (j,i) 的位置上，而平方项 x_i^2 的系数 a_{ii} 就直接放在矩阵的对角线上的 (i,i) 位置上。

显然，二次型的矩阵将是一个实对称矩阵。有人会问，为啥要把交叉项的系数一分为二的放？因为这样处理我们会得到一个对称矩阵，对称矩阵有很优良的性质，对处理二次型方便很多。再说了，不这样把系数对半分的话，一个二次型将有无数个矩阵相对应，麻烦大了。

可见，任给一个二次型，就唯一确定一个**对称矩阵**；反之，任给一个对称矩阵，也就唯一确定一个二次型。其中矩阵 $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$ 被称为二次型 $f(x_1,x_2,\cdots,x_n)$ 的矩阵。

改写其为矩阵形式的表达式 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 是为了方便。那么，今后对二次型的研究就归结到对实对称矩阵的研究上来。

7.2.2 二次型的几何及物理意义

二次型的几何图形是二次曲线或曲面，多元的二次型图形是超二次曲线或曲面，这也算是二次型的几何意义，不过是解析意义而不是向量意义，它们是向量末端集合的图形。如果在复数域上讨论向量，则可以说二次型的几何意义是向量的长度的平方，而且就是向量在不同坐标系下长度的平方。我们来看：

n 维向量在标准正交基下的坐标式为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$，那么它的长度的平方就是

$$|\boldsymbol{x}|^2=\left(\sqrt{x_1^2+x_2^2+\cdots+x_n^2}\right)^2=x_1^2+x_2^2+\cdots+x_n^2$$

这就是二次型的规范形。

这个向量如果放在一个保持单位和坐标轴正交关系都不变的新坐标系(基的正交变换)下，那么向量长度重新表示为

$$|\boldsymbol{x}|^2=(x'_1)^2+(x'_2)^2+\cdots+(x'_n)^2={x'_1}^2+{x'_2}^2+\cdots+{x'_n}^2$$

这也是二次型的规范形。

这个向量如果放在一个保持坐标轴正交关系不变但单位不同的新坐标系下，那么向量长度重新表示为

$$|\boldsymbol{x}|^2=(a_1x'_1)^2+(a_2x'_2)^2+\cdots+(a_nx'_n)^2=a_1^2{x'_1}^2+a_2^2{x'_2}^2+\cdots+a_n^2{x'_n}^2$$

这就是二次型的标准形。

这个向量如果放在一个任意的 n 维新坐标系下，这个新基的度量矩阵为 $\boldsymbol{S}$，那么向量长度由推广的内积定义重新表示为

$$|\boldsymbol{x}|^2=\left(\sqrt{\boldsymbol{x}'^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x}'}\right)^2=\boldsymbol{x}'^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x}'=(x'_1,x'_2,\cdots,x'_n)\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{12} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{1n} & a_{2n} & \cdots & a_{nn}\end{bmatrix}\begin{pmatrix}x'_1\\ x'_2\\ \vdots\\ x'_n\end{pmatrix}$$

这就是二次型的一般形式。

一句话，二次型是向量的长度的平方，这个平方值不随坐标基的变化而变化，是个不变量，是绝对性的，而二次型函数的数学表达式却随着坐标基的变化而变化。

那么二次型有物理意义吗？是的，有对应的物理解释。

因为二次型和广义内积的联系，咱自然而然地把它的物理意义与距离以及能量联系在一起。

二次圆锥曲线首先是和距离联系在一起的。比如：圆的轨迹是与到原点的距离为恒常数联系在一起的，用向量函数可表示为 $(x,y)\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=f$；椭圆的轨迹是与到两个焦点的距离的和是恒常数联系在一起的，用向量函数可表示为：$(x,y)\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=f$；双曲线的轨迹是与到两个焦点的距离的差是恒常数联系在一起的，用向量函数可表示为 $(x,y)\begin{bmatrix}a & 0\\ 0 & -b\end{bmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=f$；抛物线的轨迹是到一个点（焦点）的距离等于到一根直线（准线）的距离的点联系在一起的，用向量函数可表示为 $(x,y)\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=(d,0)\begin{pmatrix}x\\ y\end{pmatrix}$，即 $(x,y)\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\begin{pmatrix}x\\ y\end{pmatrix}+(-d,0)\begin{pmatrix}x\\ y\end{pmatrix}=0$。

实际上，上述的向量函数不是直接与距离的量联系在一起，而是与距离的乘积或距离的平方联系在一起的，或者说向量函数实际上是与内积的概念直接联系的。

前面讲过，一般的二次型向量方程如下：

$$(x_1,x_2,\cdots,x_n)\begin{bmatrix}a_1 & b_{1,2} & \cdots & b_{1,n-1}\\ b_{1,2} & a_2 & \cdots & b_{2,n-1}\\ \vdots & \vdots & & \vdots\\ b_{1,n-1} & b_{2,n-1} & \cdots & a_n\end{bmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}+(c_1,c_2,\cdots,c_n)\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}+(f)=(0) \tag{7-3}$$

这个方程细看起来有点意思：

(f) 是一个实数值；

$(c_1,c_2,\cdots,c_n)\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}$ 的形式就是内积运算，是常向量和变向量的内积，结果是一个实变量；

$(x_1,x_2,\cdots,x_n)\begin{bmatrix}a_1 & b_{1,2} & \cdots & b_{1,n-1}\\ b_{1,2} & a_2 & \cdots & b_{2,n-1}\\ \vdots & \vdots & & \vdots\\ b_{1,n-1} & b_{2,n-1} & \cdots & a_n\end{bmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}$ 的形式也是内积运算，是同一个向量与自己的内积（即长度的平方），确切的讲是推广的内积运算（参见第四章的内积推广部分），结果也是一个实变量；

如果把 (f) 也看做内积，那么这三部分都是内积，方程（7-3）就是三个内积和等于 0。

因此二次型就具有了内积所具有的物理意义，比如向量的元素都是速度的值，那么二次型就有总能量的意义。让二次型等于一个实数，就是说，向量运动总是保持这个值不变——能量守恒，这样二次型的图形就是能量守恒时向量的运动的轨迹——难怪圆锥曲线可以描述天体运动的规律。难怪狭义相对论中的度规不变量是一个二次型（进一步的解释放在了本章的最后一节）。

如果在复向量空间里定义内积，那么二次型的对称矩阵也就完全类同于内积的度量矩阵的含义了。

其实二次型和内积都属于一种更广泛意义的函数——双线性函数。如用双线性来理解二次型，我们才能安然地接受二次型也属于线性代数的研究范畴。

7.2.3 二次型函数与双线性函数的关系

就像本章开头提到的疑问：线性代数主要研究“线性”问题，如线性变换、线性空间、线性方程组等，二次型却是非线性的，形如 $f(x,y)=ax^2+bxy+cy^2$，为何能用矩阵等线性工具研究？实际上，可以从下面几个方面来理解：

一是，二次曲线或曲面本身就具有线性的性质，比如直纹曲面，一个坐标方向是直线，一个坐标方向是二次的，就可以描绘出直纹曲面（参见图）。

二是，用向量研究一元多项式，是用多项式的系数构成向量来等价地替代研究，不是直接研究多项式本身，因此化非线性为线性；同样，用二次型的系数构建矩阵来等价地替代研究它，同样化非线性为线性了。

三是，对二次型的研究利用了双线性函数的概念。二次型本来就是一个对称的双线性函数。双线性函数也是线性函数啊，既然是线性函数，当然可以用线性代数的向量及矩阵工具进行研究了。那么什么是双线性函数？

简单点说，双线性函数就是定义了某维线性空间里的双向量（两个向量）的一个运算，运算结果是一个数，这个数属于某个数域（如实数、复数域等）。比如，双向量的内积就是一个双线性函数。双线性具体是嘛意思呢？我们看 n 维线性空间下的内积运算式：

$$<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=x_1y_1+x_2y_2+\cdots+x_ny_n$$

如果固定一个向量比如 $\boldsymbol{x}$，把向量 $\boldsymbol{x}$ 改写成常向量 $\boldsymbol{x}=(k_1,k_2,\cdots,k_n)$，则内积变为

$$<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=k_1y_1+k_2y_2+\cdots+k_ny_n$$

这是一个线性函数。

相对地，如果只固定向量 $\boldsymbol{y}$，同样也是一个线性函数。或者说，其中一个变元固定时是另一个变元的线性函数，两个向量互为线性，称之为双线性。

前面讲过，推广了的内积定义是：

$$<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{y}$$

其中 $\boldsymbol{S}$ 是度量矩阵，是对称的正定方阵。

推广了的内积也是特殊的双线性函数，因为这个定义与一般的双线性函数的定义式形式是相同的，只是双线性函数矩阵 $\boldsymbol{S}$ 是一般的 n 阶方阵而已。

双线性函数的一般定义如下：

设 P^n 是数域 P 上 n 维列向量构成的线性空间，向量 $\boldsymbol{x},\boldsymbol{y}\in P^n$，再设 $\boldsymbol{A}$ 是 P 上的 n 阶方阵。令

$$f(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$$

则 $f(\boldsymbol{x},\boldsymbol{y})$ 是 P^n 上的一个双线性函数。

现在看二次型，二次型的矩阵表示式是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，显然二次型是双线性函数。实际上，二次型不过是合并变量的对称双线性函数。所谓对称的就是满足 $f(\alpha,\beta)=f(\beta,\alpha)$ 的要求，因此矩阵 $\boldsymbol{A}$ 是对称矩阵，这就与二次型的对称矩阵相一致了。不过这还不够，因为对称双线性函数的表示式仍然是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$，还需要把变向量合成一个 $\boldsymbol{x}$ 才称其为二次型。

下面我们画出两个简单的对称双线性函数及其对应的二次型的图形。希望能看到二次型的线性意义出来。

一维线性空间里，最简单的对称双线性函数为 $f(\boldsymbol{x},\boldsymbol{y})=(\boldsymbol{x})[1]\boldsymbol{y}=\boldsymbol{x}\boldsymbol{y}$，对应的二次型函数为 $f(\boldsymbol{x})=(\boldsymbol{x})[1]\boldsymbol{x}=\boldsymbol{x}^2$，它们的图形见图 7-8，可以看出，对称双线性函数是一个双曲抛物面，俗称马鞍面。

从图 7-8 中可以看出马鞍面的双线性出来：当用垂直于 x 轴的平面去截马鞍面时得到一个个直线（见图 7-8（a），就是固定 x 变量，令 $x=k$ 时的函数 $f(k,y)=ky$ 的图像）；同样地，当用垂直于 y 轴的平面去截马鞍面时得到另外一个个直线（见图 7-8（b），就是固定 y 变量，令 $y=k$ 时的函数 $f(x,k)=kx$ 的图像）。

从图上也可以看出马鞍面包含的二次型出来：当 $y=x$ 时，平面 $y=x$ 所截取的曲线图形就是抛物线 $f(x)=x^2$，这就是二次型的图形了（如图 7-8（c）所示）。

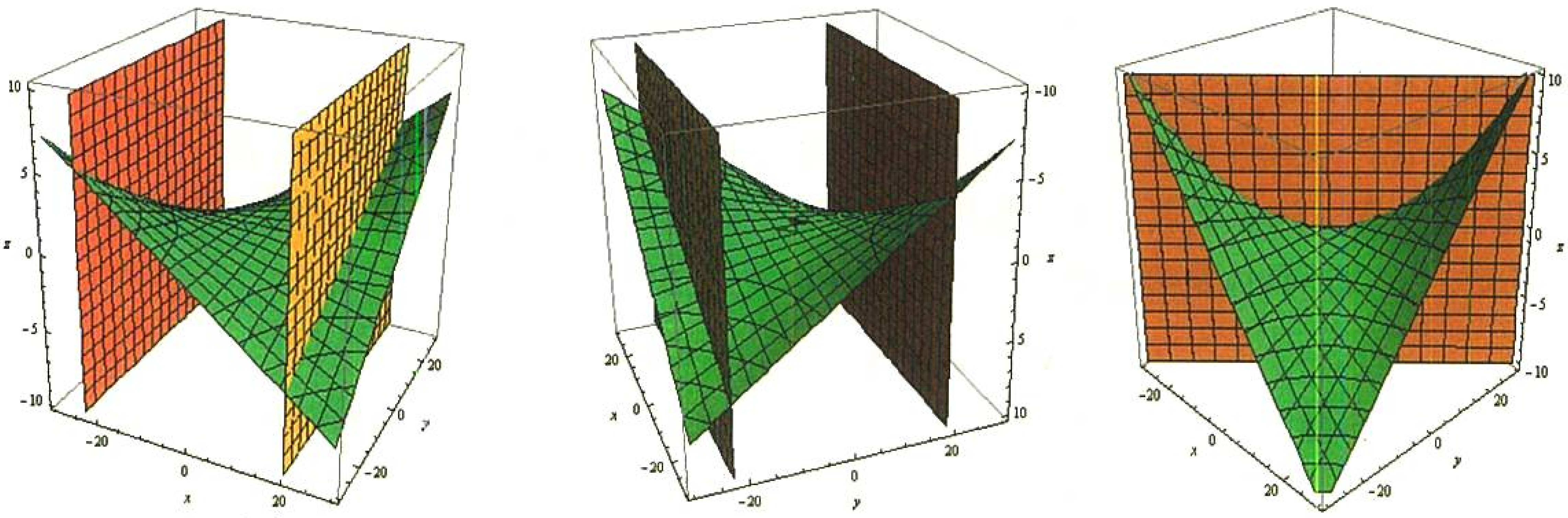

（a）（b）（c）

图 7-8 马鞍面的双线性和二次曲线性

二维线性空间里，最简单的对称双线性函数为 $\begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix}=x_1y_1+x_2y_2$，其二次型函数为 $\begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=x_1^2+x_2^2$。

二维空间里的双线性函数的图形是一个五维的几何图形，即使画出来这个图形也是一团乱麻似的难以看出几何图形的规律。若想一窥其真容，只能用切片法来想象一下了。

前面讲的一维双线性函数的图形是马鞍面，假设我们不知道这个一维双线性函数的图形是

什么，如果用切片法的话就是固定一个变量 x，垂直 x 坐标轴切出一个“直线”图片；再固定一个变量 y，垂直切出另一个“直线”图片；然后竖直斜切，切出一个抛物线；再水平斜切，切出一个双曲线。我们把这些切片组合就形成了一个天马行空的马鞍面出来。

好，如法炮制，切出二维线性空间里的双线性函数的图形片片：

固定一个向量 $\boldsymbol{x}=(x_1,x_2)=(k_1,k_2)$，就得到一个平面 $f(\boldsymbol{k},\boldsymbol{y})=k_1y_1+k_2y_2$；另外固定一个向量 $\boldsymbol{y}=(y_1,y_2)=(k_1,k_2)$，就得到另一个平面 $f(\boldsymbol{x},\boldsymbol{k})=k_1x_1+k_2x_2$；然后竖直斜切，切出一个抛物面；再水平斜切，切出一个马鞍面……

直纹面：

从这里我们终于明白了马鞍面为何也称为一种直纹面——就是可以用巨多的直线织成啊——神奇，用直线可以丝毫不差地构成一个变化多端的曲面，如图 7-9 所示。

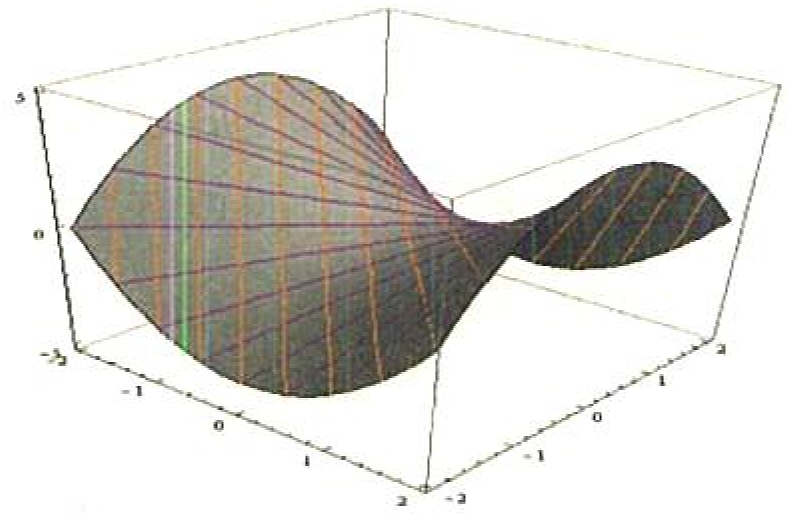


图 7-9　马鞍面也是直纹面

7.3　二次型合同对角化的几何意义

二次型对角化就是把二次型化简成标准形或规范形，或者说把二次型度量矩阵化成对角矩阵，这就是二次型的对角化问题。

对角化二次型必须要是矩阵的合同变换，或者说度量矩阵和变换后的度量矩阵（对角阵）必须是合同的。因为只有这样才能使二次型的函数值保持不变，才能使向量长度保持不变。其他方法比如矩阵的相似对角化法能使矩阵对角化，但这与二次型度量矩阵的对角化不是一回事，相似对角化法不能保证二次型的值不变。

从几何意义上讲，相似对角化法是使矩阵本身所表示的某种线性变换不变进而对角化，而二次型对角化法是保证矩阵背后所代表二次型的值不变。它们所要求的不变量是不同的。

什么样的合同变换可以保证二次型的值不变呢？其实这是一个简单的变量替换过程：

对于一个二次型：

$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \tag{7-4}$$

有一个向量替换关系 $\boldsymbol{x}=\boldsymbol{C}\boldsymbol{y}$，把它带入式（7-4），则函数值不变，得到

$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=(\boldsymbol{C}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}\boldsymbol{y}=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}\boldsymbol{y}=\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C})\boldsymbol{y} \tag{7-5}$$

式（7-5）最后的表达式中的度量矩阵是 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$，这正是一个合同变换。

确实是矩阵的“合同”变换，对矩阵 $\boldsymbol{A}$ 的左、右同时变换（左乘 $\boldsymbol{C}^{\mathrm{T}}$ 和右乘 $\boldsymbol{C}$），行和列的变换“合同”进行。

简单的变向量替换实际上就是基坐标系的变换：

基变了，原来向量的坐标系当然变了，即 $\boldsymbol{x}\to\boldsymbol{y}$，这时如果仍然保持二次型值不变的话只能把原度量矩阵变成一个新的度量矩阵 $\boldsymbol{A}\to\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$。

和广义内积的度量矩阵是一样的，选不同的基，二次型度量矩阵就不同，二次型的表达式

也就不同，但二次型的值在不同基下是不变的。因此，选何种基的问题就变成了对度量矩阵进行何种变换的问题，二次型的化简就成了度量矩阵的对角化问题。

如果新的度量矩阵是一个对角矩阵 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{AC}=\boldsymbol{\Lambda}$，我们就说把二次型合同对角化了。

合同变化的方法有多种，比如正交变换、配方法、行列初等变换法等。

要选择什么样的合同变换对矩阵 $\boldsymbol{A}$ 进行对角化呢？这与你应用目的有关系。

常见的二次型的应用主要有两类，一类是与二次曲面相关的，一类是与多元函数的极值有关的。比如，在低维空间里如果要看这个几何图形是个什么二次曲线或二次曲面的话，就只能对矩阵 $\boldsymbol{A}$ 进行正交变换，即通过旋转或镜像变换来保持原几何图形不变形；将这个二次型通过正交变换化成标准形，然后通过这个标准形判断二次型代表何种曲面。主轴定理告诉我们，标准形的参数就是二次型的矩阵的特征值。所以，根据特征值的符号或者根据惯性定理就能确定或判断曲面的类型。

如果要判断这个函数是不是在任意变量下都是正值或是负值的情况或者求这个函数的极值，就可以使用其他的可逆变换对矩阵 $\boldsymbol{A}$ 进行各种合同变换，比如初等变换、配方法等。这些变换在背景坐标系下看起来可能会改变图形的形状，比如把一个球变成一个椭球，但变换后的图形该凸的还是凸、该凹的还是凹，（复）向量的长度平方该正值仍正值、该负值还负值。

7.3.1 二次型对角化之正交变换

我们把二次型变换成标准形时，要求的是合同变换，而一般来说没有现成的方法可以直接求合同变换。但我们发现二次型对应的是一个实对称矩阵，而实对称矩阵是一定可以通过相似变换为对角阵的，更重要的是在实对称情况下，这种相似变换还可以是正交的，这时候相似变换就成了合同变换了。所以我们是用正交的相似变换代替了合同变换来求二次型的标准形的。因此，正交变换既是相似变换同时又是合同变换。

对于一个正交矩阵 $\boldsymbol{Q}$，因为 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{E}$，所以有 $\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{Q}^{-1}$。如果用正交矩阵进行合同变换，那么有向量替换关系 $\boldsymbol{x}=\boldsymbol{Qy}$ 使

$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{AQ})\boldsymbol{y}=\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{Q}^{-1}\boldsymbol{AQ})\boldsymbol{y}$$

因此，正交变换既是相似变换同时又是合同变换。

由于正交变换保持变换前后的向量内积不变，从而保持向量的长度与夹角不变，所以正交变换属于刚体变换，是代表空间的一个旋转/镜像变换。利用矩阵乘积分解的方法，这种变换的转轴、转角可以用矩阵的特征参数量化地表示出来。

和其他合同变换（又称为可逆线性替换）不保持图形的“形状”，如可以把椭圆变成圆的特点相比，正交变换则保持原图形的“形状”。

例如，椭球面的方程为

$$\frac{x^2}{1}+\frac{y^2}{4}+\frac{z^2}{9}=1$$

将其通过可逆线性替换$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$化为$x'^2+y'^2+z'^2=1$，即椭球面变成了球面，不保持图形的“形状”。

椭球面方程也通过不可逆线性变换$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 1 & & \\ & 2 & \\ & & 0 \end{pmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$化为$x'^2+y'^2=1$，即椭球面变成了圆柱面，也不保持图形的“形状”。

而正交变换保持向量长度和角度不变，因此几何图形不变。所以在讨论二次方程决定的图形时，必须用正交变换；如果只考虑它所属的类型，则可以用可逆的合同变换(当然也包括正交变换)。

用正交变换$\boldsymbol{x}=\boldsymbol{Q}\boldsymbol{y}$把三元二次型$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$化为标准形$\lambda_1y_1^2+\lambda_2y_2^2+\lambda_3y_3^2$，在几何上就是把坐标轴变换到与二次曲面$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=1$的主轴（即对称轴）重合的位置。其中$\boldsymbol{Q}$的列向量恰好给出了二次曲面的主轴的方向。因此，关于**用正交变换化二次型为标准形的定理也称为主轴定理**。这同时也是主轴定理的几何意义。

下面插一下主轴定理的定义：

对于任意一个n元二次型：

$$f(x_1,x_2,\cdots,x_n)=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

存在正交变换$\boldsymbol{x}=\boldsymbol{Q}\boldsymbol{y}$（$\boldsymbol{Q}$为$n$元正交矩阵），使得

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q})\boldsymbol{y}=\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2$$

其中：$\lambda_1,\lambda_2,\cdots,\lambda_n$是实对称矩阵$\boldsymbol{A}$的$n$个特征值；$\boldsymbol{Q}$的$n$个列向量是对应于特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$的标准正交特征向量。

因此，从几何图形上寻找二次型主轴的问题，在线性代数中就等价于让矩阵$\boldsymbol{A}$经过正交变换（同时也是相似变换）使矩阵$\boldsymbol{A}$对角化。要对角化就要找它的主轴的大小和方向，也就是找它的特征值和特征向量。

找到了特征向量就可以用它们组成一个正交矩阵，用这个正交矩阵就可以把二次型转到对角化，实际上就是把坐标轴转到二次型曲面的主轴方向，由此可以简化二次型。在许多工程问题中，可以使它的某些物理意义变得清晰起来，因而得到了广泛应用。

例 7.1 用正交变换化简下列二次曲面方程为标准形：

$$x_1^2+x_2^2+x_3^2-2x_1x_2=1$$

解 将左边的二次型表示为矩阵/向量形式：

$$x_1^2+x_2^2+x_3^2-2x_1x_2=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^T\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{7-6}$$

容易求出矩阵$\boldsymbol{A}$的特征多项式：

$$|\lambda\boldsymbol{E}-\boldsymbol{A}|=\begin{vmatrix} \lambda-1 & 0 & -1 \\ 0 & \lambda-1 & 0 \\ -1 & 0 & \lambda-1 \end{vmatrix}=\lambda(\lambda-1)(\lambda-2)$$

其实到了这里得到了特征值，我们就已经知道了二次型的标准形，如果要知道正交矩阵的

话，继续：

由每个特征值 $\lambda=1$，2，0，计算出对应的特征向量，并单位化：

$$e_1=\begin{pmatrix}0\\1\\0\end{pmatrix},e_2=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\0\\1\end{pmatrix},e_3=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}$$

因为它们是对应于不同的特征值的特征向量，因此彼此正交，是 $\boldsymbol{R}^3$ 的规范正交基，不用施密特正交化了。令 $\boldsymbol{x}=\boldsymbol{Q}\boldsymbol{y}$，其中

$$\boldsymbol{Q}=[e_1,e_2,e_3]=\begin{bmatrix}0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 1 & 0 & 0\\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

将其代入式（7-6），并化简为 $y_1^2+2y_2^2=1$。

图 7-10（a）是旧坐标系下的二次型图形，建立以特征向量为基的新坐标系。新坐标系 $\{o,y_1,y_2,y_3\}$ 的基为 $\{e_1,e_2,e_3\}$，原曲面在新坐标系下（见图 7-10（b））容易看出来是一个椭圆柱面。它的准线是 y_1y_2 平面上的椭圆，对称轴是 y_3 轴。我们看到，二次型矩阵 $\boldsymbol{A}$ 的三个特征向量 e_1、e_2、e_3 的方向就是椭圆的长轴、短轴和对称轴的方向，正是由于我们取了 $\boldsymbol{A}$ 的正交规范特征向量为新基，原方程才得以化简。由于过渡矩阵 $\boldsymbol{Q}=[e_1,e_2,e_3]$ 的行列式为 1，故新坐标系仍然是右手直角坐标系。

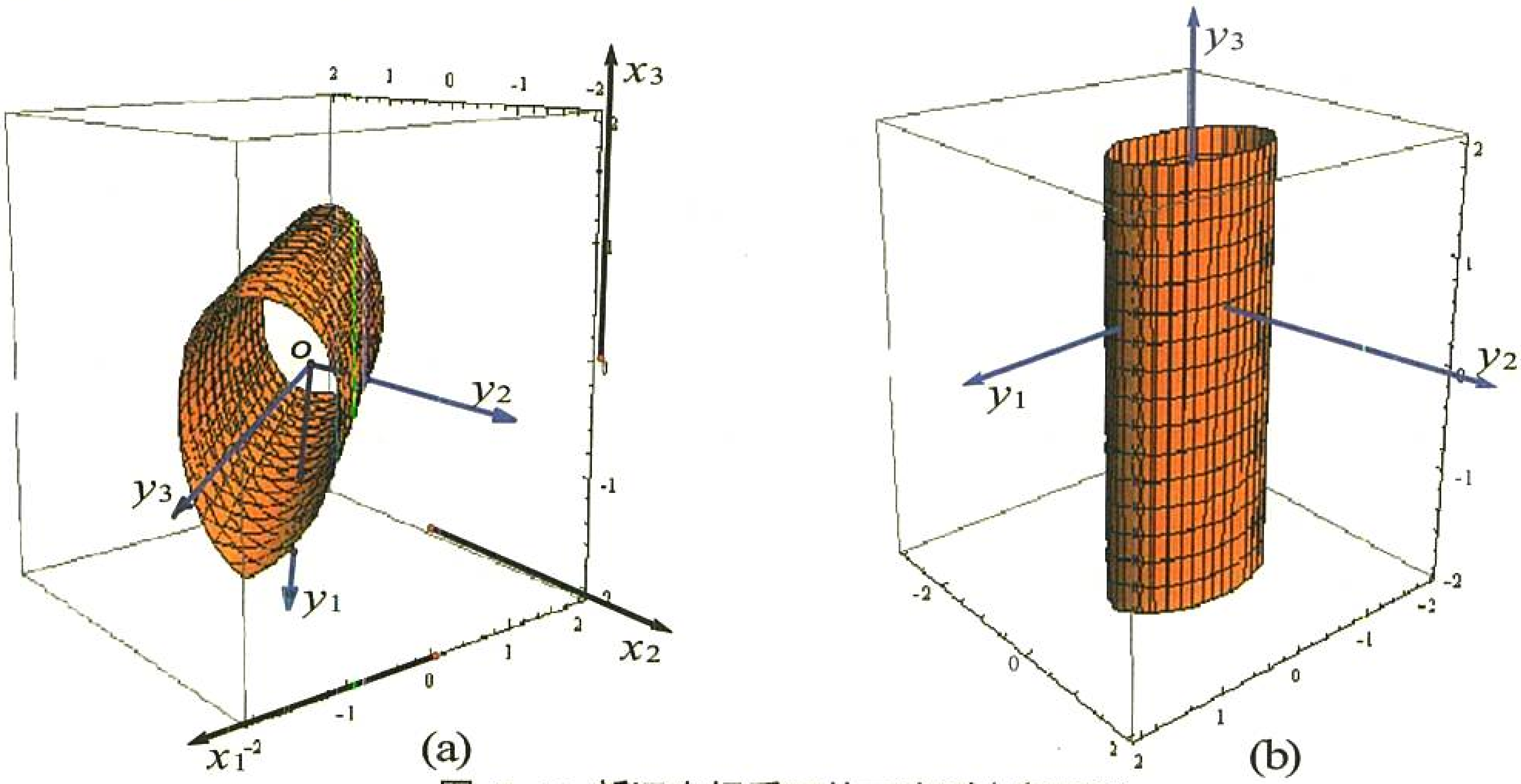


图 7-10 新旧坐标系下的二次型方程图形

合同变换和相似变换

用合同变换把矩阵对角化就等同于把函数化成了纯平方和形式。这样讨论起来函数的值的正负性质、极值等结果就非常明显，几何图形也容易推测出来。

合同变换与相似变换各有所长，前者仅适用于对称矩阵，保持了矩阵的对称性、正定性、秩等性质不变，后者适用于一般方阵，保持了矩阵的秩与特征值不变。

合同变换并不唯一，但对角线上正负元素的个数是唯一的。它保持了矩阵的秩、对称性、正定性、半正定性等。合同变换的最大特点是保持对称性不变——这是由向量长度平方值不变所决定的。

7.3.2 其他二次型对角化的方法

其他的二次型的合同对角化方法还有配方法、初等变换法等。

配方法是可逆的线性变量替换，属于合同变换。该方法咱中学都学过，用其对付简单的二次型还比较适用，不妨简单地总结一下其步骤：

(1) 如果二次型 $f(x_1,x_2,\cdots,x_n)$ 中至少有一项平方系数不为零，比如 $a_{11}x_1^2\neq 0$，那么就把所有含有 x_1 的交叉项进行配方。

(2) 配方后，进行第一次变量替换 $\begin{cases} y_1=a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n \\ y_2=x_2 \\ \vdots \\ y_n=x_n \end{cases}$，简写为 $\boldsymbol{y}=\boldsymbol{Dx}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & 1 & & 0 \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}\boldsymbol{x}$。

如果写成带入的形式就是 $\boldsymbol{x}=\boldsymbol{D}^{-1}\boldsymbol{y}=\boldsymbol{Cy}$，$\boldsymbol{D}$ 显然是可逆的。

(3) 原二次型变为 $f(x_1,x_2,\cdots,x_n)=\dfrac{1}{a_{11}}y_1^2+f'(y_2,y_3,\cdots,y_n)$ 的形式。如果二次型 $f'(y_2,y_3,\cdots,y_n)$ 中还有不为零的平方项系数，重复以上步骤，直到平方项系数为零。

(4) 对于平方项系数为零的只含有交叉乘积项的二次型，比如 $2hx_1x_2$，就使用平方差公式：

$$\begin{cases} x_1=y_1+y_2 \\ x_2=y_1-y_2 \\ x_3=y_3 \\ \vdots \\ y_n=x_n \end{cases}$$

即

$$\boldsymbol{x}=\boldsymbol{C}_i\boldsymbol{y}=\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ & & 1 & & 0 \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}\boldsymbol{y}$$

代入化成平方项，转换成前述的情况继续配方，直至化成标准形。

对于一个二次型，用上述的配方法，总可以使所有的交叉混合项消失，变成标准形。以上两种情况变量替换所用的变换矩阵都是可逆矩阵，因此，总存在一系列可逆矩阵 $\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_i,\cdots,\boldsymbol{C}_s$ 使得

$$f(x_1,x_2,\cdots,x_n)=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{C}_s^{\mathrm{T}}\cdots\boldsymbol{C}_i^{\mathrm{T}}\cdots\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{C}_1^{\mathrm{T}})\boldsymbol{A}(\boldsymbol{C}_1\boldsymbol{C}_2\cdots\boldsymbol{C}_i\cdots\boldsymbol{C}_s)\boldsymbol{y}=d_1y_1^2+d_2y_2^2+\cdots+d_ny_n^2$$

所以，配方法也是多次合同变换的结果。

初等变换法也是合同变换，有一个定理给出了二次型度量矩阵的初等变换法：

任给一个实对称矩阵 $\boldsymbol{A}$，肯定存在一连串的初等矩阵 $\boldsymbol{T}_1,\boldsymbol{T}_2,\cdots,\boldsymbol{T}_n$，使 $\boldsymbol{A}$ 对角化：

$$\boldsymbol{T}_s^{\mathrm{T}}\cdots\boldsymbol{T}_2^{\mathrm{T}}\boldsymbol{T}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{T}_1\boldsymbol{T}_2\cdots\boldsymbol{T}_s=(\boldsymbol{T}_1\boldsymbol{T}_2\cdots\boldsymbol{T}_s)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{T}_1\boldsymbol{T}_2\cdots\boldsymbol{T}_s)=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}=\begin{bmatrix} d_1 & & \\ & d_2 & \\ & & \ddots & \\ & & & d_n \end{bmatrix}$$

实际的操作是

$$T_1^{\mathrm{T}}AT_1 \to T_2^{\mathrm{T}}T_1^{\mathrm{T}}AT_1T_2 \to T_s^{\mathrm{T}}\cdots T_2^{\mathrm{T}}T_1^{\mathrm{T}}AT_1T_2\cdots T_s \to \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \tag{7-7}$$

其中矩阵 A 左乘 T_i^{T} 是对其进行初等行变换，右乘 T_i 是对其进行初等列变换。出题的考官为了圆满，常常让你给出马后炮——变换矩阵 C ；变换矩阵 $C=T_1T_2\cdots T_s$ ，如何得到？因为 $C=T_1T_2\cdots T_s=ET_1T_2\cdots T_s$ ，所以只要对 n 阶单位矩阵 E 作同样的初等列变换就可以同步求出了。

式（7-7）的运算过程是每一个步骤进行一次行变换就马上进行一次列变换，这样每个步骤都是合同的。当然，在一个步骤里可以先进行多次行变换再紧接着进行同样多次列变换。

配方法、初等变换法都是合同变换，但得到的对角矩阵的对角线上的元素一般不是特征值，这与正交变换不同。当然它们具有一致的几何意义，只是手段稍异，本质相同罢了。

7.4 惯性定理的几何及物理意义

在上面的各种对角化的不同变换中得到的标准形一般是不一样的，比如正交变换时，并没有规定对角矩阵中对角元的顺序，所以对角矩阵是不唯一的。类似地，其他合同对角化方法的每个步骤也不是唯一的，得到的对角矩阵也不是唯一的，得到的标准形自然也不是唯一的。

尽管“沧海桑田”，仍有能够“永恒”之物，那就是标准形中非零系数个数、正系数个数、负系数个数都是不变的，此即西尔维斯特(Sylvester)惯性定理：

设 n 阶二次型 $f(x)=x^{\mathrm{T}}Ax$ 的秩为 r，若可逆线性变换 $x=Cy$ 及 $x=Py$ 分别将二次型 f 化成标准形：

$$f=k_1y_1^2+k_2y_2^2+\cdots+k_ry_r^2 \quad (k_i\neq 0, i=1,2,\cdots,r)$$

及

$$f=\lambda_1z_1^2+\lambda_2z_2^2+\cdots+\lambda_rz_r^2 \quad (\lambda_i\neq 0, i=1,2,\cdots,r)$$

则 $k_1,k_2,\cdots,k_r$ 与 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 中带正号的系数个数相同（同时负号系数个数也相同）。

任何一个实二次型都可以经过一个线性变换化为标准形。惯性定理是说，标准形的正系数的个数与负系数的个数与标准形的坐标基选择无关，二次型对满秩线性替换是不变量。正因为它们是不变量，所以我们叫这个定理为惯性定理，其中正系数个数称为正惯性指数(可记为 P)，负系数个数称为负惯性指数（可记为 N）。正惯性指数加上负惯性指数等于秩，即 $P+N=r$ 。

惯性定理的几何意义：

惯性定理反映到几何上，就是经过可逆的合同变换把二次曲线/面方程化成标准方程。方程的系数与所作的线性变换有关；而曲线的类型（是椭圆型 、双曲线型等）是不会因为所作的线性变换的不同而改变的。曲线的类型在几何图形上就像是图形的轮廓，这些不同的轮廓与基的选择无关。例如，一个马鞍面无论选择的基是什么，都是一个马鞍面，尽管马鞍面可以变大变小，变陡峭或平坦，亦可横着、竖着、斜着、倒着等。马鞍面的这些改变取决于基向量的不同取法。

二次型图形的对称性是显然的，从标准形的代数式 $f=k_1y_1^2+k_2y_2^2+\cdots+k_ry_r^2$ 可以看出，对于任一个坐标轴 y_i， $k_iy_i^2$ 的图形是关于坐标零点左右对称的，这些平方项叠加起来仍然关

于每个坐标轴对称。这些对称性图形可以通过观察后续的二次型的图形来体会。

上面几何意义的解释是比较常见的说法，其实质是假定存在一个绝对的世界坐标系作为背景，它是永远都是刻度不变的标准正交坐标系，所变的是二次型的几何图形。

其实更好的解释和正交变换的坐标系旋转/镜像的几何解释类似：二次型的几何图形不变，变的是坐标系：原来的标准正交系变成了刻度和坐标轴夹角都不同的仿射坐标系。想象二次型所表示的几何图形是一个物理实体，各种合同变换所变换的是坐标系。因为坐标系的变化向量之坐标和度量矩阵一起变，因而二次型的函数表达式也相应地改变了——每一个坐标系下对应着一个函数表达式；然而仍要记住这个物理实体是不变的。这个说法看起来像二次型的物理意义的解释。

惯性定理又叫主轴定理，因为二次型的图形在化成标准形体现出了最完美的对称性，每一个基向量所在的直线都成为二次齐次函数图形的对称轴，我们称这些对称轴为二次型的主轴。

惯性定理或主轴定理的命名由来更像是直接来自于力学中的惯性张量矩阵二次型。惯性矩二次型是描述刚体绕定点转动的总惯量。转动物体的转动惯量和直线运动物体的质量一样都是描述惯性大小的标量，因此具有不变性。惯性矩二次型经过正交变换得到了由三个主轴转动惯量所组成的对角阵。

惯性定理在相对论中的应用

正如前面所讲，二次型惯性定理的更本质一些的几何意义是二次型图形本身可以看做不变的物理实体，二次型在选择不同的基/坐标轴时并不能改变这个物理实体，只是这个物理实体的数学描述式 $f=k_1y_1^2+k_2y_2^2+\cdots+k_ry_r^2$ 在随着不同的基而改变而已（呵呵，数学是不是有点唯心）。这个数学描述式再改变也是描述的同一个物理实体啊，因此，在线性映射下，一个球体再怎么投影也就变成个小球、大球或椭圆球面，不会变成马鞍面那样的波澜起伏的丘陵加山壑。

二次型的这个不变的“惯性”被闵可夫斯基成功地运用于描述狭义相对论的四维时空线性空间（被称为闵可夫斯基空间）里的度规不变量，这个二次型在不同的惯性系（坐标系间保持静止或匀速直线运动）变换下是不变量。

闵可夫斯基空间的二次型不变的来源是因为假设光速不变。假设三维空间坐标零点发出一个光脉冲，光线向四面八方随着时间 t 的延长而扩散开来，最外面的光波阵面是一个迅速扩张的球面，这个球面的半径等于光速和时间的积 ct，因此这个球面方程为

$$\sqrt{x^2+y^2+z^2}=ct$$

两边平方得到

$$x^2+y^2+z^2=c^2t^2$$

那么我们把这个关系式定义为一个二次型

$$f(x,y,z,t)=x^2+y^2+z^2-c^2t^2=0$$

在洛仑兹变换（坐标变换）下，坐标系从 $\{o,x,y,z,t\}$ 变换为 $\{o,x',y',z',t'\}$，这个二次型 f 的值仍然等于 0，也就是总有

$$f\left(x,y,z,t\right)=x^2+y^2+z^2-c^2t^2=x'^2+y'^2+z'^2-c^2t'^2=f\left(x',y',z',t'\right)$$

7.5 二次型正定性的几何意义

二次型的定义式 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 看起来与内积的推广定义式 $<\boldsymbol{x},\boldsymbol{y}>=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{y}$ 是类似的；同时，二次型矩阵 $\boldsymbol{A}$ 和内积的度量矩阵 $\boldsymbol{S}$ 都是实对称矩阵，看起来渊源颇深啊。

泛泛地看来，二次型只是内积的特殊形式——是一个向量与自身的内积，或者说是向量长度的平方。实际上，因为在实线性空间里内积度量矩阵是一个正定对称方阵，任意一个正定对称方阵 $\boldsymbol{S}$ 可一一对应一个内积 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{y}$；如果一个二次型方阵 $\boldsymbol{A}$ 也是一个正定实阵，则根据二次型和实对称方阵的一一对应性，我们当然可以把一个正定二次型看做一个内积了，甚至更自然的是向量长度的平方了。

正定具体是指什么意思，通过对二次型进行分类，事情可能更清晰一些。

7.5.1 二次型正定性的几何意义

对二次型的分类主要有五类：正定，负定，半正定，半负定，不定。它们的几何意义都是什么呢？先看定义：

设有二次型 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$。

- 若对任何 $\boldsymbol{x}\neq\boldsymbol{0}$，都有 $f(\boldsymbol{x})>0$，则称 $f(\boldsymbol{x})$ 为正定二次型，并称对称矩阵 $\boldsymbol{A}$ 是正定矩阵，记为 $\boldsymbol{A}>\boldsymbol{0}$。
- 若对任何 $\boldsymbol{x}\neq\boldsymbol{0}$，都有 $f(\boldsymbol{x})<0$，则称 $f(\boldsymbol{x})$ 为负定二次型，并称对称矩阵 $\boldsymbol{A}$ 是负定矩阵，记为 $\boldsymbol{A}<\boldsymbol{0}$。
- 若对任何 $\boldsymbol{x}\neq\boldsymbol{0}$，都有 $f(\boldsymbol{x})\geqslant 0$，则称 $f(\boldsymbol{x})$ 为半正定的；若 $-f(\boldsymbol{x})$ 为半正定的，则称 $f(\boldsymbol{x})$ 为半负定（负半定）的；若 $f(\boldsymbol{x})$ 既不是半正定又不是半负定的，则称 $f(\boldsymbol{x})$ 为不定的。

对上述四平八稳的定义麻木了就看不到细节了，换个讲法：

在欧氏空间 $\mathbf{R}^n$，实二次型 $f(x_1,x_2,\cdots,x_n)$ 作为定义在 $\mathbf{R}^n$ 的一个二次实函数 $z=f(x_1,x_2,\cdots,x_n)$，则当 $f(x_1,x_2,\cdots,x_n)$ 为正定时就是只要取 $x_1,x_2,\cdots,x_n$ 为一组不全为 0 的实数，均有 $z>0$（只有当 $x_1=x_2=\cdots=x_n=0$ 时，$z=0$）；也就是说，这个函数的值域是 $z\geqslant 0$，但只有在 $\mathbf{R}^n$ 的原点取得 $z=0$，而在其他点上均有 $z>0$。这个地方好多人都忽略了。

半正定呢，就是在此二次型在整个定义域 $\mathbf{R}^n$ 上也有值域 $z\geqslant 0$，但除了原点使 $z=0$ 外，尚有其他一些点也使 $z=0$，此外便是 $z>0$（负定和半负定相应地有类似的结果，只是 $z\leqslant 0$）。

当二次型为不定时，则函数值域既有正实数也有负实数和 0。

下面举几个正定、半正定和不定的例子及其三维图形供参考。

例 7.2 $z=f(x,y)=\dfrac{1}{4}x^2+y^2$ 是一个正定二次型，它的图像如图 7-11（a）所示，是一个开口向上的椭圆抛物面，整个曲面除一点（顶点）在坐标平面 xoy 的原点上，其余均在 xoy 平面的上方。

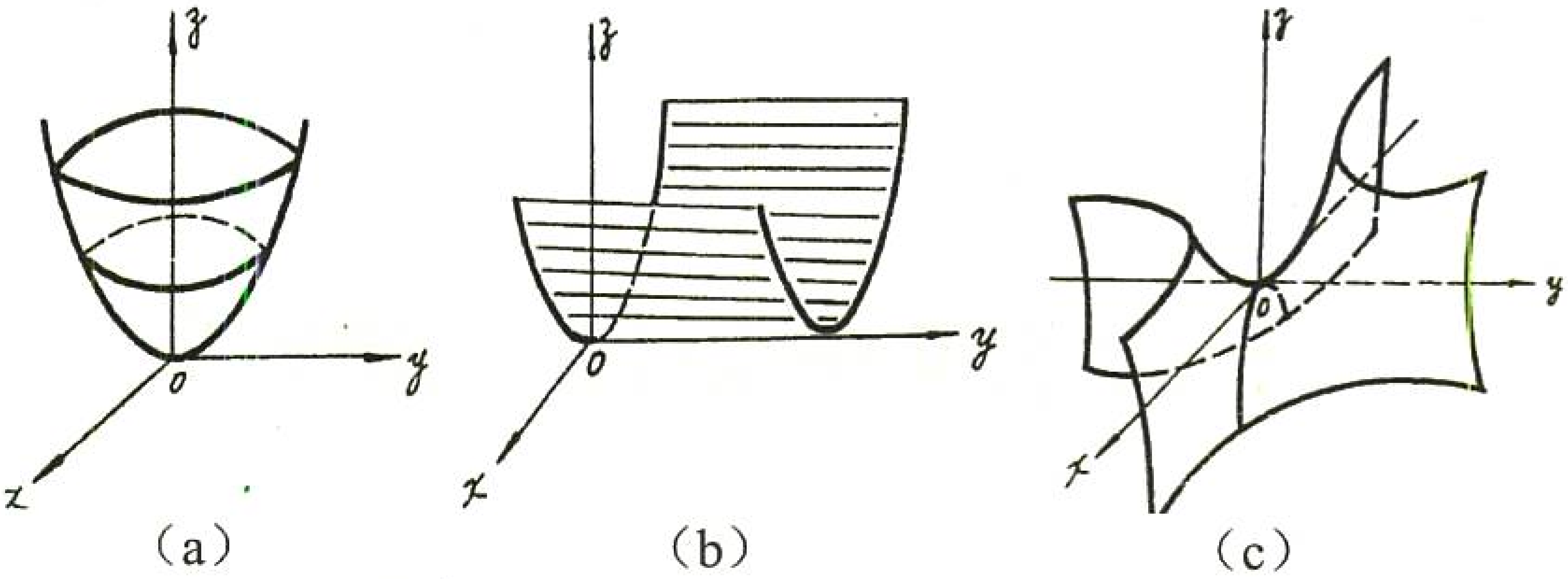


图 7-11 正定、半正定和不定二次型的曲面图形

正定二次型有着明显的几何意义，当 $f(x,y)$ 为二维的正定二次型时，它可由可逆线性替换（坐标变换）化为系数全为正的标准形，于是 $f(x,y)=c\ (c>0)$ 的图形是以原点为中心的椭圆，当把 c 看做任意正常数时，则是一族椭圆线，这族椭圆随着 c（沿着抛物线）减小且趋于零而收缩到原点（见图 7-11（a））；当 f 为三维正定二次型时，$f(x,y,z)=c\ (c>0)$ 的图形则是一族椭球面，这族椭球面随着 c（沿着抛物面）减小且趋于零而收缩到原点。

例 7.3　$z=f(x,y)=2x^2$ 是半正定二次型，它的图形如图 7-11（b）所示，是一个开口向上，以 xoy 面上的抛物线 $z=2x^2$ 为准线，以平行于 y 轴的直线为母线的抛物柱面。整个曲面有一条母线在 xoy 面上（即 y 轴，轴上二次型等于零），其余均在 xoy 平面上方。

例 7.4　$z=f(x,y)=xy$ 是一个不定二次型，其图形如图 7-11（c）所示，是一个过原点的马鞍面。显然，马鞍面上正值、负值和零都有。

7.5.2 二次型正定性判别法的直观理解

知道了二次型的函数图形当然比较容易判定二次型的正定性，一旦碰到超多元的二次型就麻烦了，还是需要代数的计算手段。我们知道了一个实二次型对应一个实对称矩阵，是否从矩阵本身就可以知道这个二次型的正定性？

应该如此。因为合同变换不改变二次型函数的值，原来大于零的还是大于零，所以合同对角化矩阵就可以了，如果对角线上的元素是大于零的数就是正定的了（这说明标准形的系数都是正值）。

对角化判定方法之一是特征值法，因为特征值是正交变换（也是合同变换）对角化矩阵的对角线上的元素。特征值大于零，当然二次型也是正定的了。

还有个矩阵正定性判定的方法，比如顺序主子式法：

(1)　$\boldsymbol{A}$ 是正定矩阵的充要条件是它的所有主子式都大于 0。

(2)　$\boldsymbol{A}$ 是负定矩阵的充要条件是它的所有奇数阶主子式都小于 0，并且它的所有偶数阶主子式都大于 0。

顺序主子式法能理解就好了，我们看一个例子，见表 7-1。

表 7-1　各阶二次型矩阵顺序主子式的特点分析

顺序	二次型各阶函数式	二次型度量矩阵	行列初等变换后的合同对角阵	行列式的极性	二次型标准形函数式
1	$f(x_1)=a_{11}x_1^2$　是正定的	$\begin{bmatrix}a_{11}\end{bmatrix}$	$\begin{bmatrix}a_{11}\end{bmatrix}$	$a_{11}>0$	$f(x_1)=a_{11}x_1^2$
2	$f(x_1,x_2)=a_{11}x_1^2+a_{22}x_2^2+2a_{12}x_1x_2$ 是正定的	$\begin{bmatrix}a_{11} & a_{12}\\ a_{12} & a_{22}\end{bmatrix}$	$\begin{bmatrix}a_{11} & \\ & d_2\end{bmatrix}$	$[a_{11}]d_2>0$	$f(x_1,x_2)=a_{11}x_1^2+d_2x_2^2$
3	$f(x_1,x_2,x_3)$ $=\underline{a_{11}x_1^2+a_{22}x_2^2+2a_{12}x_1x_2}+a_{33}x_3^2$ $+2a_{13}x_1x_3+2a_{23}x_2x_3$ 是正定的	$\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{12} & a_{22} & a_{23}\\ a_{13} & a_{23} & a_{33}\end{bmatrix}$	$\begin{bmatrix}a_{11} & & \\ & d_2 & \\ & & d_3\end{bmatrix}$	$[a_{11}d_2]d_3>0$	$f(x_1,x_2,x_3)$ $=a_{11}x_1^2+d_2x_2^2+d_3x_3^2$
⋮	⋮	⋮	⋮	⋮	⋮

n	$f(x_1,x_2,0)$ $=a_{11}x_1^2+a_{22}x_2^2+2a_{12}x_1x_2+0$	$\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} a_{11} & & \\ & d_2 & \\ & & 0 \end{bmatrix}$	$[a_{11}d_2]>0$	$f(x_1,x_2,0)$ $=a_{11}x_1^2+d_2x_2^2+0$

咱来看看：

顺序 1 中，一阶二次型是正定的等价于$a_{11}>0$。

顺序 2 中，二阶二次型矩阵$\boldsymbol{A}$经过行和列同时的初等变换，得到了包含a_{11}和d_2的对角阵$\boldsymbol{\Lambda}$。注意，$\boldsymbol{A}$和$\boldsymbol{\Lambda}$的行列式是同符号的，因为$|\boldsymbol{\Lambda}|=|\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}|=|\boldsymbol{C}|^2|\boldsymbol{A}|$。

显然，a_{11}和d_2要同时大于 0 的话，二次型就是正定；反之亦然。

a_{11}和d_2要同时大于 0 等价于$a_{11}>0$和$a_{11}d_2>0$，换句话说，$\boldsymbol{A}$的顺序主子式全大于 0。

随着变量逐步地增加，显然各阶主子式必须大于 0，才是各阶二次型正定的充要条件。

我相信，通过仔细地比对上述的二次型函数式、顺序主子式，各位能顺利地理解正定的主子式判定定理。

另外，表 7-1 给出了顺序 n 是为了和顺序 3 进行比对，应该看出，各阶顺序主子式也可认为是变量x_1、x_2、x_3中批量取零得到矩阵行列式。如对于三阶二次型，令$x_3=0$得到二阶主子式，令$x_2=x_3=0$得到一阶主子式。

俺衷心希望你能顺利地理解下面的判断。

例 7.5 判别二次型$f=-3x^2-5y^2-7z^2+4xy+4xz$的正定性。

解 f的矩阵为

$$\boldsymbol{A}=\begin{bmatrix} -3 & 2 & 2 \\ 2 & -5 & 0 \\ 2 & 0 & -7 \end{bmatrix}$$

因为

$$a_{11}=-3<0,\quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=\begin{vmatrix} -3 & 2 \\ 2 & -5 \end{vmatrix}=11>0,\quad |\boldsymbol{A}|=-57<0$$

故由上述顺序主子式法定理知f为负定的。

7.6 二次型的分类与二次曲面的分类

对二次型分类方式，除了前面讲到的正定性的分法，几何上也有多种说法，比如投影分类、仿射分类或者度量分类等，无论怎么分类，最好都要把n元二次型化为标准形再说：

$$f(x_1,x_2,\cdots,x_n)=f(y_1,y_2,\cdots,y_r)=k_1y_1^2+k_2y_2^2+\cdots+k_ry_r^2(k_i\neq 0)$$

化成如上的标准形，我们再看这个二次型正定性就一目了然：

- 正定$\Leftrightarrow P=r=n$，就是所有的系数k_i大于 0，系数个数是n。
- 负定$\Leftrightarrow P=0,r=n$，就是所有的系数k_i小于 0，系数个数是n。
- 半正定$\Leftrightarrow P=r<n$，就是所有的系数k_i大于 0，但系数个数小于n。
- 半负定$\Leftrightarrow P=0,r<n$，就是所有的系数k_i小于 0，但系数个数小于n。
- 不定$\Leftrightarrow 0<P<r$（$\leqslant n$），就是系数k_i既有大于 0 的也有小于 0 的。

这里 P 是二次型的正惯性指数，n 是二次型阶数，r 是二次型的秩。

好了，二次型的分类确定了，由二次型的切片或截面得到的二次曲面/线也得到了分类。所谓二次型的切片，就是使二次型等于一个实数 d，即令 $f(x_1,x_2,\ldots,x_n)=d$。这个实数 d 可以是大于 0、小于 0 或等于 0。

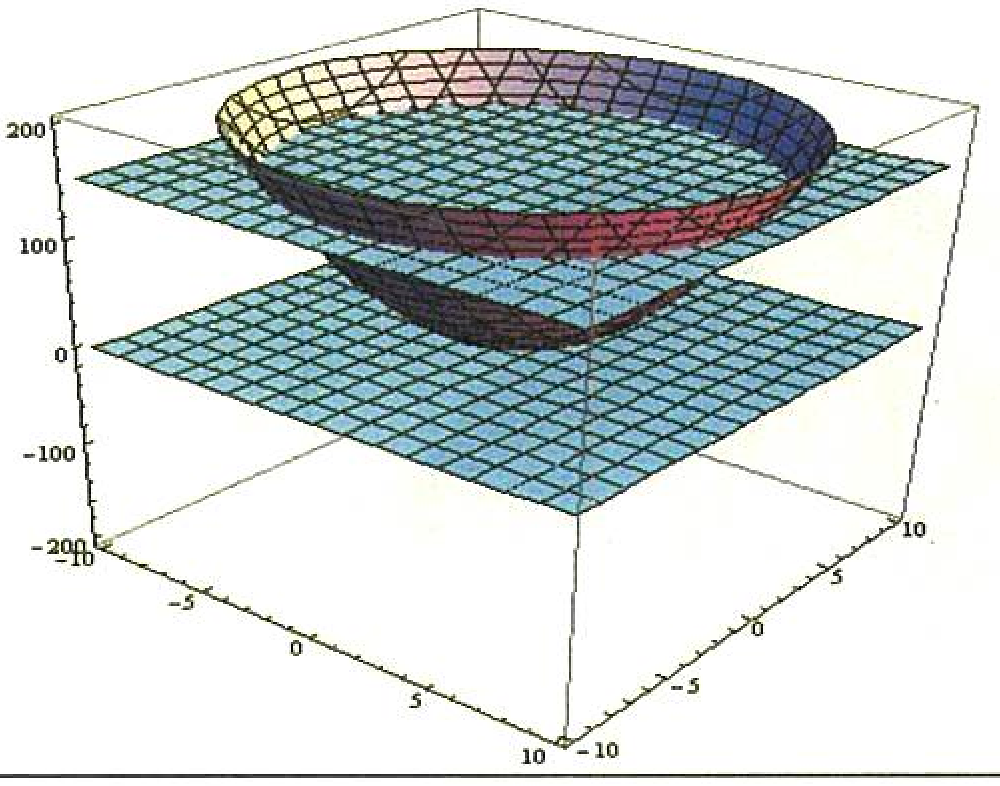
当然，为了让切片能在实数空间里切到实物，正定或半正定的二次曲面就让 d 大于 0 或 d 大于等于 0；负定或半负定的二次曲面就让 d 小于 0 或 d 小于等于 0；不定的二次曲面就让 d 取三个值：大于、小于和等于 0。这样一来，图形就全面了。好，我们先以简单的二元二次型给出对比，见表 7-2 (为了方便，这里假定 k_i 大于 0)。

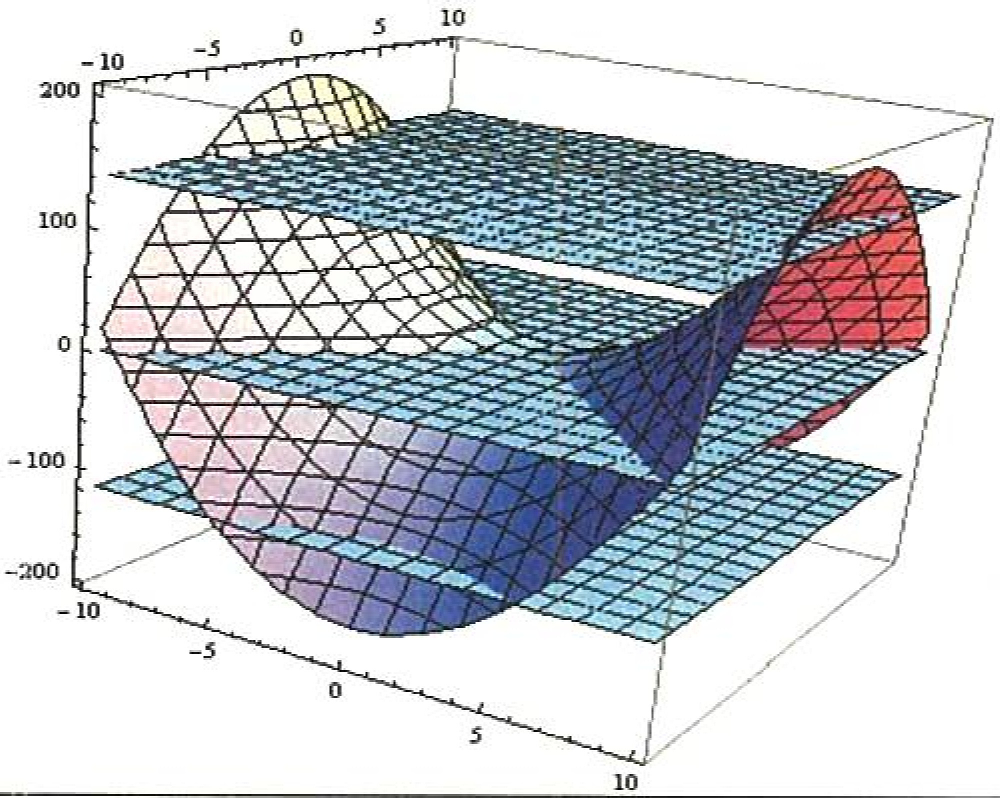
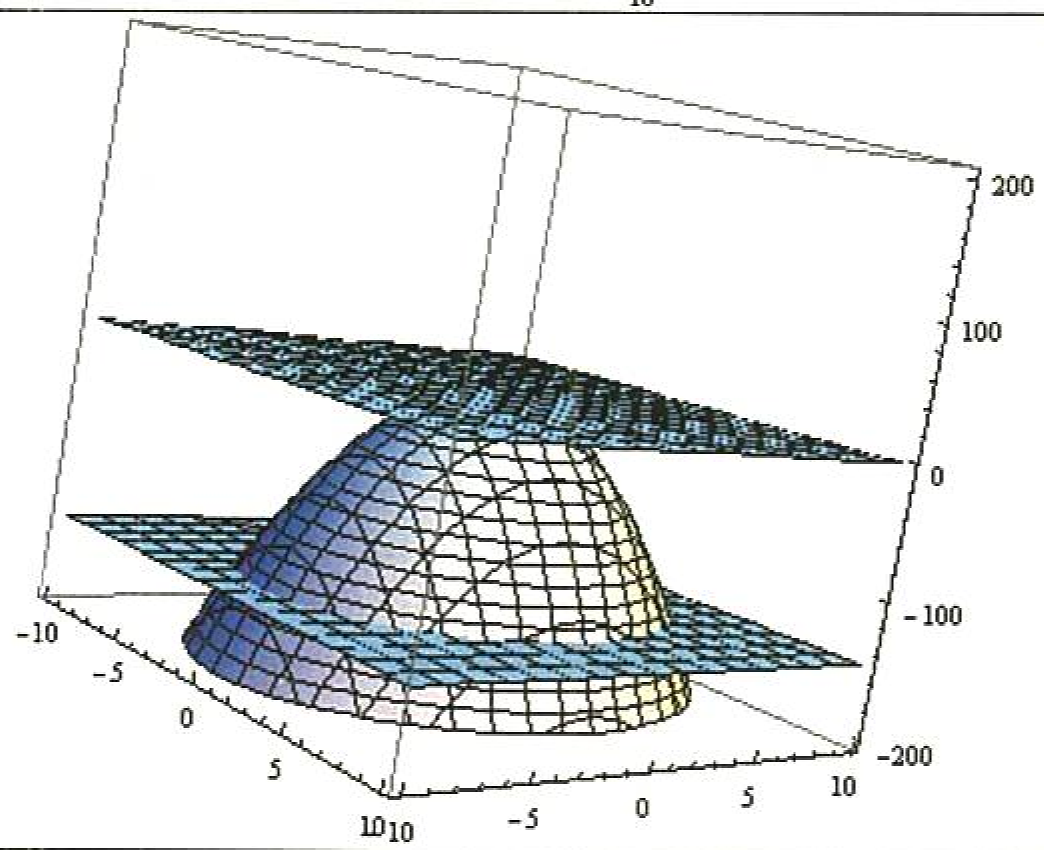
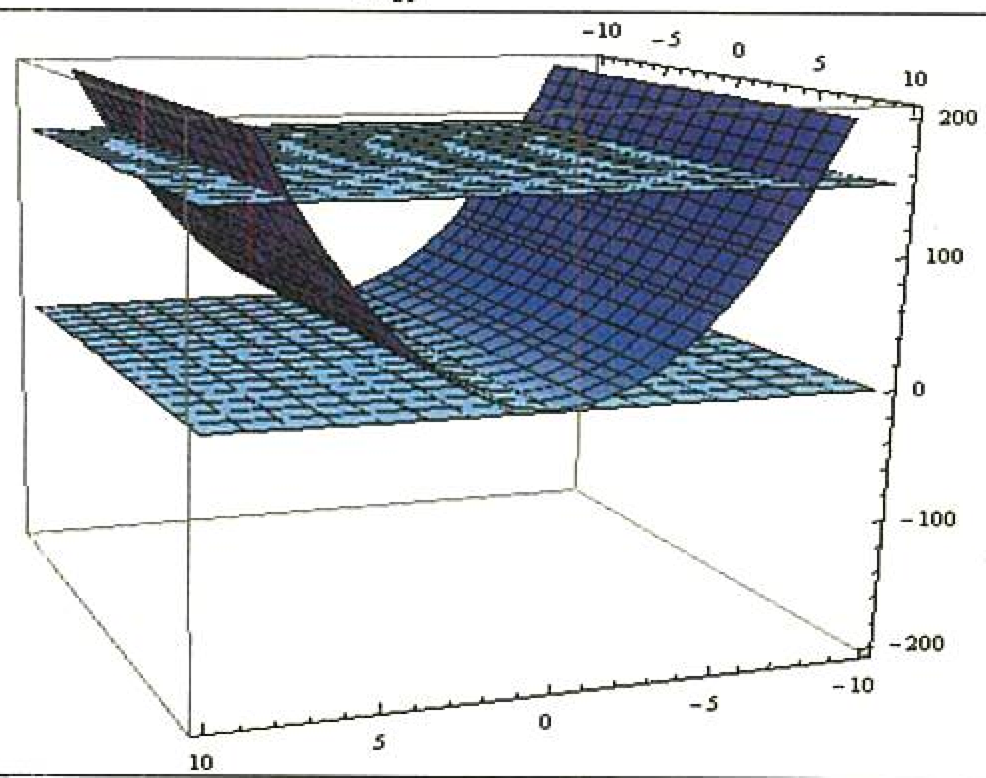
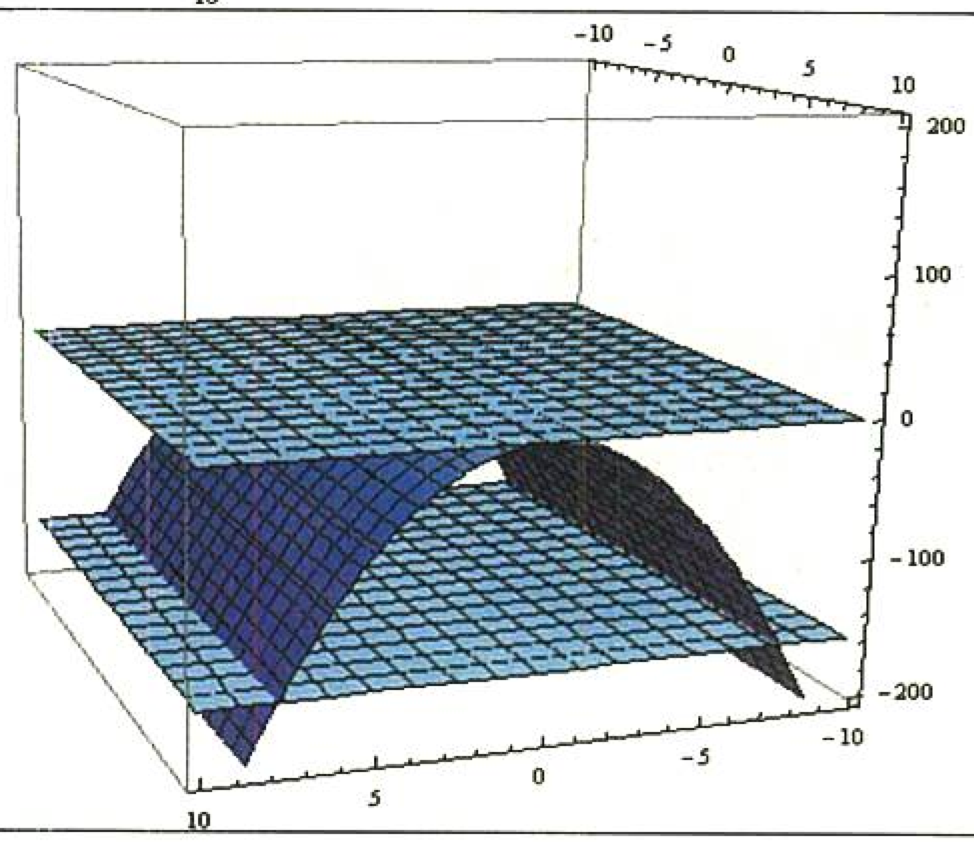
表 7-2　二元二次型的分类

序号	r	p	k_1	k_2	二次标准形	二次型分类	二次曲面方程/二次型切片	二次曲面图形分类
1	2	2	+	+	$k_1y_1^2+k_2y_2^2$	正定	$k_1y_1^2+k_2y_2^2=1$	圆/椭圆
2	2	1	+	-	$k_1y_1^2-k_2y_2^2$	不定	$k_1y_1^2-k_2y_2^2=1$	双曲线
							$k_1y_1^2-k_2y_2^2=0$	相交 2 直线
							$k_1y_1^2-k_2y_2^2=-1$	双曲线
3	2	0	-	-	$-k_1y_1^2-k_2y_2^2$	负定	$-k_1y_1^2-k_2y_2^2=-1$	圆/椭圆
4	1	1	+	0	$k_1y_1^2$	半正定	$k_1y_1^2=1$	平行 2 直线
							$k_1y_1^2=0$	重合 2 直线
5	1	0	-	0	$-k_1y_1^2$	半负定	$-k_1y_1^2=-1$	平行 2 直线
							$-k_1y_1^2=0$	重合 2 直线
6	0	0	0	0	0	—	（0，0）	坐标零点

下面绘出图形，见表 7-3。

表 7-3　二元二次曲面的图形分类

序号	二次标准形函数	二次型图形及切片	二次曲面方程/二次型切片（截线）名称
1	$k_1y_1^2+k_2y_2^2$，正定		$k_1y_1^2+k_2y_2^2=1$，平面的截痕为圆/椭圆

2	$k_1y_1^2-k_2y_2^2$， 不定		$k_1y_1^2-k_2y_2^2=1$，平面的截痕为双曲线。 $k_1y_1^2-k_2y_2^2=0$，平面的截痕为相交 2 直线。 $k_1y_1^2-k_2y_2^2=-1$，平面的截痕为双曲线
3	$-k_1y_1^2-k_2y_2^2$， 负定		$-k_1y_1^2-k_2y_2^2=-1$，平面的截痕为圆/椭圆。
4	$k_1y_1^2$， 半正定		$k_1y_1^2=1$，平面的截痕为平行两直线。 $k_1y_1^2=0$，平面的截痕为重合两直线
5	$-k_1y_1^2$， 半负定		$-k_1y_1^2=-1$，平面的截痕为平行两直线。 $-k_1y_1^2=0$，平面的截痕为重合两直线。

从表 7-3 可以看出，二元二次曲面的图形大致可以分类为椭圆、双曲线和直线。

下面给出三元二次型的分类及其二次曲面的分类，见表 7-4。

表 7-4　三元二次型及其二次曲面的分类

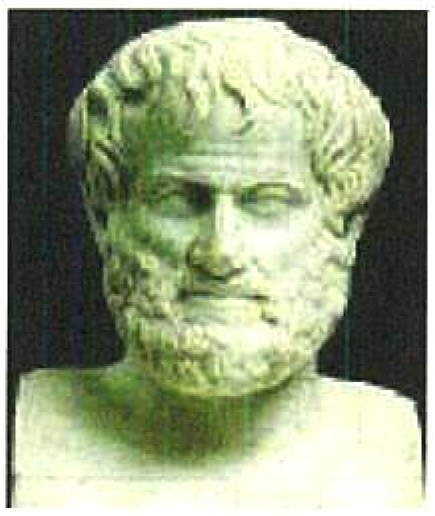
序号	r	p	k_1	k_2	k_3	二次标准形	二次型分类	二次曲面方程/二次型切片方程	二次曲面分类
1	3	3	+	+	+	$k_1y_1^2+k_2y_2^2+k_3y_3^2$	正定	$k_1y_1^2+k_2y_2^2+k_3y_3^2=1$	球面/椭球面
2	3	2	+	+	-	$k_1y_1^2+k_2y_2^2-k_3y_3^2$	不定	$k_1y_1^2+k_2y_2^2-k_3y_3^2=1$	单叶双曲面
								$k_1y_1^2+k_2y_2^2-k_3y_3^2=0$	圆锥面
								$k_1y_1^2+k_2y_2^2-k_3y_3^2=-1$	双叶双曲面
3	3	1	+	-	-	$k_1y_1^2-k_2y_2^2-k_3y_3^2$	不定	$k_1y_1^2-k_2y_2^2-k_3y_3^2=1$	双叶双曲面
								$k_1y_1^2-k_2y_2^2-k_3y_3^2=0$	圆锥面
								$k_1y_1^2-k_2y_2^2-k_3y_3^2=-1$	单叶双曲面，
4	3	0	-	-	-	$-k_1y_1^2-k_2y_2^2-k_3y_3^2$	负定	$-k_1y_1^2-k_2y_2^2-k_3y_3^2=-1$	球面/椭球面
5	2	2	+	+	0	$k_1y_1^2+k_2y_2^2$	半正定	$k_1y_1^2+k_2y_2^2=1$	椭圆柱面
								$k_1y_1^2+k_2y_2^2=0$	一条直线
6	2	1	+	-	0	$k_1y_1^2-k_2y_2^2$	不定	$k_1y_1^2-k_2y_2^2=1$	双曲柱面
								$k_1y_1^2-k_2y_2^2=0$	双相交平面
								$k_1y_1^2-k_2y_2^2=-1$	双曲柱面
7	2	0	-	-	0	$-k_1y_1^2-k_2y_2^2$	半负定	$-k_1y_1^2-k_2y_2^2=-1$	椭圆柱面
								$-k_1y_1^2-k_2y_2^2=0$	一条直线
8	1	1	+	0	0	$k_1y_1^2$	半正定	$k_1y_1^2=1$	一对平行面
								$k_1y_1^2=0$	一对重合面
9	1	0	-	0	0	$-k_1y_1^2$	半负定	$-k_1y_1^2=-1$	一对平行面
								$-k_1y_1^2=0$	一对重合面
10	0	0	0	0	0	0	----	（0，0，0）	坐标零点

从表 7-4 可以看出，三元二次曲面的图形大致可以分类为椭球面、双曲面、圆锥面、柱面、平面及直线等，我们就不再给出图形了，部分图形请参考本章开始章节中的内容。

附录 线性代数简史和名师学习指点

1. 线性代数主要内容及其发展简史

下面的线性代数的简史摘编自网络百科《线性代数》的内容，包括向量、行列式、矩阵、线性方程组及二次型。资料难得，并遍索网络得巨人头像，供上瞻仰。

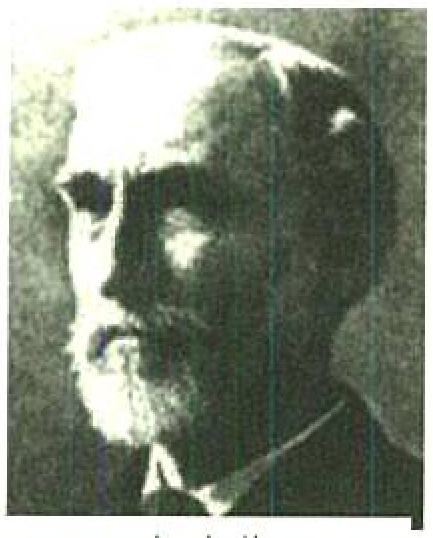

亚里士多德

向量

向量又称为矢量，最初应用于物理学。很多物理量如力、速度、位移以及电场强度、磁感应强度等都是向量。大约公元前 350 年前，古希腊著名学者亚里士多德就知道了力可以表示成向量，两个力的组合作用可用著名的平行四边形法则来得到。

“向量”一词来自力学、解析几何中的有向线段。最先使用有向线段表示向量的是英国科学家牛顿。向量进入数学并得到发展的阶段是 18 世纪末期，挪威测量学家威塞尔首次利用坐标平面上的点来表示复数 $a+bi$，并利用具有几何意义的复数运算来定义向量的运算。把坐标平面上的点用向量表示出来，并把向量的几何表示用于研究几何问题与三角问题。人们逐步接受了复数，也学会了利用复数来表示和研究平面中的向量，向量就这样平静地进入了数学。 但复数的利用是受限制的，因为它仅能表示平面，若有不在同一平面上的力作用于同一物体，则需要寻找所谓三维“复数”以及相应的运算体系。19 世纪中期，英国数学家哈密尔顿发明了四元数（包括数量部分和向量部分），以代表空间的向量。他的工作为向量代数和向量分析的建立奠定了基础。随后，电磁理论的发现者、英国的数学物理学家麦克斯韦把四元数的数量部分和向量部分分开处理，从而创造了大量的向量分析。


哈密尔顿


吉布斯

三维向量分析的开创，以及与四元数的正式分裂，是美国的吉布斯（Gibbs）和海维塞德于 19 世纪 80 年代各自独立完成的。他们提出，一个向量不过是四元数的向量部分，但不独立于任何四元数。他们引进了两种类型的乘法，即数量积和向量积。并把向量代数推广到变向量的向量微积分。从此，向量的方法被引进到分析和解析几何中来，并逐步完善，成为了一套优良的数学工具。

一般日常生活中使用的向量是一种带几何性质的量，除零向量外，总可以画出箭头表示方向。但是在高等数学中还有更广泛的向量。例如，把所有实系数多项式的全体看成一个多项式空间，这里的多项式都可看成一个向量。在这种情况下，要找出起点和终点甚至画出箭头表示方向是办不到的。这种空间中的向量比几何中的向量要广泛得多，可以是任意数学对象或物理对象。这样，就可以指导线性代数方法应用到广阔的自然科学领域中去了。因此，向量空间的概念已成为数学中最基本的概念和线性代数的中心内容，它的理论和方法在自然科学的各领域中得到了广泛的应用。而向量及其线性运算也为“向量空间”这一抽象的概念提供了一个具体的模型。

从数学发展史来看，历史上很长一段时间，空间的向量结构并未被数学家们所认识，直到 19 世纪末 20 世纪初，人们才把空间的性质与向量运算联系起来，使向量成为具有一套优良运算通性的数学体系。

行列式

行列式出现于线性方程组的求解，它最早是一种速记的表达式，现在已经是数学中一种非常有用的工具。行列式是由日本数学家关孝和以及德国的莱布尼茨发明的。关孝和于 1683 年在其著作《解伏题之法》中第一次提出了行列式的概念与展开算法。同时代的莱布尼茨是欧洲第一个提出行列式概念的人。他


关孝和

在1693年4月写给洛必达的一封信中使用了行列式，并给出方程组的系数行列式为零的条件。


克莱姆

1750年，瑞士数学家克莱姆（G. Cramer，1704－1752）在其著作《线性代数分析导言》中对行列式的定义和展开法则给出了较完整、明确的阐述，并给出了现在我们所称的解线性方程组的克莱姆法则。稍后，法国数学家贝祖（E. Bezout，1730－1783)将确定行列式每一项符号的方法进行了系统化，利用系数行列式概念指出了如何判断一个有 n 个未知量的 n 个齐次线性方程组有非零解的方法，就是系数行列式等于零是这个方程组有非零解的条件。

总之，在很长一段时间内，行列式只是作为解线性方程组的一种工具使用，并没有人意识到它可以独立于线性方程组之外，单独形成一门理论加以研究。

在行列式的发展史上，第一个对行列式理论做出连贯、逻辑的阐述，即把行列式理论与线性方程组求解相分离的人，是法国数学家范德蒙（A.T. Vandermonde，1735－1796)。范德蒙自幼在父亲的指导下学习音乐，但对数学有浓厚的兴趣，后来终于成为法兰西科学院院士。特别地，他给出了用二阶子式和它们的余子式来展开行列式的法则。就对行列式本身这一点来说，他是这门理论的奠基人。1772年，拉普拉斯在一篇论文中证明了范德蒙提出的一些规则，推广了他的展开行列式的方法。


柯西

继范德蒙之后，在行列式的理论方面，又一位做出突出贡献的就是法国大数学家柯西(Cauchy)。1815年，柯西在一篇论文中给出了行列式的第一个系统的、几乎是近代的处理。其中主要结果之一是行列式的乘法定理。另外，他第一个把行列式的元素排成方阵，采用双足标记法；引进了行列式特征方程的术语；给出了相似行列式概念；改进了拉普拉斯的行列式展开定理并给出了一个证明等。

继柯西之后，在行列式理论方面最多产的人就是德国数学家雅可比(J.Jacobi，1804－1851)，他引进了函数行列式，即"雅可比行列式"，指出函数行列式在多重积分的变量替换中的作用，给出了函数行列式的导数公式。雅可比的著名论文《论行列式的形成和性质》标志着行列式系统理论的建成。行列式在数学分析、几何学、线性方程组理论、二次型理论等多方面的应用，促使了行列式理论在19世纪得到了很大发展。整个19世纪都有行列式的新结果。除了一般行列式的大量定理之外，还有许多有关特殊行列式的其他定理都相继得到。

矩 阵


西尔维斯特

矩阵是数学中的一个重要的基本概念，是代数学的一个主要研究对象，也是数学研究和应用的一个重要工具。"矩阵"这个词是由西尔维斯特（James Joseph Sylvester，1814－1897）首先使用的，他是为了将数字的矩形阵列区别于行列式而发明了这个术语。矩阵这个词来源于拉丁语，代表一排数。而实际上，矩阵这个课题在诞生之前就已经发展的很好了。从行列式的大量工作中明显地表现出来，为了很多目的，不管行列式的值是否与问题有关，方阵本身都可以研究和使用，矩阵的许多基本性质也是在行列式的发展中建立起来的。在逻辑上，矩阵的概念应先于行列式的概念，然而在历史上发生的次序正好相反。


凯莱

英国数学家凯莱(A.Cayley，1821－1895)一般被公认为是矩阵论的创立者，因为他首先把矩阵作为一个独立的数学概念提出来，并首先发表了关于这个题目的一系列文章。凯莱在研究线性变换下的不变量相结合时，首先引进矩阵以简化记号。1858年，他发表了关于这一课题的第一篇论文《矩阵论的研究报告》，系统地阐述了关于矩阵的理论。文中他定义了矩阵的相等、矩阵的运算法则、矩阵的转置以及矩阵的逆等一系列基本概念，指出了矩阵加法的可交换性与可结合性。他用单一的字母 $\boldsymbol{A}$ 来表示矩阵是对矩阵代数发展至关重要的，其公式 $\det(\boldsymbol{AB})=\det(\boldsymbol{A})\det(\boldsymbol{B})$ 为矩阵代数和行列式间提供了一种联系。另外，凯莱还给出了方阵的特征方程和特征根（特征值）以及有关矩阵的一些基本结果。凯莱出生于一个古老而有才能的英国家庭，剑桥大学三一学院大学毕业后留校讲授数学，三年后他转从律师职业，工作卓有成效，并利用业余时间研究数学，发表了大量的数学论文。

1855年，埃米特(Cherie，1822－1901)证明了别的数学家发现的一些矩阵类的特征根的特殊性质，如现在称为埃米特矩阵的特征根性质等。后来，克莱伯施(A. Clebsch，1831－1872)、布克海姆

(A. Buchheim)等证明了对称矩阵的特征根性质。泰伯(H. Taber)引入矩阵的迹的概念并给出了一些有关的结论。

在矩阵论的发展史上，弗罗伯纽斯(G.Frobenius，1849－1917)的贡献是不可磨灭的。他讨论了最小多项式问题，引进了矩阵的秩、不变因子和初等因子、正交矩阵、矩阵的相似变换、合同矩阵等概念，以合乎逻辑的形式整理了不变因子和初等因子的理论，并讨论了正交矩阵与合同矩阵的一些重要性质。1854 年，约当研究了矩阵化为标准型的问题。

矩阵本身所具有的性质依赖于元素的性质。矩阵由最初作为一种工具经过两个多世纪的发展，现在已成为独立的一门数学分支——矩阵论。而矩阵论又可分为矩阵方程论、矩阵分解论和广义逆矩阵论等矩阵的现代理论。矩阵及其理论现已广泛地应用于现代科技的各个领域。

矩阵的发展是与线性变换密切相连的。到 19 世纪它还仅占线性变换理论形成中有限的空间。现代向量空间的定义是由皮亚诺（Peano）于 1888 年提出的，皮亚诺以公理的方式定义了有限维或无限维向量空间。二次世界大战后随着现代数字计算机的发展，矩阵又有了新的含义，特别是在矩阵的数值分析等方面。由于计算机的飞速发展和广泛应用，许多实际问题可以通过离散化的数值计算得到定量的解决。于是作为处理离散问题的线性代数，成为从事科学研究和工程设计的科技人员必备的数学基础。

线性方程组

刘徽

线性方程组的解法早在中国古代的数学著作《九章算术》的方程一章中已作了比较完整的论述。《九章算术》是综合性的历史著作，原作者不详，据研究西汉的张苍、耿寿昌曾经做过增补，魏晋时的数学家刘徽做过详细注解。刘徽定义了若干数学概念，全面论证了《九章算术》的公式解法，提出了许多重要的思想、方法和命题。在这部书的手稿中解释了如何消去变元的方法求解带有三个未知量的三方程系统，其中所述方法实质上相当于现代的对方程组的增广矩阵施行初等行变换从而消去未知量的方法，即高斯消元法。在西方，线性方程组的研究是在 17 世纪后期由莱布尼茨开创的。他曾研究含两个未知量的三个线性方程组组成的方程组。麦克劳林在 18 世纪上半叶研究了具有二、三、四个未知量的线性方程组，得到了现在称为克莱姆法则的结果。克莱姆不久也发表了这个法则。18 世纪下半叶，法国数学家贝祖对线性方程组理论进行了一系列研究，证明了 n 个 n 元齐次线性方程组有非零解的条件是系数行列式等于零。

19 世纪，英国数学家史密斯(H.Smith)和道奇森(C.L.Dodgson)继续研究线性方程组理论，前者引进了方程组的增广矩阵和非增广矩阵的概念，后者证明了 n 个未知数 m 个方程的方程组相容的充要条件是系数矩阵和增广矩阵的秩相同。这正是现代方程组理论中的重要结果之一。

道奇森

大量的科学技术问题，最终往往归结为解线性方程组。因此在线性方程组的数值解法得到发展的同时，线性方程组解的结构等理论性工作也取得了令人满意的进展。现在，线性方程组的数值解法在计算数学中占有重要地位。

二次型

高斯

二次型也称为“二次形式”。二次型的系统研究是从 18 世纪开始的，它起源于对二次曲线和二次曲面的分类问题的讨论。将二次曲线和二次曲面的方程变形，选有主轴方向的轴作为坐标轴以简化方程的形状，这个问题是在 18 世纪引进的。柯西在其著作中给出结论：当方程是标准形时，二次曲面用二次项的符号来进行分类。然而，那时并不太清楚，在化简成标准形时，为何总是得到同样数目的正项和负项。西尔维斯特回答了这个问题，他给出了 n 个变数的二次型的惯性定理，但没有证明。这个定理后被雅可比重新发现和证明。 1801 年，高斯在《算术研究》中引进了二次型的正定、负定、半正定和半负定等术语。

2. 怎样学习线性代数

（注：本文是一篇我国的代数名家丘维声教授在电大讲授线性代数课程时关于如何学好线性代数的综合论述，超牛。特摘录如下，与诸君共赏。）

初学线性代数的同学们一定想了解，线性代数的主要内容是什么？学习线性代数有哪些用处？这门课程有什么特点？学习时应注意些什么？本文想就这几个方面的问题谈谈自己的体会。

一、 线性代数的内容简介

同学们在中学里都学过代数（即初等代数）。初等代数的主要内容是：研究数及其运算。由于用字母表示数，因此应用题可以列方程来解，解方程就成为初等代数的一个中心问题。

由于初等代数研究的是数的运算和方程，所以代数的方法其优点在于可以运算，正因为这样，人们很早就运用代数的方法去研究几何图形的性质，即建立坐标系，用坐标表示点，用方程表示图形，这就是解析几何。线性代数正是随着解析几何的研究而发展起来的。

在解析几何里我们知道：在平面上直线的方程是一次方程，反之，任一个二元一次方程在平面上表示一条直线。由于这个几何直观，通常把一次方程称为线性方程。当然，这里“线性”的意思是指变量的次数是一次，并不是说，线性方程都表示直线。事实上，在空间中，一次方程表示一个平面。在几何里，经常要求几个平面的交点，这转化为代数上就是要解三元一次方程组，即三个未知量的线性方程组。数学的各个分支、物理及许多实际问题则进一步提出要解几个未知量的线性方程组：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\; \vdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n = b_s \end{cases} \tag{1}$$

其中：$x_1, x_2, \ldots, x_n$ 是未知量；a_{ij} 表示第 i 个方程中 x_j 的系数；b_j 称为常数项。

线性方程组（1）在什么时候有解？如何求解？解的结构如何？这就是线性代数要研究的第一个问题。在求解线性方程组时，行列式是一个重要的工具，所以先要研究行列式。

线性方程组完全被它的全部系数和常数项所决定，至于未知量用什么符号表示是没有什么关系的。因此，线性方程组（1）可以用下面

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{sn} & b_s \end{bmatrix} \tag{2}$$

这张表来表示。这种由一些数排列成的若干行（横的）、若干列（纵的）所组成的表称为一个矩阵。于是研究线性方程组就转化为研究矩阵。

除了线性方程组之外，还有大量的各种各样的问题也都提出矩阵的概念，譬如，解析几何中，平面直角坐标系的转轴公式为

$$\begin{cases} x = x'\cos\theta - y'\sin\theta \\ y = x'\sin\theta + y'\cos\theta \end{cases} \tag{3}$$

显然，新旧坐标之间的关系完全可以通过公式（3）中系数排列成的矩阵

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4}$$

表示出来。

矩阵是线性代数的一个主要研究对象。矩阵已经不是数，而是由一些数排列成的一张表，但是矩阵可以像数那样进行运算。矩阵的运算有加法、数量乘法、乘法等。矩阵由于可以进行运算，因此它的用途就很广，特别是矩阵的乘法很独特，矩阵的应用很多就是由矩阵的乘法而来的。矩阵除了可以进行运算外，还可以进行一些交换变成较简单的矩阵，根据不同问题的需要，对矩阵进行不同的变换。矩阵的变换有初等变换、合同变换、相似变换等。由于矩阵可以进行运算和交换，因此它成为了一个强有力的数学工具，希望同学们像熟练掌握微积分一样熟练掌握矩阵。

在解析几何中，平面上二次曲线的一般方程是

$$ax^2+2bxy+cy^2+2dx+2ey+f=0 \tag{5}$$

为了便于研究这条二次曲线的性质，需要坐标交换，把方程（5）化成标准方程。如果方程（5）中没有交叉项（即 xy 项），那么只需要配方，然后作移轴就可以化成标准方程了，于是关键在于如何消去交叉项，而这实质就是要把

$$ax^2+2bxy+cy^2 \tag{6}$$

化成平方和

$$a'x'^2+c'y'^2 \tag{7}$$

的形式。式（6）是 x 和 y 的二次齐次多项式，简称为 x，y 的二次齐次或二次型。于是把二次曲线方程（5）化成标准方程的关键是要把二次型（6）化成平方和（7）的形式。因此，二次曲线问题的提出需要研究二次型。此外，数学的其他分支以及物理力学中也常常会碰到二次型。在研究二次型时，矩阵是一个有力的工具。

在解析几何中研究了向量。我们知道向量可以进行运算，有加法、数量乘法运算等，并满足加法交换律等八条运算规律（以下简称八条规律）。

在研究线性方程组时，一个线性方程 $a_1x_1+a_2x_2+\cdots+a_nx_n=b$ 可以用有序数组 $(a_1, a_2,\cdots,a_n, b)$ 表示，这种有序数组也称为向量。一个 n 元有序数组 $(a_1, a_2,\cdots,a_n)$ 称为一个 n 维向量。n 维向量之间也可以规定加法和数量乘法两种运算，这两种运算同样满足八条规律。矩阵也有加法、数量乘法，并且这两种运算也满足八条规律。像向量、矩阵这些对象已经不是数，但是它们都有加法、数量乘法这两种运算（加法、数量乘法称为线性运算），并且满足加法交换率等八条规律。这样的一些对象很多，譬如，多项式、连续函数等也都具有加法、数量乘法两种运算，且满足八条规律。代数是研究运算的，线性代数就是研究具有加法、数量乘法这两种线性运算的集合。像向量组成的集合、矩阵组成的集合，都是具有两种线性运算，且满足八条规律的集合，这样的集合就称为线性空间。线性空间是线性代数研究的最基本的对象。解析几何里，全体向量组成的集合就是一个线性空间。我们都知道，若取定了三个不共面的向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$，则空间中任何一个向量 $\boldsymbol{a}$ 都可以表示成 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$ 的线性组合：$\boldsymbol{a}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+x_3\boldsymbol{e}_3$。这说明通过三个不共面的向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$ 就能表示出空间中的全部向量，于是称这空间是三维的，并且称这三个不共面的向量 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$ 是这空间的一组基。维数和基的概念可以推广到抽象的线性空间中。

线性空间是某一类事物从量的方面的一个抽象。我们认识客观事物，固然要弄清它们单个的和总体的性质，但更重要的是研究它们之间的各种各样的联系。在线性空间中，事物之间的联系就反映为线性空间的映射。线性空间 V 到自身的映射通常称为 V 的一个变换。其中最简单的一种变换就是线性变换，即保持线性运算的变换：

$$\begin{cases} \Gamma(\boldsymbol{\alpha}+\boldsymbol{\beta})=\Gamma(\boldsymbol{\alpha})+\Gamma(\boldsymbol{\beta}) \\ \Gamma(k\boldsymbol{\alpha})=k\Gamma(\boldsymbol{\alpha}) \end{cases} \tag{8}$$

譬如在几何空间中，关于平面 Π 的投影就是一个线性变换。这是因为，若向量 $\boldsymbol{\alpha}$ 的投影是 $\boldsymbol{\alpha}'$，向量 $\boldsymbol{\beta}$ 的投影是 $\boldsymbol{\beta}'$，则易看出 $\boldsymbol{\alpha}+\boldsymbol{\beta}$ 的投影是 $\boldsymbol{\alpha}'+\boldsymbol{\beta}'$。这表明投影是保持向量加法的。同样，容易看出 $k\boldsymbol{\alpha}$ 的投影是 $k\boldsymbol{\alpha}'$，这表明投影是保持数量乘法的。

线性变换是线性代数的重要研究对象。线性变换与矩阵有着非常密切的联系，在线性空间中若取定一组基，则一个线性变换可以对应到一个矩阵，这个对应是一一对应，并且这个对应是保持运算的。因此矩阵的好些问题可以通过研究线性变换来得到解决，而线性变换的好些问题又可以利用矩阵的工具来解决，它们互相渗透，密切相连。

以上所述的线性方程组、矩阵、二次型、线性空间和线性变换就是线性代数研究的主要对象。由于电大“线性代数”这门课程的学时比较少，我们就不讲抽象的线性空间了，而讲具体的 n 维向量空间；也不讲线性变换的概念了，而把线性变换的主要问题用矩阵的语言来讲。在这里顺便指出，高等代数是包括线性代数和多项式理论这两部分内容的，而我们仅学习线性代数这部分内容。

二、 学习线性代数的用处

由于线性代数研究的是具有加法和数量乘法这两种运算的集合，而具有这两种线性运算的对象是很多的，因此线性代数的内容有广泛的应用。

线性方程组是最简单的方程组，在数学的各个分支、物理以及工程技术实际问题中会大量遇到线性方程组。因此给了一个线性方程组，它有没有解？若有解，如何求解？这都是必须明确的问题，以便于我们在数学的各个分支以及实际工作中能正确而迅速地解决所遇到的线性方程组的问题。

矩阵是非常重要的数学工具，许多数学问题、实际问题都可以引出矩阵，并且可以通过对矩阵进行运算或变换得到解决。例如，在微分方程组中，要解微分方程组：

$$\begin{cases}\dfrac{dy_1}{dx}=y_1+4y_2\\ \dfrac{dy_2}{dx}=y_1+y_2\end{cases}\tag{9}$$

其中 y_1、y_2 是未知函数。

式（9）右边的系数可以排成一个矩阵 $\boldsymbol{A}=\begin{bmatrix}1&4\\1&1\end{bmatrix}$。如果记 $\dfrac{d}{dx}\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}\dfrac{dy_1}{dx}\\ \dfrac{dy_2}{dx}\end{pmatrix}$，则利用矩阵的乘法可以把式（9）写成

$$\frac{d}{dx}\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}1&4\\1&1\end{bmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix}\tag{10}$$

作未知函数的线性替换

$$\begin{cases}y_1=t_{11}z_1+t_{12}z_2\\ y_2=t_{21}z_1+t_{22}z_2\end{cases}\tag{11}$$

式（11）也可以利用矩阵的乘法写成

$$\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}t_{11}&t_{12}\\t_{21}&t_{22}\end{bmatrix}\begin{pmatrix}z_1\\z_2\end{pmatrix}\tag{12}$$

记 $\boldsymbol{T}=\begin{bmatrix}t_{11}&t_{12}\\t_{21}&t_{22}\end{bmatrix}$，将式（12）代入式（10），得

$$\frac{d}{dx}\left(\boldsymbol{T}\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)=\boldsymbol{AT}\begin{pmatrix}z_1\\z_2\end{pmatrix}$$

即

$$\boldsymbol{T}\frac{d}{dx}\begin{pmatrix}z_1\\z_2\end{pmatrix}=\boldsymbol{AT}\begin{pmatrix}z_1\\z_2\end{pmatrix}\tag{13}$$

在式（13）两边左乘 $\boldsymbol{T}^{-1}$（$\boldsymbol{T}$ 的逆矩阵），得

$$\frac{d}{dx}\begin{pmatrix}z_1\\z_2\end{pmatrix}=\boldsymbol{T}^{-1}\boldsymbol{AT}\begin{pmatrix}z_1\\z_2\end{pmatrix}\tag{14}$$

如果能选择矩阵$\boldsymbol{T}$，使得$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$成为对角矩阵$\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$，那么式（14）变成

$$\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{15}$$

利用矩阵的乘法，式（15）就成为

$$\begin{cases} \dfrac{\mathrm{d}z_1}{\mathrm{d}x} = a_1 z_1 \\ \dfrac{\mathrm{d}z_2}{\mathrm{d}x} = a_2 z_2 \end{cases} \tag{16}$$

显然，微分方程组（16）是很容易求解的。这个例子说明，通过矩阵的运算和矩阵的变换，可以把一个复杂的问题变成一个较简单的问题，所以矩阵是一个很有用的工具。

二次型理论首先的一个应用是解决了二次曲线，特别是二次曲面的分类问题。其次，系数是实数的二次型的正定、负定、不定性被用在数学分析中解决多元函数的极大值、极小值问题。二次型理论在概率论与数理统计以及物理力学中也有重要应用。

值得指出的是，我们学习线性代数，除了线性代数的内容有广泛的应用以外，还在于线性代数这门学科考虑问题、解决问题的思想方法对我们有很大的帮助。譬如，代数学总是要对所研究的对象进行分类，在每类中取一个最简单的对象作为代表来进行研究。关于这点，举一个例子，对于两个n阶矩阵（n行n列的矩阵）$\boldsymbol{A}$、$\boldsymbol{B}$，如果存在一个可逆矩阵$\boldsymbol{T}$，使得

$$\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$$

则$\boldsymbol{A}$、$\boldsymbol{B}$称为相似。按照相似这个关系，全体n阶矩阵就被分成好多类，每一类里的矩阵是彼此相似的。有人会问：在每一类里能否找到一个最简单的矩阵作为代表？上面提到的微分方程组的例子中，关于选择$\boldsymbol{T}$，使得$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$成为对角矩阵，就是这里所说的问题。这个问题在第五章矩阵的标准形中将得到解决。代数学的上述这种研究问题的方法对我们解决其他数学分支中的问题以至于实际中的一些问题都是可取的。又譬如在代数学里，经常要在一类对象里找出具有某一特定性质的特殊对象，举一个例子：什么样的n阶矩阵是可逆矩阵？在解决这种问题时，总是先考虑如果矩阵$\boldsymbol{A}$是可逆矩阵，那么它将具有哪些性质；然后再考虑具有这些性质的矩阵是否一定是可逆矩阵。这种考虑问题的方法是：先缩小范围，然后在这缩小了的范围内来审查这些对象是否都是具有某一特定性质的对象。这种考虑问题的方法也是可取的。像这种在学习一门课程的时候，除了学习这门课程的内容外，还注意学习这门学科考虑问题的一些方法也是很有必要的。这样我们学到的就不仅是一些现成的结论，而且有考虑问题、解决问题的一些方法。

三、 线性代数的特点

代数学研究问题的基本方法是，先从某些具有公共性质的对象抽象出它们最基本的几条性质作为定义，然后从这些定义出发进行逻辑推理，揭示出这类对象的新的性质。譬如线性空间这一概念就是这样抽象出来的，把具有加法和数量乘法这两种运算，并且满足八条规律作为线性空间的定义，然后从线性空间的定义出发进行逻辑推理，揭示出线性空间的许多性质，这些性质由于是只用到线性空间的定义进行逻辑推理得到的，并没有用到线性空间里的元素的具体特性，因此这样推导出来的性质就对所有的线性空间都适用，不管是几何空间、n维向量空间、还是矩阵组成的空间、连续函数组成的空间，都具有这些性质。可见，这种方法的好处在于它能使应用非常广泛。代数学的这种研究问题的方法称为公理化的方法或抽象的方法，这种公理化的方法使得线性代数具有如下几个特点：

第一，概念多而且抽象。譬如在第三章线性方程组中，将遇到向量组的线性相关、线性无关，等价的向量组，向量组的极大无关组，向量组的秩，矩阵的秩等一系列概念，而且这些概念都比较抽象。概念上的高度抽象性是近代数学的特点之一。由此而来的则是应用上的极其广泛性，这也是近代数学的一个特点。在线性代数中，由线性方程组抽象出n维向量和矩阵这两个概念，又从n维向量空间抽象出线性空间的概念，一次比一次

抽象。抽象的东西不如具体的东西好学，但是抽象的东西比具体的东西应用要广，所以还是需要认真下功夫把它学好。

第二，逻辑性强。由于代数学的基本方法是由具体事务抽象出一般概念，再运用这些概念逻辑推理揭示出事物新的性质，因此线性代数这门课程的逻辑性强，一环扣一环。

第三，明显的几何背景。这在前面介绍线性代数的内容时已多次提到。这个特点使我们在学习线性代数里的抽象的概念时，可以借助几何直观来帮助我们理解这些概念，应该充分利用这一特点。

四、学习线性代数的方法

根据线性代数的上述特点，我们在学习线性代数时应当注意以下几方面。

第一，要把概念弄清楚，理解确切并且记住。如果概念不清楚，模模糊糊，那就没有办法运用概念去进行逻辑推理，做题时就不知道如何下手。因此在学习中，应当首先复习概念、定理，弄明白了，然后再开始做作业。表面上看起来，似乎比一开始就做作业要慢，实际上，先复习概念、定理、例题，然后再做作业，可以使作业做得比较顺利，反倒省时间。更何况，我们如果没有弄清楚概念，那么只会做跟例题差不多的题，稍灵活一些，变换一下题目，就不会做了。所以希望同学们务必先复习概念、定理、例题，然后再做作业。由于线性代数逻辑性强，后面的内容需要用到前面的好些概念、定理，如果每次课，没有及时复习、消化，那么时间越长，学的概念、定理越多，脑子里就会堆积着一大堆模模糊糊的东西，这样学起后面的内容就很吃力。而如果每次课后，都能及时复习、及时消化，就会越学越顺利。那么怎样才能把概念弄清楚呢？一般来说应当从以下几方面着手:

(1) 先要弄清楚这个概念是怎样提出来的？它的背景是什么？

(2) 这个概念的确切内容是什么？

(3) 多举一些具体的例子来帮助理解抽象的概念，特别是举一些几何上的例子，比较直观、形象。

譬如，向量组线性相关、线性无关这个概念，它是怎么提出来的呢？为什么要学习这个概念呢？这是因为我们要研究线性方程组什么时候有解。由于每个线性方程可用一个向量表示，于是线性方程组是否有解必然跟表示这些方程的一组向量是怎么样的向量组有关系。从几何上看，两个向量 $\boldsymbol{a}$、$\boldsymbol{b}$ 组成的向量组有两种可能，即 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线，或者 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不共线。前一情形，$\boldsymbol{a}$ 与 $\boldsymbol{b}$ 在一条直线上，我们称向量组 $\boldsymbol{a}$、$\boldsymbol{b}$ 线性相关；后一情形，$\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不在一条直线上，我们称 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 组成的向量组线性无关。为了能把线性相关、线性无关的概念推广到 n 维向量空间中，需要对几何上的向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线在代数上如何表示加以分析。在解析几何里知道，$\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线的充分必要条件是其中一个向量是另一个向量的倍数，即 $\boldsymbol{a}=\lambda\boldsymbol{b}$ 或者 $\boldsymbol{b}=\mu\boldsymbol{a}$，其中 λ、μ 是实数，也就是 $1\cdot\boldsymbol{a}-\lambda\boldsymbol{b}=\boldsymbol{0}$ 或者 $\mu\boldsymbol{a}-1\cdot\boldsymbol{b}=\boldsymbol{0}$。换言之,存在不全为 0 的数 k_1、k_2，使得 $k_1\boldsymbol{a}+k_2\boldsymbol{b}=\boldsymbol{0}$。从这几何背景，引出向量组线性相关的定义：对于向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$（$s\geqslant1$），如果存在一组不全为 0 的数 $k_1,k_2,\cdots,k_s$，使得

$$k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\cdots+k_s\boldsymbol{a}_s=\boldsymbol{0} \tag{17}$$

则称向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ （$s\geqslant1$）线性相关。

在理解这个定义时，要注意“不全为 0”这几个字。因为对于任何向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ 来说，都有

$$0\cdot\boldsymbol{a}_1+0\cdot\boldsymbol{a}_2+\cdots+0\cdot\boldsymbol{a}_s=\boldsymbol{0}$$

所以如果在线性相关的定义中不添加上“不全为 0”的限制，那么就毫无意义了。何况从几何背景上我们知道，$\boldsymbol{a}$ 与 $\boldsymbol{b}$ 共线的充分必要条件是存在不全为 0 的数 k_1、k_2，使得 $k_1\boldsymbol{a}+k_2\boldsymbol{b}=\boldsymbol{0}$，所以只有当向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ 存在不全为 0 的数 $k_1,k_2,\cdots,k_s$，使得式（17）成立时，才是线性相关。一个向量组不线性相关就称这个向量组线性无关。例如，几何上两个向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 若不共线，就称向量组 $\boldsymbol{a}$、$\boldsymbol{b}$ 线性无关。既然向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ 不线性相关，就称为线性无关，因此向量组线性无关的定义又可叙述如下:

对于向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$，如果从

$$k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\cdots+k_s\boldsymbol{a}_s=\boldsymbol{0}$$

可推出 $k_1=k_2=\cdots=k_s=0$，则称向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ 线性无关。这就是说，对于向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$，如果只有一组全为 0 的数，使得式（17）成立，则称向量组 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_s$ 线性无关。

为了更好地把握向量线性无关的定义，我们在学习中，还应该多通过一些具体的例子去深入体会。

第二，要努力培养逻辑推理的能力，即运用概念和已知的定理、性质进行推理、判断的能力。为此，形式逻辑的一些基本常识是应当熟悉的。譬如，命题有四种形式：原命题、逆命题、否命题、逆否命题。若原命题正确，则逆否命题一定正确，但是逆命题和否命题不一定正确。

要能进行逻辑推理，就必须熟记概念和定理、性质，否则就没有武器，遇到一个题就不知道如何下手。

第三，学习每一章、每一节时，都要明确这章、这节要研究什么问题，是如何解决的。这样做，学习中头脑就是清醒的；如果不这样做，就会稀里糊涂，不知道现在在干什么。在学习一个定理时也要这样，要明确为什么要有这个定理，这个定理是如何证明的，主要是抓住证明的思路，弄清是怎样去证明的。如果坚持这么做了，就能不断学到一些方法，从而提高分析问题、解决问题的能力。

第四，学习线性代数跟学习任何一门数学课一样，必须适当多做一些习题。光听课、光看书，自己不动手做题，那是学不好数学的。只有通过做题，才能加深对概念、定理的理解，才能学到一些方法。做习题时，一定要自己动脑子想，不要轻易翻习题解答，只有当想了很久确实想不出来时，才翻一下习题解答，稍微看一两眼，得到一点提示就可以不看了，自己再去想。只有自己动脑子做出来的题才是真的会了，如果是自己动脑子不够，轻易就看习题解答，那么很快就会忘的。做完题要注意小结，看看这样一类问题应当如何下手，千万不要任务观点，赶紧做完作业就放到一边。应当想一想，通过做这几个题有什么收获，学到什么方法。这样日积月累，能力就逐渐提高了。还要指出一点，做计算题时要算到底，不要因为算起来较麻烦就不愿意往下算了，认为反正我方法会了。这样下去是不行的。其实算得对、算得快，这也是需要培养的一种能力，不在平时做作业时训练，就会在今后实际工作中算题时尽出错。

线性代数概念较多，也比较抽象一点，但是只要我们下决心，并且注意学习方法，是能够学好的。

主要参考文献

[1] Lay. C. 线性代数及其应用. 刘深泉,等,译. 北京: 机械工业出版社，2005.
[2] .Strang. C 线性代数及其应用. 侯自新,等,译. 天津: 南开大学出版社，1990.
[3] 索普.J,佩尔.P 线性代数基础. 钱辉镜, 杨宗仁,等,译. 北京: 中央广播电视大学出版社，1988.
[4] 李尚志. 线性代数. 北京: 高等教育出版社，2006.
[5] 张远达. 线性代数原理. 上海: 上海教育出版社，1998.
[6] 陈怀琛，龚杰民. 线性代数实践及 MATLAB 入门. 北京:电子工业出版社，2005.
[7] 刘学质. 线性代数的数学思想方法. 北京: 中国铁道出版社，2006.
[8] 吕世虎, 李军. 基本初等矩阵的几何意义及其在教学中的运用. 数学教育学报，2008.
[9] Comap. 数学的原理与实践. 申大维,等,译. 北京: 高等教育出版社，1998.
[10] 王兵. 几何学的思想与方法. 济南: 山东大学出版社，2009.
[11] 居余马,等. 线性代数. 北京: 清华大学出版社，1995.
[12] 柯朗.R,约翰.F. 微积分和数学分析引论. 北京: 科学出版社，2001.
[13] 贺霖. 实用矢量代数. 济南: 山东科学技术出版社，1983.
[14] 陈维翰. 线性关系及其应用. 重庆: 重庆大学出版社，1989.
[15] 蓝以中. 线性代数引论. 北京: 北京大学出版社，1981.
[16] 邱森. 高等代数. 武汉: 武汉大学出版社，2008.
[17] 温书. 坐标变换的 Jacobian 的几何意义及其应用. 淮阴工业专科学校学报, 2000, 9(1).

后　记

本书原想写得又薄又精炼，没想打印下来还挺厚的，删除谁都有点可惜，就请诸位将其权当线性代数概念的几何意义手册参阅。再说了，如果东西写得清晰而详尽的话，读起来就感觉不到厚度了。因此，后期的想法是不求本子薄，只求理儿明，不知道做到没有。

本书的第一稿倒是写得不太慢，从儿子出生（07 年 9 月）开始写，一年多就弄完了。放在网上折腾了两年，后来网友们反应热烈纷纷建议出版，就联系到了母校——西安电子科技大学出版社。在毛红兵编辑的悉心指导下开始着手修改文稿。

常常在安静的深夜，守着一叠厚厚的处处用红笔勾画过的稿子，俺一边修改着稿子，一边感叹着出书的不易，一边感念着编辑老师的辛劳。

排版修改太费时间了，几乎全部图片、公式至少重打了 3 遍以上，工作一忙有时一两个月都没有动一个字。本书的姗姗来迟令俺愧对网友的热望，有时都不敢再到网上去浏览网友的热切留言。

感谢毛红兵编辑的耐心和信任，感谢诸位网友的批评指正，感谢辛勤的合作者谢聪博士、胡翠芳老师，感谢家人在背后默默的支持让我有闲情逸趣来写这本纯粹的“闲书”。事项进行到这总算给热情的读者有点交待了。

最后，一个老生常谈却必须要说的是：作者业余的数学水平有限，缺憾谬误和笔误一定很多（虽然已修正了很多），请诸位亲们勿怕麻烦，把这些脸上雀、肉中刺指点出来。俺无它以报，当赠书为念。

联系邮箱：renguangqian@126.com。

作者　任广千

2015 年 3 月

几何意义名言录

没有任何东西比几何图形更容易印入脑际了，因此用这种方式来表达事物是非常有意义的。

——笛卡儿

算术符号是文字化的图形，而几何图形则是图像化的公式；没有一个数学家能缺少这些图像化的公式。

——希尔伯特

如果代数与几何各自分开发展，那它的进步十分缓慢，而且应用范围也很有限。但若两者互相结合而共同发展，则就会相互加强，并以快速的步伐向着完善化的方向猛进。

——拉格朗日

不会几何学就不会正确的思考，而不会正确思考的人不过是行尸走肉。

——柏拉图

无论是从事数学教学或研究， 我是喜欢直观的。学习一条数学定理及其证明，只有当我能把定理的直观含义和证明的直观思路弄明白了， 我才认为真正懂了。

——中国当代数学家徐利治