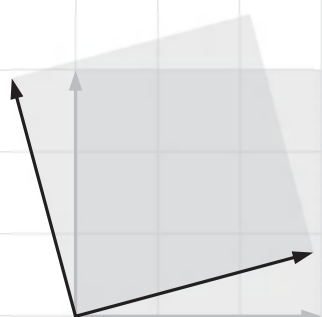




线性代数入门

梁 鑫 田 垠 杨一龙 编著

清华大学出版社
北 京

内 容 简 介

线性代数是当代大学生的必修科目,也是当前科学技术领域的数学基础和通用语言.随着数据科学的发展,线性代数的地位越来越重要.本书前六章是线性代数的基础,主要讨论 $\mathbb{R}^n$ 的结构,内容包括向量、矩阵、子空间、内积、行列式、特征值等;后两章是对基础内容的升华,主要讨论抽象的线性空间、线性映射、内积等内容.

本书可作为线性代数课程的教材或参考书.

本书封面贴有清华大学出版社防伪标签,无标签者不得销售.

版权所有,侵权必究.举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn.

图书在版编目(CIP)数据

线性代数入门/梁鑫,田垠,杨一龙编著. —北京:清华大学出版社, 2022.6

ISBN 978-7-302-60971-1

I. ①线… II. ①梁… ②田… ③杨… III. ①线性代数 IV. ①O151.2

中国版本图书馆 CIP 数据核字(2022)第 089506 号

责任编辑:刘颖

封面设计:傅瑞学

责任校对:王淑云

责任印制:丛怀宇

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-83470000

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:艺通印刷(天津)有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:21

字 数:468千字

版 次:2022年7月第1版

印 次:2022年7月第1次印刷

定 价:59.80元

产品编号:097005-01

前言

线性代数是当代高等院校学生的必修科目，也是当前科学技术领域的数学基础和通用语言。随着时代的发展和社会的进步，计算机技术和数据科学逐步渗透、应用甚至开始主导不少科学研究和工程技术领域。在新形势下，线性代数变得越来越重要。另一方面，为了更好地建设国家、服务社会，教师和学生都迫切地希望能更快更好地教会、学会基础课程。教好、学好线性代数，离不开体系完整、简明易懂的入门教材。本书正是为此而编写的。

线性代数研究线性空间和线性映射理论，其基础是向量和矩阵的性质。本书精心选择了教学内容，重点阐述最主要的、最常用的知识点，并辅以常见而易于理解的工程实例。

绪论简述了线性代数的整体框架，作者希望读者能够在学前和学后认真阅读，以便纲举目张，对线性代数有一个整体性认识。第 0 章主要用来应对中小学数学教育内容不断减少的问题，供读者了解数学的基本逻辑工具。第 1 章至第 4 章是最基本的主干内容，可以认为是中学数学的简单拓展，是各个专业的学生都务求精熟的部分；第 5 章至第 6 章是另一部分基本内容，是自然科学、工程技术、社会科学等专业的学生应当掌握的内容。这两部分是线性代数的初等部分，是抽象理论的具体实例。第 7 章至第 8 章是线性代数中抽象理论的介绍，是线性代数的真正入门。读者应注意随时对照初等部分，发现相同和不同，加深对抽象理论的理解。掌握这部分内容的读者应能在学习其他课程中不断发现和应用线性空间和线性映射理论。这样安排教学内容，既利于读者由浅入深，自具体而抽象地学习线性代数，又便于课时不足的读者只学习部分内容就能了解到线性代数的概貌。

“熟能生巧”“练习铸就大师”。学好线性代数，必须得有习题辅助。本书安排了大量不同类型、由易到难的习题，以供读者使用。只有当学生清楚地知道自己答题是否正确时，才能谈得上学会相应的知识。“学而不思则罔，思而不学则殆。”中学生常有些不好的学习习惯，如只看正文不做题，或不看正文只做题。作者希望读者能在练习中不断熟悉和复习相关知识，当习题做不出时反复阅读思考正文中的知识点，尝试使用相关知识点来解决问题。另外，读者应当认识到，有些较难的习题不是每个人都能做出，但只要尽力思考，即使没有找到解决问题的办法，也已经在思考过程中复习了相关知识，锻炼了分析能力，并非一无所获。

本书将定义、定理、命题、推论、例、注等统一编号为 $a.b.c$ ，表示第 a 章第 b 节第 c 项定义、定理或其他内容，其中加  的章节属于选讲内容；将习题、阅读材料等集中放在每节之后，并统一编号为 $a.b.c$ ，表示第 a 章第 b 节第 c 项习题或阅读材料，其中

加 ☞ 或 ☞☞ 的习题是较难或更难的选做题. 另外, 为提示读者, 本书中的例以 ☺ 结束, 证以 □ 结束, 其他部分材料换用不同于正文的字体.

本书最后列出名词索引、人名表、符号表, 以供读者查阅. 正文中出现的人名除华人外一般以拉丁转写形式出现, 并在人名表中提供了原文和常见译法.

本书可作为高等院校线性代数课程的教材, 也可作为自学或复习的参考书. 本着体系的完整性和严谨性, 作者对所有命题都提供了证明, 有些命题还有多个证明, 从而使知识体系在逻辑上尽量扎实可靠. 教师可以根据教学要求和课时限制自行调整. 根据作者的授课经验, 完整讲述本书内容, 第 1 章至第 4 章用 35~40 学时, 第 5 章至第 6 章用 15~20 学时, 第 7 章至第 8 章用 16~20 学时, 供参考.

本书参考了不少国内外教材, 部分习题也来源于此. 现列出供读者参阅:

- [1] G. Strang. 线性代数 [M]. 5 版. 北京: 清华大学出版社, 2019.
- [2] A.И. Кострикин. 代数学引论 (一) [M]. 2 版. 张英伯, 译. 北京: 高等教育出版社, 2006.
- [3] A.И. Кострикин. 代数学引论 (二) [M]. 3 版. 牛凤文, 译. 北京: 高等教育出版社, 2008.
- [4] 丘维声. 简明线性代数 [M]. 北京: 北京大学出版社, 2002.
- [5] 蓝以中. 高等代数简明教程 (上、下) [M]. 北京: 北京大学出版社, 2002.
- [6] 史明仁. 线性代数六百证明题详解 [M]. 北京: 北京科学技术出版社, 1985.

本书是在清华大学数学科学系和丘成桐数学科学中心的领导与线性代数教学团队的大力支持下完成的, 并在 2020 年初稿完成后在清华大学大范围试用了两年, 获得了大量颇有价值的意见. 作者感谢张友金、朱敏娴、刘思齐、王浩然、曹晋、蔚辉、孙晟昊、袁瑶、刘余及其他同事对写书的倡议和支持、对初稿的指正与对体系的建议, 并感谢探微书院化 12 班刘恒瑀同学对书稿的建议. 作者感谢本书责任编辑刘颖为本书付出的辛勤劳动.

本书得到了清华大学教学改革项目的资助.

作者始终努力使本书准确可靠, 但精力和水平所限, 书中的错误在所难免. 欢迎广大读者对本书提出宝贵意见, 指出本书的逻辑上的、文字上的、排版上的任何错误. 作者向所有关心本书的人致以诚挚的感谢!

梁 鑫 田 垠 杨一龙

2022 年 1 月于北京

目 录

绪论	1
第 0 章 预备知识	7
0.1 逻辑与集合	7
0.2 间接证明法	9
0.3 映射	13
第 1 章 线性映射和矩阵	16
1.1 基本概念	16
1.2 线性映射的表示矩阵	27
1.3 线性方程组	37
1.4 线性映射的运算	49
1.4.1 线性运算	49
1.4.2 复合	52
1.5 矩阵的逆	65
1.5.1 初等矩阵	65
1.5.2 可逆矩阵	66
1.5.3 等价关系	71
1.6 分块矩阵	76
1.7 LU 分解	87
第 2 章 子空间和维数	94
2.1 基本概念	94
2.2 基和维数	104
2.2.1 向量的线性关系	104
2.2.2 基和维数	108
2.3 矩阵的秩	112
2.4 线性方程组的解集	120

第 3 章 内积和正交性	130
3.1 基本概念	130
3.1.1 内积	130
3.1.2 标准正交基	133
3.2 正交矩阵和 QR 分解	139
3.2.1 正交矩阵	139
3.2.2 QR 分解	142
3.3 子空间和投影	149
3.3.1 正交补	151
3.3.2 正交投影	153
3.3.3 最小二乘问题	155
第 4 章 行列式	162
4.1 引子	162
4.2 行列式函数	165
4.3 行列式的展开式	176
第 5 章 特征值和特征向量	187
5.1 引子	187
5.2 基本概念	191
5.3 对角化和谱分解	199
5.4 相似	208
第 6 章 实对称矩阵	219
6.1 实对称矩阵的谱分解	219
6.2 正定矩阵	227
6.3 奇异值分解	234
6.3.1 基本概念	235
6.3.2 低秩逼近	239
第 7 章 线性空间和线性映射	246
7.1 线性空间	246
7.2 基和维数	253
7.3 线性映射	260
7.4 向量的坐标表示	266
7.5 线性映射的矩阵表示	274

第 8 章 内积空间	287
8.1 欧氏空间	287
8.2 欧氏空间上的线性映射	297
8.3 酉空间	305
8.3.1 基本概念	305
8.3.2 快速 Fourier 变换	310
名词索引	318
人名表	324
符号表	325

清华大学出版社

绪 论

线性代数这门学科，以研究**线性系统**为目的，是一个从理论到应用都相当完善的数学分支。线性系统是什么呢？

我们先从系统说起。系统是指由若干个相互作用和相互依赖的事物组合而成的具有特定功能的一个整体。例如一个由若干部件组成的机器，可以对“输入”进行运算或处理，得到“输出”：

输入 $\rightarrow$ 系统/机器 $\rightarrow$ 输出.

而线性系统是指，这个系统的输入与输出符合叠加原理，即如下两条性质：输入放大或缩小某一倍数，则产生的输出也放大或缩小相同倍数，这称为**齐次性**；两组输入所产生的总输出是二者分别产生的独立输出之和，这称为**可加性**。

下面来看几个具体的例子。

- 一个大型代工厂是一个系统：以工人的工时作为输入，以不同产品的数量作为输出。二者的关系满足叠加原理。
- 由刚性杆组成的复杂支架是一个系统：在支架上端放置若干重物，将重物重量作为输入；支架下端以若干地秤支撑，将地秤测得的重量作为输出。根据 Newton 定律，二者的关系满足叠加原理。
- 由电阻组成的电路是一个系统：在若干对节点间接入电压表，将电压值作为输入；在若干节点处接入电流表，将电流值作为输出。根据 Kirchhoff 定律，二者的关系满足叠加原理。

这些例子是大量实际问题中的几个代表。今后读者会发现，日常生产生活中总会遇到类似的问题。这些问题有时规模很大、未知数很多，但都可以利用线性代数知识使用计算机来解决。我们将看到，线性系统虽然应用十分丰富，但它的数学本质并不复杂。

给定一个线性系统，它内部的运作方式我们还不清楚。能够清楚知道的是，它需要 n 个输入 $x_1, x_2, \dots, x_n$ ，并生成 m 个输出 $y_1, y_2, \dots, y_m$ 。换句话说，目前它还是一个黑盒子，给出任意输入，我们不知道它会输出什么。

想要搞清楚系统如何运作，我们可以对它进行测试，即给定一组输入，测量其输出。我们先考虑最简单的输入，即只对某一个输入项输入：

- 先指定 $x_1 = 1$ ，其他输入项是 0，这时我们测出输出项依次为 $y_1 = a_{11}, y_2 = a_{21}, \dots, y_m = a_{m1}$ 。
- 再指定 $x_2 = 1$ ，其他输入项是 0，这时输出项依次为 $y_1 = a_{12}, y_2 = a_{22}, \dots, y_m =$

a_{m2} .

- 依次类推,直到指定 $x_n = 1$,其他输入项是 0,这时输出项依次为 $y_1 = a_{1n}, y_2 = a_{2n}, \dots, y_m = a_{mn}$.

根据叠加原理中的齐次性,我们可以得到以下结论:

- 当 x_1 任意,其他输入项是 0 时,输出项必然为 $y_1 = a_{11}x_1, y_2 = a_{21}x_1, \dots, y_m = a_{m1}x_1$.
- 当 x_2 任意,其他输入项是 0 时,输出项必然为 $y_1 = a_{12}x_2, y_2 = a_{22}x_2, \dots, y_m = a_{m2}x_2$.
- 依次类推,当 x_n 任意,其他输入项是 0 时,输出项必然为 $y_1 = a_{1n}x_n, y_2 = a_{2n}x_n, \dots, y_m = a_{mn}x_n$.

再根据叠加原理中的可加性,我们就知道,任意输入项 $x_1, x_2, \dots, x_n$ 与其输出项 $y_1, y_2, \dots, y_m$ 满足:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = y_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = y_2, \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = y_m. \end{cases} \quad (0.0.0)$$

可以看到,任意输入项和输出项之间的关系构成了一个多元一次方程组,又称为线性方程组.这就是线性系统得名的原因.

容易看到,整个线性系统由 a_{ij} 这 mn 个数唯一确定.换句话说,只要进行了前述的 n 组测试,得到了这些 a_{ij} ,这个线性系统对我们来说就变得透明,不再是一个黑盒子了.

四个基本问题 现在假设某个线性系统已知,或者说任意输入项与输出项之间的关系已知.如果系统有一个未知的输入,能否通过观察其输出来反推这个输入呢?对于不同的系统,可能会出现以下三种情形:

1. 存在唯一一组输入得到给定的输出.例如,由电阻组成的电路这一系统,对输出的电流,只可能有唯一的输入电压.
2. 存在不只一组输入得到给定的输出.那么根据叠加原理,一定有无穷多组输入都能得到给定的输出,这是因为如果两组输入 $u_1, u_2, \dots, u_n$ 和 $v_1, v_2, \dots, v_n$ 得到同一输出,那么 $tu_1 + (1-t)v_1, tu_2 + (1-t)v_2, \dots, tu_n + (1-t)v_n$ 也将会得到相同的输出.例如,服装厂这一系统,对输出指定数量的衣服,既可以是这个工人花费工时做的,也可以是那个工人花费工时做的,也可能是这些工人按照任意比例分配工时完成的.这里的输入有无数种可能.
3. 不存在输入得到给定的输出.事实上这意味着,无论输入什么都不可能得到这种输出!尽管听起来略显怪异,但实践中仍有可能发生.例如,CT 扫描机这一

系统, 输入人, 输出 X 光片. 受灰尘、机械错位、计算机误差等因素影响, 可能得到的 X 光片和真实情况有微小偏差. 这个微小偏差很可能使得这个 X 光片不能百分之百地符合任何人的真实情况. 另一方面, 实际上这个误差并不影响医生诊断, X 光片不需要完全真实, 只要它和真实情况足够接近就足够了. 换句话说, 测量总有误差, 因此测量结果并不一定是系统的真实输出. 但理论上讲, 测量值和真实的输出应该非常接近. 尽管测量结果不可能从输入得到, 但是我们可以所有可能的输出结果中找到最接近测量结果的那个输出. 它应该是最有可能的真实的系统输出.

在数学上, 我们是在研究线性方程组 (0.0.0), 其中输出 $y_1, y_2, \dots, y_m$ 已知, 要求解的是未知的输入 $x_1, x_2, \dots, x_n$.

在上面提到的三种情形中, 前两种情形分别意味着线性方程组 (0.0.0) 有唯一解或有无穷多解, 此时我们希望计算出解. 对这两种情形的考察构成了线性代数的一个核心问题——**线性方程组求解问题**.

而后一种情形意味着线性方程组 (0.0.0) 无解, 这时我们希望计算出某组 $x_1, x_2, \dots, x_n$, 使得方程组左端的值和右端给定的值最接近. 对这种情形的考察构成了线性代数的一个重要问题——**最小二乘问题**.

在很多实际问题中, 我们常常需要考虑自反馈系统, 即将给定输入通过系统演化产生的输出作为新的输入继续演化, 并不断重复这一过程直到达到我们需要的演化次数为止, 或者直到系统稳定下来, 即输出和输入相同. 例如, 对某个范围的人群, 将每一年的入口数量作为输入, 都可以得到作为输出的第二年的入口数量, 而这个输出可以再次作为输入, 得到第三年的入口数量, 以此类推. 我们希望得到第三十年的人口数量, 或者当人口数量不再变化时的人口数量. 我们可以提出问题:

1. 如何计算出达到演化次数时的输出? 当然我们可以逐次计算, 但是否有快速计算的方法?
2. 如何计算系统稳定时的输入或输出?

对这两个问题的考察构成了线性代数的另一个核心问题——**特征值问题**.

有时我们希望用简单的系统来模拟一个复杂的系统, 即用远少于 mn 个数来描述由前述 mn 个数所代表的系统. 例如, 一张数码照片, 由于计算机存储不足或网络带宽限制, 我们希望对它进行压缩. 压缩的过程中可能会丢失一些信息, 但我们希望, 在保证压缩程度的同时丢失的信息尽量少. 这需要考虑这样的问题:

1. 如何计算出这样的简单系统?
2. 如何衡量模拟的好坏?

对这两个问题的考察构成了线性代数的另一个重要问题——**奇异值问题**.

线性方程组求解问题、最小二乘问题、特征值问题和奇异值问题, 是线性代数理论与应用中的**基本问题**, 矩阵计算就是围绕着这四个问题展开的。

两个核心概念 数学上, 为了便于开展对前述的系统、输入、输出等理念及上面提出的基本问题进行研究, 我们需要将其数学化.

在考虑线性系统时, 我们常常把一组输入或一组输出称为一个**向量**; 所有可以叠加的向量构成的集合称为一个**线性空间**. 由于每个输入项和输出项都是一个数, 当它们都只有有限多个时, 向量, 作为一组输入或者一组输出, 常常可以写成一个数组.

向量, 也就是线性空间中的元素, 最重要的特点是可以进行线性运算. 这是叠加原理赋予向量的运算.

而线性系统的处理过程, 或者说从输入到输出的对应关系, 称为一个**线性映射**. 特别地, 如果输入项和输出项都只有有限多个, 对应的线性映射常常可以写成一个数表, 称为**矩阵**.

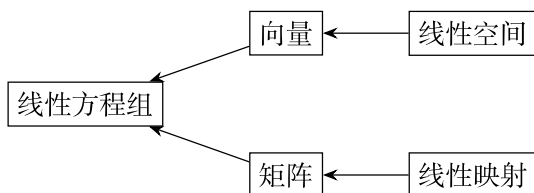
以线性方程组 (0.0.0) 为例, 输入 x_j 组成的数组就是向量 $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, 输出 y_i 组成的

数组也是向量 $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$. 而描述系统只需要知道 a_{ij} 的值, 它们组成的数表就是矩阵

$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. 我们还可以把线性方程组 (0.0.0) 写成:

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

可以看到, 线性方程组可以把向量和矩阵有机地联系起来.



上面介绍了许多概念. 这些概念总结如下:

- 线性空间和线性映射是对所研究的问题提出的高层次概念, 是我们推导的基石和思考问题的凭借;
- 向量和矩阵是所研究的问题自然导出的低层次概念, 是我们计算的对象和获得结果的媒介;
- 线性方程组是向量和矩阵之间的联系, 很多问题的解决方案常常归结于对一个线性方程组求解.

线性空间和线性映射是线性代数研究中的**核心概念**, 可以说线性代数就是关于线性空间和线性映射的代数学.

研究思想和常用技巧 在学习、掌握到精通线性代数的过程中, 在利用线性代数尝试解决实际问题时, 我们常常使用许多数学思想来攻克问题, 获得解决方案.

(1) 简化问题, 化全局为局部 例如, 前面提到线性方程组的解如果不止一组, 就必然有无穷多组, 但是这无穷多组解是由两组解获得的. 反过来问, 如果我们知道有无穷多组解, 那么这些解能否由有限组解获得? 又如前所述, 我们获得一个有 n 个输入项的线性系统的全部信息, 并不需要知道所有可能的输入对应的输出, 只需要知道 n 组输入对应的输出就足够了. 这二者启发了线性空间的基和子空间分解等概念, 落实在矩阵上就是矩阵分块的技巧.

(2) 分解问题, 化复杂为简单 如前述, 一个有 n 个输入项、 m 个输出项的线性系统可以用 mn 个数来表示, 这可能是很复杂的; 但有些线性系统的演化可能产生简单的效果. 如果一个线性系统可以分解成一系列简单的系统的组合, 那么线性系统的特性将容易把握得多. 这启发了映射标准形的概念, 落实在矩阵上就是矩阵分解的技巧.

(3) 归约问题, 化未知为已知 面对一个解决方案尚不明确的问题, 最直接的指导方针并不是随便尝试以期起效, 而是将问题划归成一个解决方案已经存在的问题. 有些读者熟知的数学归纳法就是一例. 如果我们能把某个问题都化归成已获解决的若干类简单问题之一, 例如已提及的标准形, 那么问题将容易获得解决, 这启发了等价类的概念, 落实在矩阵上就是矩阵变换的技巧.

矩阵分块、矩阵分解、矩阵变换是线性代数中最常用的推理和计算技巧. 可以看到, 矩阵是利用线性代数解决问题的**计算重点**.

总的来说, 我们这门课程将围绕着上述四个基本问题、两个核心概念展开, 利用上述研究思想和推理技巧, 来阐述完善的数学理论, 推导有效的计算方法, 并为后续的理论科学和应用技术课程提供必要的基础.

下面我们再举一些小例子, 读者可以先自行思考一下问题的解决之道.

- 解方程组: 一百根参数已知的相同弹簧串联在一起, 两端挂在给定间距的天花板和地板上, 如何计算每根弹簧拉伸的长度.
- 最小二乘: 有人说严刑峻法能威慑潜在犯罪分子从而遏制犯罪; 有人说严刑峻法变相鼓励犯罪分子犯重罪; 现有世界上 58 个国家死刑执行率与重罪率的统

计数据，如何用数据来判断不同说法的正确性.

- 特征值：甲有 100 元，乙和丙没有钱；甲将自己的钱的一半给乙，乙再将自己的钱的三分之一给丙，丙再将自己的钱的四分之一给甲，然后甲再将自己的钱的一半给乙，依次循环；如何计算最终达到稳定状态时三人的钱数.
- 奇异值：给定十个村庄的位置，修一条笔直的高速公路，再修一些小路将村庄和高速公路连通，如何修路能够使各小路长度的平方和最小.

在学习过线性代数后，读者将能够用一张纸徒手计算，得到上述问题的答案.

还有一些读者在中学时可能思考过但不曾得到结论的数学问题，在学习过线性代数后，将完全能够口述其中的原理.

- 多元一次方程组的所有解如何只用有限多个解表示；
- 如何获得一个一元高次方程的所有解的近似值；
- 平面上的二次曲线只有三种非退化的情形；
- 平面上保持距离的变换只有平移、旋转、镜射和滑移镜射；
- 平面上不改变共线关系的变换一定是保持距离的变换和正压缩的复合.

现在就让我们开始学习精巧而美妙的线性代数吧！

第 0 章 预备知识

0.1 逻辑与集合

首先回顾数学中常用的逻辑术语.

数学中, 可以判断真假的陈述句称为**命题**. 其中判断为真的命题称为**真命题**, 判断为假的命题称为**假命题**. 一个命题为真, 常称为命题成立.

在进行逻辑推理时, 最常使用的语句是

如果命题 A 成立, 那么命题 B 成立 或简写成 若 A , 则 B .

其中命题 A 常称为**条件**, 命题 B 常称为**结论**. 如果由 A 通过推理可以得出 B , 则上述语句是真命题. 此时称 A **蕴涵** B , 又称 A 是 B 的**充分条件**, 也称 B 是 A 的**必要条件**, 记作 $A \Rightarrow B$, 或 $B \Leftarrow A$. 如果 A 既是 B 的充分条件又是 B 的必要条件, 则称 A 是 B 的**充分必要条件**, 也称 A 和 B **等价**, 或 A **当且仅当** B , 记作 $A \Leftrightarrow B$.

例如, 平面上两个三角形, 如果全等, 则一定相似. 因此全等是相似的充分条件, 相似是全等的必要条件, 即“全等 $\Rightarrow$ 相似”. 反过来, 如果两个三角形相似, 并不一定全等. 因此相似不是全等的充分条件, 全等也不是相似的必要条件. 进而可知, 相似和全等并不是充分必要条件.

命题添加逻辑联结词可以得到新的命题. 用“且”联结命题 A, B , 得到新命题“ A 且 B ”, 记作 $A \wedge B$. A 且 B 成立当且仅当 A 和 B 同时成立. 用“或”联结命题 A, B , 得到新命题“ A 或 B ”, 记作 $A \vee B$. A 或 B 成立当且仅当 A 和 B 至少有一个成立. 用“非”联结命题 A , 得到新命题“非 A ”, 记作 $\neg A$. 非 A 成立当且仅当 A 不成立.

含有变量的陈述句不一定能判断真假, 也就不一定是命题, 例如, “矩形是正方形”. 如果在这种句子中添加一些对变量的限定词 (称为**量词**), 就能使其成为命题. 例如, “存在一个矩形是正方形”或“至少有一个矩形是正方形”是真命题. “存在”和“至少有一个”这种限定词称为**存在量词**, 用“ $\exists$ ”表示. 又如, “所有矩形都是正方形”或“任意矩形都是正方形”或“每一个矩形都是正方形”是假命题. “所有”和“任意”这种限定词称为**全称量词**, 用“ $\forall$ ”表示.

其次回顾关于集合的基本知识.

在数学研究中, 研究对象常称为**元素**, 而所研究对象的全体称为一个**集合**. 集合最重要的基本性质有两条:

1. 给定集合的元素是确定的, 即给定一个集合和一个元素, 必须能够对元素是否在集合中这一问题做出明确的判断;
2. 给定集合中的元素互不相同, 即集合中的元素不重复出现.

不含任何元素的集合称为**空集合**, 记作 $\emptyset$. 最常用的集合是由数组成的集合. 自然数的全体构成的集合称为**自然数集**, 记作 $\mathbb{N}$. 类似地, 有整数集 $\mathbb{Z}$ 、有理数集 $\mathbb{Q}$ 、实数集 $\mathbb{R}$ 、复数集 $\mathbb{C}$.

集合可以用列举法或描述法表示. 例如, 在平面上取定平面直角坐标系 Oxy 之后, 以 O 点为圆心的单位圆上横纵坐标都是整数的点所组成的集合 A 可以用如下记号表示:

$$A = \{(1, 0), (0, 1), (-1, 0), (0, -1)\} = \{(x, y) \mid x, y \in \mathbb{Z}, x^2 + y^2 = 1\}.$$

如果 x 是集合 X 的元素, 就称 x **属于** 集合 X , 记作 $x \in X$; 如果 x 不是集合 X 的元素, 称 x **不属于** 集合 X , 记作 $x \notin X$.

给定集合 X 和 Y , 如果 X 中的任意元素都在 Y 中, 则称 X 是 Y 的**子集**, 或 Y **包含** X , 记作 $X \subseteq Y$ 或 $Y \supseteq X$. 如果 X 和 Y 互为子集, 则两个集合的元素完全相同, 则称 X 与 Y **相等**, 记作 $X = Y$. 显然, 它与 $X \subseteq Y$ 且 $Y \subseteq X$ 等价. 证明 $X \subseteq Y$ 和 $Y \subseteq X$ 都成立是今后证明 X, Y 两个集合相等的一个基本方法. 如果 $X \subseteq Y$, 但 $X \neq Y$, 则称 X 是 Y 的**真子集**, 记作 $X \subsetneq Y$. 空集认为是任何集合的子集.

集合 X 与 Y 的公共元素组成的集合称为 X 与 Y 的**交 (集)**, 记作 $X \cap Y$. 显然, $X \cap Y \subseteq X, X \cap Y \subseteq Y$. 如果 X 与 Y 没有公共元素, 则 $X \cap Y = \emptyset$. 容易看出, $X \cap Y = X$ 与 $X \subseteq Y$ 等价. 把 X 与 Y 中的所有元素合并在一起组成的集合称为 X 与 Y 的**并 (集)**, 记作 $X \cup Y$. 显然, $X \subseteq X \cup Y, Y \subseteq X \cup Y$. 容易看出, $X \cup Y = Y$ 与 $X \subseteq Y$ 等价. 从集合 X 中去掉属于 Y 的元素之后剩下的元素组成的集合, 称为 X 对 Y 的**差 (集)**, 记作 $X \setminus Y$. 特别地, 如果 $Y \subseteq X$, 则称 $X \setminus Y$ 是 X 中 Y 的**补 (集)**.

容易发现, 逻辑运算的且、或、非, 直接对应于集合运算中的交、并、补. 而命题的蕴涵关系对应于集合的包含关系. 我们有以下类比.

例 0.1.1 给定命题 A, B 和集合 S , 令

$$X = \{s \in S \mid s \text{ 满足条件 } A\}, \quad Y = \{s \in S \mid s \text{ 满足条件 } B\}.$$

1. $X \cap Y = \{s \in S \mid s \text{ 满足条件 } A \wedge B\}.$
2. $X \cup Y = \{s \in S \mid s \text{ 满足条件 } A \vee B\}.$
3. $S \setminus X = \{s \in S \mid s \text{ 满足条件 } \neg A\}.$
4. $X \subseteq Y$ 和 $A \Rightarrow B$ 等价.

设 X, Y, Z 是集合 S 的子集, 容易验证以下运算律:

1. 交满足交换律和结合律, 即 $X \cap Y = Y \cap X, (X \cap Y) \cap Z = X \cap (Y \cap Z);$
2. 并满足交换律和结合律, 即 $X \cup Y = Y \cup X, (X \cup Y) \cup Z = X \cup (Y \cup Z);$

3. 交对并满足分配律, 即 $(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z)$;
4. 并对交满足分配律, 即 $(X \cap Y) \cup Z = (X \cup Z) \cap (Y \cup Z)$;
5. De Morgan 定律, 即 $S \setminus (X \cap Y) = (S \setminus X) \cup (S \setminus Y)$, $S \setminus (X \cup Y) = (S \setminus X) \cap (S \setminus Y)$;
6. 包含关系具有传递性, 即如果 $X \subseteq Y, Y \subseteq Z$, 那么 $X \subseteq Z$.

类似地, 对命题 A, B, C , 有以下运算律:

1. 且满足交换律和结合律, 即 $A \wedge B \Leftrightarrow B \wedge A, (A \wedge B) \wedge C \Leftrightarrow A \wedge (B \wedge C)$;
2. 或满足交换律和结合律, 即 $A \vee B \Leftrightarrow B \vee A, (A \vee B) \vee C \Leftrightarrow A \vee (B \vee C)$;
3. 且对或满足分配律, 即 $(A \vee B) \wedge C \Leftrightarrow (A \wedge C) \vee (B \wedge C)$;
4. 或对且满足分配律, 即 $(A \wedge B) \vee C \Leftrightarrow (A \vee C) \wedge (B \vee C)$;
5. De Morgan 定律, 即 $\neg(A \wedge B) \Leftrightarrow (\neg A) \vee (\neg B), \neg(A \vee B) \Leftrightarrow (\neg A) \wedge (\neg B)$.
6. 蕴涵关系具有传递性, 即如果 $A \Rightarrow B, B \Rightarrow C$, 那么 $A \Rightarrow C$.

以上运算规律的证明均留给读者.

☺

习题

练习 0.1.1 画出例 0.1.1 中的所有集合运算律的 Venn 图.

练习 0.1.2 对任意两个命题 A, B , $A \Rightarrow B$ 可以定义为 $B \vee (\neg A)$. 直观上可以如此理解: $B \vee (\neg A)$ 成立等价于 A 不成立或 B 成立, 于是当 A 成立时必有 B 成立. 运用例 0.1.1 中的逻辑运算律, 证明 $(A \Rightarrow B) \Leftrightarrow ((\neg B) \Rightarrow (\neg A))$.

练习 0.1.3 对任意三个命题 A, B, C , 运用例 0.1.1 中的逻辑运算律, 证明逻辑运算符合模律, 即当 A 是 C 的必要条件时, $(A \wedge B) \vee C \Leftrightarrow A \wedge (B \vee C)$.

练习 0.1.4 对任意两个命题 A, B , 我们定义运算异或 $A \oplus B$ 为 $(A \wedge (\neg B)) \vee (B \wedge (\neg A))$.

1. 按例 0.1.1 的条件画出 $A \oplus B$ 对应的 Venn 图.
2. 证明异或满足交换律和结合律.

0.2 间接证明法

数学结论的正确性必须通过逻辑推理加以证明. 除直接从条件出发推出结论外, 也可以利用逻辑进行间接证明.

反证法 容易看出, “如果命题 A 成立, 则命题 B 成立”, 和 “如果命题 B 不成立, 则命题 A 不成立” 这两个命题等价. 这就是反证法的理论基础. 反证法是间接证明的方法之一, 即当需要证明 $A \Rightarrow B$ 时, 先假设命题 B 不成立, 再想办法证明命题 A 也不成立. 特别地, 当没有命题 A , 只有命题 B 时, 只需假设命题 B 不成立, 然后推出与已知条件或事实相抵触的条件 (称为矛盾) 即可.

例 0.2.1 证明: $\sqrt{2}$ 是无理数.

☺

证 假设命题不成立, 即 $\sqrt{2}$ 是有理数, 则可令 $\sqrt{2} = \frac{n}{m}$, 其中 n, m 是互素的整数. 易得 $2m^2 = n^2$, 于是 n 是偶数. 设 $n = 2s$, 则 $2m^2 = 4s^2, m^2 = 2s^2$, 因此 m 也是偶数. 这与 n, m 互素的假设矛盾. $\square$

例 0.2.2 证明: 素数有无穷多个. $\odot$

证 假设命题不成立, 即只有有限个素数, 设为 $p_1, p_2, \dots, p_n$. 令 $N = p_1 p_2 \cdots p_n + 1 > 1$. 显然任何 p_i 都不能整除 N . 这与任何大于 1 的正整数都能写成素数乘积这一结论矛盾. 根据反证法推出, 一定存在无穷多个素数. $\square$

由上面两个例子可以看出, 反证法往往适用于“命题 B 不成立”比较好处理的情形.

数学归纳法 数学归纳法是一种数学证明方法, 通常被用于证明某个给定命题在整个自然数范围内成立, 包括第一数学归纳法和第二数学归纳法.

(1) **第一数学归纳法**: 给定一个与自然数 n 有关的命题 $P(n)$, 如果

1. 可以证明当 n 取初始值 n_0 时命题 $P(n_0)$ 成立;
2. 假设当 $n = k (k \geq n_0)$ 时命题 $P(k)$ 成立, 可以证明当 $n = k + 1$ 时命题 $P(k + 1)$ 也成立;

则对一切自然数 $n \geq n_0$, 命题 $P(n)$ 都成立.

例 0.2.3 证明平方求和公式: $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$. $\odot$

证 当 $n = 1$ 时, 命题显然成立.

假设当 $n = k$ 时命题成立, 即 $1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}$. 当 $n = k + 1$ 时, 可得 $1^2 + 2^2 + \cdots + k^2 + (k+1)^2 = \frac{k(k+1)(2k+1)}{6} + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6}$, 此时命题也成立.

根据第一数学归纳法可知, 命题对所有正整数 n 都成立. $\square$

(2) **第二数学归纳法**: 给定一个与自然数 n 有关的命题 $P(n)$, 如果

1. 可以证明当 n 取初始值 n_0 时命题 $P(n_0)$ 成立;
2. 假设当 $n \leq k (k \geq n_0)$ 时命题 $P(n)$ 成立, 可以证明当 $n = k + 1$ 时命题 $P(k + 1)$ 也成立;

则对一切自然数 $n \geq n_0$, 命题 $P(n)$ 都成立.

例 0.2.4 考虑 Fibonacci 数列 $F_0 = 0, F_1 = 1, F_n = F_{n-1} + F_{n-2}, n = 2, 3, \dots$, 证明其通项公式 $F_n = \frac{x_1^n - x_2^n}{x_1 - x_2}$, 其中 $x_1 = \frac{1 + \sqrt{5}}{2}, x_2 = \frac{1 - \sqrt{5}}{2}$ 是方程 $x^2 - x - 1 = 0$ 的两个根. $\odot$

证 当 $n = 0$ 和 $n = 1$ 时, 命题显然成立.

假设当 $n \leq k$ ($k \geq 2$) 时命题成立, 特别地, 当 $n = k-1, k$ 时命题成立. 当 $n = k+1$ 时, 有

$$\begin{aligned} F_{k+1} &= F_k + F_{k-1} \\ &= \frac{x_1^k - x_2^k + x_1^{k-1} - x_2^{k-1}}{x_1 - x_2} \\ &= \frac{x_1^{k-1}(x_1 + 1) - x_2^{k-1}(x_2 + 1)}{x_1 - x_2} \\ &= \frac{x_1^{k-1}x_1^2 - x_2^{k-1}x_2^2}{x_1 - x_2} = \frac{x_1^{k+1} - x_2^{k+1}}{x_1 - x_2}, \end{aligned}$$

即命题也成立.

根据第二数学归纳法可知, 命题对所有自然数 n 都成立. □

由这个例子可以看出, 对比第一数学归纳法, 第二数学归纳法第二条中的假设条件更强, 因此适用范围更广.

读者需要注意, 由于在例 0.2.4 的递推步骤中, $n = k+1$ 被化简成了 $n = k-1, k$ 的情形, 这就要求 $k \geq 2$. 也就是说, 在递推步骤中, $k = 0, 1$ 都是不适用的. 因此, 在验证初始值时, 既需要验证 $k = 0$, 也需要验证 $k = 1$. 总之, 凡是递推步骤不适用的 n 的取值, 全都需要单独进行验证.

下面举一个错误使用数学归纳法的例子, 其错误的核心就是忘记检验全部初始条件. 读者务必引以为戒!

例 0.2.5 我们将 (错误地) 证明, 世界上所有的马都是同一种颜色的.

当 $n = 1$ 时, 只有一匹马, 命题显然成立.

假设当 $n \leq k$ 时命题成立. 当 $n = k+1$ 时, 在 $k+1$ 匹马中, 除第一匹马外, 剩下的 k 匹马是同一种颜色的, 另一方面, 除第二匹马外, 剩下的 k 匹马也是同一种颜色的. 这样一来, 全部 $k+1$ 匹马都只能是同一种颜色的.

于是命题对所有正整数 n 都成立.

这个证明的错误之处就在于, 递推步骤中实际上隐含了假设 $n \geq 3$ (读者可以思考一下为什么). 因此这个归纳过程既需要验证 $n = 1$ 的情形, 也需要验证 $n = 2$ 的情形. 而验证一下很容易发现, $n = 2$ 时命题为假. 因此后面的递推步骤毫无意义. ☹

我们这里再给两个例子. 这两个例子仅需验证一种初始条件, 但是递推步骤时, 每个 $n = k$ 都需要用到各个 k 之前所有情形.

例 0.2.6 假设有一个 $a \times b$ 的巧克力排块. 那么需要掰几次才能把巧克力排块掰成 1×1 的小块呢? 下面证明, 无论怎样掰, 都需要 $ab - 1$ 次才能做到 (参见图 0.2.1).

当 $a = b = 1$ 时, 命题显然成立.

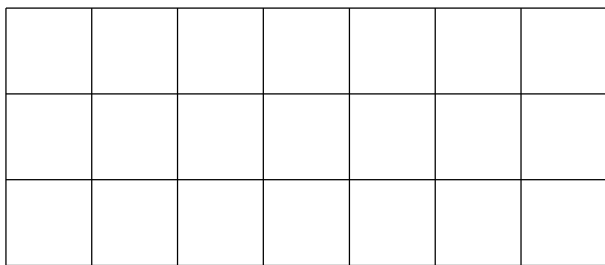


图 0.2.1 掰巧克力

假设 a, b 中至少有一个大于 1, 不妨设 $a > 1$. 如果将 a 行排块掰成 k 行和 $a - k$ 行, 那么还剩 $k \times b$ 的排块和 $(a - k) \times b$ 的排块需要掰. 根据归纳假设, 还需要 $(kb - 1) + ((a - k)b - 1)$ 次才能掰成 1×1 的小块. 于是共需要掰 $1 + (kb - 1) + ((a - k)b - 1) = ab - 1$ 次. ☺

例 0.2.7 令 p_n 为自然数中的第 n 个素数. 我们来证明, 第 n 个素数 p_n 一定小于等于 $2^{2^{n-1}}$. ☺

证 当 $n = 1$ 时, 命题显然成立.

假设当 $n \leq k$ 时命题成立, 下面考虑 p_{k+1} . 由例 0.2.2 可得, $p_{k+1} \leq p_1 p_2 \cdots p_k + 1$. 根据归纳假设, 我们有

$$p_{k+1} \leq p_1 p_2 \cdots p_k + 1 \leq 2^{2^0} 2^{2^1} \cdots 2^{2^{k-1}} + 1 \leq 2^{2^0 + 2^1 + \cdots + 2^{k-1}} + 1 \leq 2^{2^k - 1} + 1 \leq 2^{2^k}.$$

因此命题成立. □

习题

练习 0.2.1 请用反证法证明以下命题.

1. 若 $p > 2$ 是素数, 则 p 是奇数.
2. 对正整数 n , 若 n^2 是奇数, 则 n 是奇数.
3. 不存在最小的正有理数.

练习 0.2.2 经典逻辑的基本公理之一是排中律, 即对任意命题 A , A 不能既不真又不假. 反证法可以看作排中律的推论, 即如果我们发现 A 不能是错的, 那么 A 就只能是对的. 因此, 想要证明命题 A 真, 不妨挑一个命题 B , 先证明 $B \Rightarrow A$, 再证明 $(\neg B) \Rightarrow A$. 那么 A 就必须是真的了. 给定命题: 存在两个无理数 a, b , 满足 a^b 是一个有理数. 请先假设 $\sqrt{2}^{\sqrt{2}}$ 是有理数, 再假设 $\sqrt{2}^{\sqrt{2}}$ 是无理数, 针对两种情形讨论来证明这个命题.

练习 0.2.3 证明 $2^{2^{n-1}} - 1$ 必然至少有互异的 $n - 1$ 个奇素因数. (因此第 n 个素数 p_n 一定小于等于 $2^{2^{n-1}}$.)

提示: 运用因式分解 $x^2 - 1 = (x + 1)(x - 1)$; 并注意对任意 $m \neq n$, $2^{2^m} + 1$ 与 $2^{2^n} + 1$ 互素.

练习 0.2.4 假设有一个 $a \times b$ 的巧克力排块, 我们需要将它掰成 1×1 的小块, 在掰的同时会被打分. 假设掰一次将 $x + y$ 块掰成了 x 块和 y 块, 则得分 xy . 证明: 当巧克力被完全掰成 1×1 的小块时, 总得分永远为 $\frac{1}{2}ab(ab - 1)$.

0.3 映射

最后回顾关于映射的基本概念.

设有集合 X 和 Y , 如果 X 中的任意元素 x , 都以某种法则 f 对应于 Y 中唯一一个元素, 则称这个对应的法则 f 是集合 X 到集合 Y 的一个**映射**, X 中的元素 x 所对应的 Y 中的元素常记作 $f(x)$.

映射用如下记号表示:

$$\begin{aligned} f: X &\rightarrow Y, \\ x &\mapsto f(x). \end{aligned}$$

这里, 集合 X 称为映射 f 的**定义域**, 集合 Y 称为映射 f 的**陪域**. 设 $x \in X, y = f(x) \in Y$, 我们称 y 为 x 在映射 f 下的**像**, x 为 y 在 f 下的**原像**. 注意, Y 中的元素不一定都有原像; 有原像时, 原像也可能不唯一. 定义域 X 中所有元素在 f 下的像构成 Y 的一个子集 $\{f(x) \mid x \in X\}$, 称为映射 f 的**像集或值域**, 记作 $f(X)$.

两个映射 $f, g: X \rightarrow Y$ 相等, 当且仅当对任意 $x \in X$, $f(x) = g(x)$.

根据映射的像集和原像的性质, 我们考虑如下三种特殊情形:

1. 如果对 X 中的任意两个不同元素 x_1, x_2 , 都有 $f(x_1) \neq f(x_2)$, 则称 f 是一个**单射**; 即, 任何元素的原像至多只有一个.
2. 如果对 Y 中的任意元素 y , 都存在 X 中的某个 x , 使得 $y = f(x)$, 则称 f 是一个**满射**; 即, f 的像集 $f(X) = Y$.
3. 映射 f 既是单射又是满射, 则称 f 是一个**双射**. 双射 f 给出了集合 X 和 Y 之间的一个一一对应.

映射之间最重要的运算是两个映射的复合. 设 X, Y, Z 是三个集合, $f: X \rightarrow Y$, $g: Y \rightarrow Z$ 是两个映射, 则可以定义一个映射

$$\begin{aligned} h: X &\rightarrow Z, \\ x &\mapsto h(x) = g(f(x)), \end{aligned}$$

称其为 g 与 f 的**复合**, 记作 $h = g \circ f$. 可以如下表示:

$$\begin{aligned} X &\xrightarrow{f} Y \xrightarrow{g} Z, \\ x &\mapsto f(x) \mapsto g(f(x)). \end{aligned}$$

注意, 只有当 f 的陪域与 g 的定义域相同时, 才能定义二者的复合. 设 X, Y, Z, W 是四个集合, $f: X \rightarrow Y, g: Y \rightarrow Z, h: Z \rightarrow W$ 是三个映射, 则两种不同顺序的复合是同一个 X 到 W 的映射: $h \circ (g \circ f) = (h \circ g) \circ f$, 即映射的复合运算满足结合律. 因此, 这个映射可以写成 $h \circ g \circ f$ 而不会引起混淆.

当映射 f 的定义域和陪域都是 X 时, 称 f 是集合 X 上的一个**变换**. 容易看到, 一个变换永远可以跟自身复合, 记 $f^2 = f \circ f, f^n = f^{n-1} \circ f, n \geq 1$. 任意集合 X 上都有

一个特殊的变换, 它把 X 中的任意元素 x 映到 x 本身, 称为 X 上的**恒同变换**, 记作 id_X , 即 $\text{id}_X(x) = x$, 对任意 $x \in X$ 都成立. 特别地, 恒同变换与映射的复合不改变映射: 对任意映射 $f: X \rightarrow Y$, 都有 $f \circ \text{id}_X = f = \text{id}_Y \circ f$.

定义 0.3.1 设 $f: X \rightarrow Y$ 是一个映射, 若存在另一个映射 $g: Y \rightarrow X$, 满足 $f \circ g = \text{id}_Y, g \circ f = \text{id}_X$, 则称映射 f **可逆**, 称 g 是 f 的一个**逆 (映射)**.

容易看出, 当 f 可逆时, 它的逆映射是唯一的, 称为 f 的逆, 记作 f^{-1} .

利用映射复合和恒同变换, 可以给出单射、满射和双射的等价描述.

命题 0.3.2 对映射 $f: X \rightarrow Y$, 若 $X \neq \emptyset$, 则有

1. f 是单射当且仅当存在另一个映射 $g: Y \rightarrow X$, 使得 $g \circ f = \text{id}_X$;
2. f 是满射当且仅当存在另一个映射 $g: Y \rightarrow X$, 使得 $f \circ g = \text{id}_Y$;
3. f 是双射当且仅当 f 可逆, 即存在另一个映射 $g: Y \rightarrow X$, 使得 $f \circ g = \text{id}_Y$, $g \circ f = \text{id}_X$.

命题的证明留作习题.

习题

练习 0.3.1 判断下列映射是否为单射、满射、双射, 并写出双射的逆映射:

1. $f: \mathbb{R} \rightarrow \mathbb{R}$,
 $x \mapsto x + 1$.
2. $f: \mathbb{R} \rightarrow \mathbb{R}$,
 $x \mapsto 2x$.
3. $f: \mathbb{R} \rightarrow \mathbb{R}$,
 $x \mapsto 3$.
4. $f: \mathbb{R} \rightarrow \mathbb{R}$,
 $x \mapsto x^2$.
5. $f: (-\infty, 0] \rightarrow \mathbb{R}$,
 $x \mapsto x^2$.
6. $f: \mathbb{R} \rightarrow (0, \infty)$,
 $x \mapsto e^x$.
7. $f: \left[-\frac{3\pi}{2}, -\frac{\pi}{2}\right] \rightarrow [-1, 1]$,
 $x \mapsto \sin x$.

练习 0.3.2 对复合映射 $f = g \circ h$, 证明或举出反例.

1. g, h 都是满射, 则 f 是满射.
2. g, h 都是单射, 则 f 是单射.
3. h 不是满射, 则 f 不是满射.
4. g 不是满射, 则 f 不是满射.
5. g 不是单射, 则 f 不是单射.
6. h 不是单射, 则 f 不是单射.
7. g, h 都不是双射, 则 f 不是双射.

练习 0.3.3

1. 在不改变对应法则和定义域的前提下, $\mathbb{R}$ 上的变换 $f(x) = x^2 + 2x + 3$, 把陪域换成哪个集合, 得到的映射是满射?
2. 证明映射 $f: [0, 1] \rightarrow \mathbb{R}, f(x) = x$ 是单射.

练习 0.3.4 下列 $\mathbb{R}$ 上的变换, 哪些满足交换律 $f \circ g = g \circ f$?

1. $f(x) = x + 1, g(x) = 2x$.
2. $f(x) = x^2, g(x) = x^3$.
3. $f(x) = 2^x, g(x) = 3^x$.
4. $f(x) = 2x + 1, g(x) = 3x + 2$.
5. $f(x) = 2x + 1, g(x) = 3x + 1$.
6. $f(x) = \sin x, g(x) = \cos x$.

练习 0.3.5 对 $\mathbb{R}$ 上的变换 $f(x) = 2x + 1$ 和 $g(x) = ax + b$, 求实数 a, b , 使得 $f \circ g = g \circ f$.

练习 0.3.6 在化简函数 $\arccos \circ \sin: \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow [0, \pi]$ 的下列步骤中, 分别利用了映射的哪些性质?

1. 如果 $y = (\arccos \circ \sin)(x)$, 则 $\cos y = (\cos \circ (\arccos \circ \sin))(x)$.
2. $\cos \circ (\arccos \circ \sin) = (\cos \circ \arccos) \circ \sin$.
3. $(\cos \circ \arccos) \circ \sin = \text{id}_{[0,1]} \circ \sin$.
4. $\text{id}_{[0,1]} \circ \sin = \sin$.

综上得 $\sin x = \cos y$, 由此推断化简结果.

练习 0.3.7 用数学归纳法证明, 任取有限个映射 $f_1, f_2, \dots, f_n$, 如果对 $1 \leq i \leq n-1$, f_{i+1} 的定义域都等于 f_i 的陪域, 则复合映射 $f_n \circ \dots \circ f_2 \circ f_1$ 不因映射复合的计算次序而改变.

练习 0.3.8 证明, 任意映射 f 都存在分解 $f = g \circ h$, 其中 h 是满射, g 是单射.

练习 0.3.9 给定映射 h, g 和 $f = g \circ h$, 证明, 若 f 是双射, 则 h 是单射, g 是满射.

练习 0.3.10 证明命题 0.3.2.

第 1 章 线性映射和矩阵

线性代数这门学科，以研究**线性系统**为目的，是一个从理论到应用都相当完善的数学分支。我们先从系统说起。系统是指由若干个互相作用和互相依赖的事物组合而成的具有特定功能的一个整体。比较简单的一种是集合之间的映射：

$$\text{输入} \xrightarrow{\text{系统}} \text{输出}.$$

读者在中学学习的函数是一种特殊的映射，即从实数集到实数集的映射。在本书中，我们用 $\mathbb{R}$ 来表示实数集，而 $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{C}$ 分别表示自然数集、整数集、有理数集和复集。

1.1 基本概念

我们从一个简单的系统开始。

例 1.1.1 (桥墩载荷) 如图 1.1.1 所示，有一座桥，桥长为 d ，假设桥两端各有一桥墩，不考虑桥面和桥墩的重力、厚度。在距左侧 S_1 桥墩 l_1 处放置一重物，重力为 F_1 。问两个桥墩各承担了多少载荷？

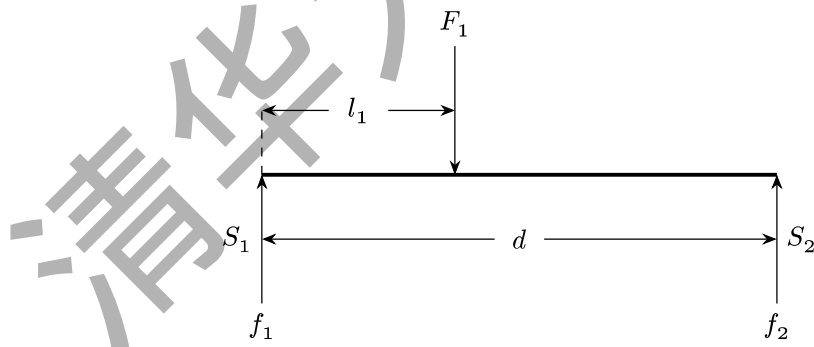


图 1.1.1 桥墩载荷：单重物

只需杠杆原理我们就可以求出答案。考虑以 S_1 为支点、桥面为杆臂的杠杆，显然 $F_1 l_1 = f_2 d$ ，因此 $f_2 = \frac{l_1}{d} F_1$ 。再考虑以 S_2 为支点的杠杆，有 $F_1 (d - l_1) = f_1 d$ ，因此 $f_1 = \frac{d - l_1}{d} F_1$ 。也可以利用力平衡， $f_1 = F_1 - f_2$ 求出 f_1 。

下面考虑两个重物的情形。如图 1.1.2 所示，在距左侧 S_1 桥墩 l_1 处放置一重物，重力为 F_1 ；在距左侧 S_1 桥墩 l_2 处放置一重物，重力为 F_2 。问两个桥墩各承担了多少载荷？

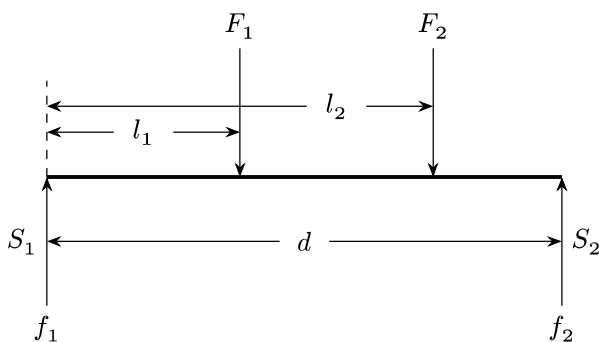


图 1.1.2 桥墩载荷：双重物

利用杠杆原理, 考虑以 S_1 为支点、桥面为杆臂的杠杆, 显然 $F_1 l_1 + F_2 l_2 = f_2 d$, 因此 $f_2 = \frac{l_1}{d} F_1 + \frac{l_2}{d} F_2$. 考虑以 S_2 为支点的杠杆, 有 $F_1(d - l_1) + F_2(d - l_2) = f_1 d$, 因此 $f_1 = \frac{d - l_1}{d} F_1 + \frac{d - l_2}{d} F_2$. 也可以利用力平衡, $f_1 = F_1 + F_2 - f_2$ 求出 f_1 .

固定重物的位置, 把重力 F_1, F_2 视为系统的输入, 两个桥墩的载荷 f_1, f_2 视为系统的输出. 系统中的输入和输出满足**叠加原理**, 即,

1. 输入放大或缩小某一倍数时, 输出也放大或缩小同一倍数;
2. 两组输入所产生的输出是二者分别产生的独立输出之和.

满足叠加原理的系统称为**线性系统**.

更一般地, 在任意多位置上放置任意重物, 我们都可以直接写出两个桥墩的载荷的表达式. ☺

上面描述的线性系统里, 输入和输出只有有限多个, 二者都可以表示成有序数组, 我们把有序数组称为**向量**.

定义 1.1.2 (向量) 一个 m 元有序数组 $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ 称为一个 m 维**向量**, 实数

$a_1, a_2, \dots, a_m$ 称为向量 $\mathbf{a}$ 的**分量**或**坐标**.

分量都是实数的 m 维向量的全体构成的集合记为 $\mathbb{R}^m$.

注 1.1.3 分量都是复数的 m 维向量的全体构成的集合记为 $\mathbb{C}^m$. 分量也可以取其他范围内的数, 本书主要讨论分量为实数和复数的两种情形.

两个向量**相等**是指二者的每个分量都相等. 向量常用黑体小写字母表示^①, 如 $\mathbf{a}$. 由于分量纵向排列, 向量又称为**列向量**. 根据需要, 向量也可以把分量横向排列, 称为**行向量**. 为

^① 手写时可以直接写字母, 也可以在字母上加箭头, 如 $\vec{a}$.

了与列向量区分, 用符号 $\mathbf{a}^T$ 表示行向量, 即如果 $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$, 则 $\mathbf{a}^T = [a_1 \ a_2 \ \cdots \ a_m]$.

现在, 例 1.1.1 中两个重物的线性系统可以表示成如下映射:

$$\begin{aligned} \mathbb{R}^2 &\rightarrow \mathbb{R}^2, \\ \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} &\mapsto \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \frac{d-l_1}{d}F_1 + \frac{d-l_2}{d}F_2 \\ \frac{l_1}{d}F_1 + \frac{l_2}{d}F_2 \end{bmatrix}. \end{aligned}$$

一般地, n 个输入 m 个输出的线性系统可以表示成如下映射:

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \mapsto \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}. \quad (1.1.1)$$

我们将在下节说明该表示的正确性.

线性系统中的输入和输出, 也就是向量, 需要满足叠加原理: 可加性要求任意两个向量之间可以相加, 而齐次性要求任意一个向量可以放缩某一倍数.

定义 1.1.4 (线性运算) 为 $\mathbb{R}^m$ 中的向量定义两种运算^①

1. 两个 m 维向量的向量加法: $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} := \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_m + b_m \end{bmatrix};$
2. 一个 m 维向量与一个实数的数乘: $k \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} := \begin{bmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_m \end{bmatrix};$

向量的加法和数乘统称向量的线性运算.

带有线性运算的集合 $\mathbb{R}^m$, 称为向量空间 $\mathbb{R}^m$ 或线性空间 $\mathbb{R}^m$.

注意, “线性空间 $\mathbb{R}^m$ ” 中的向量能做加法和数乘, 而 “集合 $\mathbb{R}^m$ ” 则不能, 二者并不是同一个数学对象. 另外, 线性空间 $\mathbb{R}^n$ 和线性空间 $\mathbb{R}^m$ 中的向量在 $m \neq n$ 时无法做加法; 线性空间 $\mathbb{R}^m$ 上的数乘运算中的数, 需要是实数.

^① 表达式中, $A := B$ 表示令 $A = B$ 或者把 B 记作 A ; 类似地, $A =: B$ 表示 $B := A$.

线性运算本身需要满足一定的运算法则.

命题 1.1.5 线性空间 $\mathbb{R}^m$ 中的向量加法和数乘满足如下八条运算法则:

1. 加法结合律: $(a + b) + c = a + (b + c)$;

2. 加法交换律: $a + b = b + a$;

3. 零向量: 存在 m 维零向量 $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, 满足 $a + \mathbf{0} = a$;

4. 负向量: 对任意 $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$, 记 $-a = \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_m \end{bmatrix}$, 则 $a + (-a) = \mathbf{0}$, 称 $-a$ 为

a 的负向量;

5. 单位数: $1a = a$;

6. 数乘结合律: $(kl)a = k(la)$;

7. 数乘关于数的分配律: $(k + l)a = ka + la$;

8. 数乘关于向量的分配律: $k(a + b) = ka + kb$.

其中, a, b, c 是 $\mathbb{R}^m$ 中的向量, k, l 是实数.

命题的验证是直接的, 我们留给读者. 下面讨论运算法则的一些结果和注意事项.

1. 根据负向量性质, 我们可以定义向量的**减法**: $a - b = a + (-b)$.

2. 根据加法结合律, 我们不需要区分加法运算的先后次序, 因此可以写 $a + b + c = (a + b) + c = a + (b + c)$. 类似地, $kla = (kl)a = k(la)$.

3. 上面的运算法则使得我们能够“合并同类项”. 例如

$$2a + 4(2a + (3b + \mathbf{0})) + 4((a + (2 + 3)c) - a) = 10a + 12b + 20c.$$

4. 有关数乘的其他结果: $0a = \mathbf{0}$, $(-1)a = -a$.

5. 特别注意, 向量之间**没有**直接的乘除法. 有时我们会写 $\frac{a}{k}$, 这时 k 必须是实数而非向量. 向量出现在分母上, 意味着已经犯了十分严重的概念错误.

例 1.1.6 读者在中学学习过平面向量. 把平面向量的起点固定在原点, 它的坐标表示就是二维数组; 而平面向量线性运算的坐标表示就是上面定义的两线性运算. 因此, 配备上线性运算, 所有以原点为起点的平面向量全体构成的集合就是线性空间 $\mathbb{R}^2$. $\odot$

有了线性空间中的线性运算, 线性系统 (1.1.1) 的叠加原理可以具体地描述如下:

$$\begin{aligned}
 f(\boldsymbol{x} + \boldsymbol{x}') &= \begin{bmatrix} a_{11}(x_1 + x'_1) + \cdots + a_{1n}(x_n + x'_n) \\ a_{21}(x_1 + x'_1) + \cdots + a_{2n}(x_n + x'_n) \\ \vdots \\ a_{m1}(x_1 + x'_1) + \cdots + a_{mn}(x_n + x'_n) \end{bmatrix} \\
 &= \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} + \begin{bmatrix} a_{11}x'_1 + \cdots + a_{1n}x'_n \\ a_{21}x'_1 + \cdots + a_{2n}x'_n \\ \vdots \\ a_{m1}x'_1 + \cdots + a_{mn}x'_n \end{bmatrix} = f(\boldsymbol{x}) + f(\boldsymbol{x}'), \\
 f(k\boldsymbol{x}) &= \begin{bmatrix} a_{11}kx_1 + \cdots + a_{1n}kx_n \\ a_{21}kx_1 + \cdots + a_{2n}kx_n \\ \vdots \\ a_{m1}kx_1 + \cdots + a_{mn}kx_n \end{bmatrix} = k \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = kf(\boldsymbol{x}),
 \end{aligned}$$

其中^①, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\boldsymbol{x}' = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} \in \mathbb{R}^n$, $k \in \mathbb{R}$. 由观察可知, 等式左边的 $\boldsymbol{x} + \boldsymbol{x}'$, $k\boldsymbol{x}$ 是 $\mathbb{R}^n$

中的线性运算, 而等式右边的 $f(\boldsymbol{x}) + f(\boldsymbol{x}')$, $kf(\boldsymbol{x})$ 是 $\mathbb{R}^m$ 中的线性运算.

定义 1.1.7 (线性映射) 映射 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, 如果满足

1. 任取 $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n$, 都有 $f(\boldsymbol{x} + \boldsymbol{x}') = f(\boldsymbol{x}) + f(\boldsymbol{x}')$;
2. 任取 $\boldsymbol{x} \in \mathbb{R}^n, k \in \mathbb{R}$, 都有 $f(k\boldsymbol{x}) = kf(\boldsymbol{x})$;

则称 f 为从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性映射.

因此, (1.1.1) 式中的映射是线性映射.

线性映射把定义域上的加法映射成陪域上的加法, 把定义域上的数乘映射成陪域上同一个数的数乘. 简单来说, 线性映射保持线性运算. 例如, 对任意线性映射 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, 都有 $f(k\boldsymbol{x} + k'\boldsymbol{x}' + k''\boldsymbol{x}'') = kf(\boldsymbol{x}) + k'f(\boldsymbol{x}') + k''f(\boldsymbol{x}'')$. 特别地, 线性映射把零向量映射到零向量: 在定义中取常数 $k = 0$, 可以得到 $f(\mathbf{0}_n) = \mathbf{0}_m$, 其中 $\mathbf{0}_n \in \mathbb{R}^n$ 是 $\mathbb{R}^n$ 中的零向量, 而 $\mathbf{0}_m \in \mathbb{R}^m$ 是 $\mathbb{R}^m$ 中的零向量. 利用这个性质, 可以对一些映射简单有效地做出其并非线性的判断.

例 1.1.8 线性映射在生活中随处可见.

^① 注意, 这里 $\boldsymbol{x}'$ 表示不同于 $\boldsymbol{x}$ 的另一个向量, 而不是对向量 $\boldsymbol{x}$ 做了某种运算的结果. 类似地, 后文出现的 $\boldsymbol{\alpha}, \boldsymbol{\alpha}$ 等都表示不同的向量.

1. 去超市买水果, 在苹果、香蕉、樱桃之间选择. 潜在的购买空间是由这三种水果

组成的一个线性空间 $\mathbb{R}^3$, 设苹果 $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, 香蕉 $\mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, 和樱桃 $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

而结账这个过程就是一个映射, 它将水果组合, 映射到对应的总价上. 如果苹果、香蕉和樱桃的每千克价格分别是 12.98 元、9.98 元和 129.98 元, 那么“结账”这个映射可以表达为:

$$\begin{aligned} \text{结账: } \mathbb{R}^3 &\rightarrow \mathbb{R}, \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\mapsto 12.98x_1 + 9.98x_2 + 129.98x_3. \end{aligned}$$

容易看出, 这是一个线性映射. 例如, $\mathbf{x}, \mathbf{x}'$ 分别代表两种水果组合, 对 $\mathbf{x} + 2\mathbf{x}'$ 结账时, 总价等于对 $\mathbf{x}$ 结账的总价加上对 $\mathbf{x}'$ 结账的总价的 2 倍. 即,

$$\text{结账}(\mathbf{x} + 2\mathbf{x}') = \text{结账}(\mathbf{x}) + 2\text{结账}(\mathbf{x}').$$

2. 假设每千克苹果含有 12 克纤维素和 135 克糖, 每千克香蕉含有 12 克纤维素和 208 克糖, 每千克樱桃含有 3 克纤维素和 99 克糖. 所有纤维素和糖的含量的组合组成一个二维线性空间 $\mathbb{R}^2$, 设纤维素 $\mathbf{f} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 糖 $\mathbf{s} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. 把每种水果组合映射到对应的纤维素和糖的含量, 是一个从 $\mathbb{R}^3$ 到 $\mathbb{R}^2$ 的映射, 称为“营养”:

$$\begin{aligned} \text{营养: } \mathbb{R}^3 &\rightarrow \mathbb{R}^2, \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\mapsto \begin{bmatrix} 12x_1 + 12x_2 + 3x_3 \\ 135x_1 + 208x_2 + 99x_3 \end{bmatrix}. \end{aligned}$$

容易看出, 这是一个线性映射.

注意, 这两个例子中的线性映射都具有形如 (1.1.1) 式的表达式. ☺

定义 1.1.9 (线性变换) 从 $\mathbb{R}^n$ 到自身的线性映射称为 $\mathbb{R}^n$ 上的**线性变换**.

特别地, $\mathbb{R}^n$ 上的**恒同变换**

$$\begin{aligned} \mathbf{I} = \text{id: } \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \mathbf{x} &\mapsto \mathbf{x} \end{aligned} \tag{1.1.2}$$

是线性变换.

例 1.1.10 考虑例 1.1.6 中由平面向量构成的线性空间 $\mathbb{R}^2$, 下面讨论平面^①上的几类线性变换.

1. **旋转变换**: 对任意实数 θ , 将所有向量绕原点逆时针旋转 θ 大小的角. 这定义了一个 $\mathbb{R}^2$ 上的变换, 记为 R_θ . 图 1.1.3 证明了 $R_\theta(kx) = kR_\theta(x)$, $R_\theta(x+x') = R_\theta(x) + R_\theta(x')$.

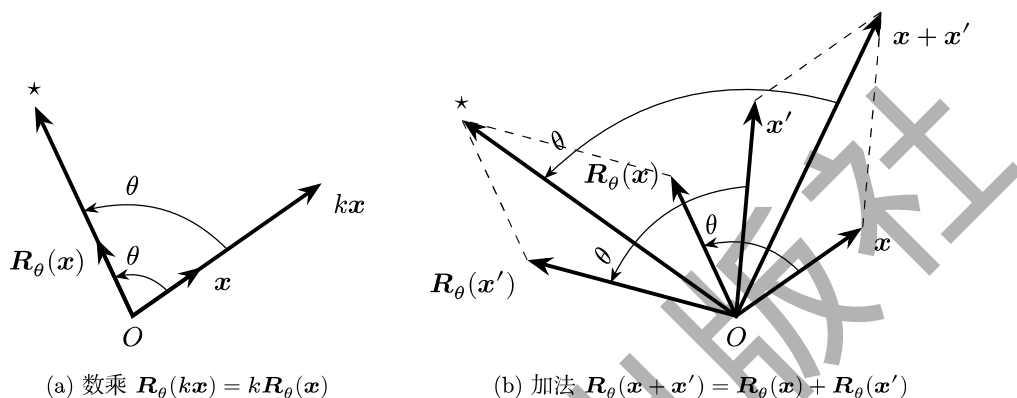


图 1.1.3 旋转变换

因此, 旋转变换 R_θ 是线性变换. 那么, 它是否具有形如 (1.1.1) 式的表达式呢? 由于 R_θ 是线性的, 我们有

$$R_\theta \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = R_\theta \left(x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = x_1 R_\theta \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + x_2 R_\theta \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

容易知道

$$R_\theta \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad R_\theta \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} \cos(\theta + \frac{\pi}{2}) \\ \sin(\theta + \frac{\pi}{2}) \end{bmatrix},$$

因此

$$R_\theta \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = x_1 \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} + x_2 \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} = \begin{bmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{bmatrix}.$$

注意, 线性映射 R_θ 仅仅由它在 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 这两个向量上的取值决定. 这是

线性映射最特殊的地方. 我们称 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 这两个向量为 $\mathbb{R}^2$ 的**标准坐标向量**.

^① 我们以后经常会直接把 $\mathbb{R}^2$ 称为平面, 但要注意多数情形中 $\mathbb{R}^2$ 中的元素是向量, 而不是点.