

线性代数

浅思录

作者：匡睿

时间：July 1, 2025

版本：1.0 正式版

邮箱：qjtykr65536@gmail.com



现实是此岸，理想是彼岸。中间隔着湍急的河流，行动则是架在川上的桥梁。——克雷洛夫

序言：从等待走向默契

这本书推荐使用电子介质阅读，或使用彩印书籍以获得最佳观感。

感谢你翻开这本书！

当年我的数学老师告诉我说，书籍可是作者的亲生儿子，你需要花费很多很多时间和精力去创作它，而不是脑子一热说我想写一本书。But，这本书确实是脑子一热诞生的一个项目，只不过开坑一张嘴，填坑跑断腿，不知不觉间这本书已经陪我走过半年时间。出这部书出自于偶然，并且承载了一份很长的等待，从立项开始，内心可以算是挣扎着是否要继续写下去；恍惚间感觉一直都在强加给这件事一个意义。

所以意义我认为算是没有等到，不过等到了一个机会，那就是理科迷的公益课邀请我与大家来聊聊线性代数的意义与价值。如果大家已经接触过线性代数，或许会有和我当年一样学习线性代数时的困惑：

1. 上来就是行列式，它们如何得来，为何这么计算？
2. 什么是矩阵，它们的相加，相乘都有什么意义？
3. 为什么要引入线性代数这门数学分支，它能够做什么？

经过我们本科的线性代数的学习，我们老师告诉我们线性代数（工科¹）的本质围绕着“解方程”²展开，行列式与矩阵不过是解决这类问题的工具，当我们尝试追问，行列式为什么这么计算的时候，数学老师告诉我们，这就像 $1 + 1 = 2$ 一样，是一个规定，如果你尝试严格证明，那牵扯的东西会很多。事实上的确如此，不过这本书打算以一个抽象的方式来描述“解方程”，实际上解方程也不过是一个从 $\mathbb{R}^n \rightarrow \mathbb{R}^m$ 的线性映射；那么除了解方程，线性代数也有很多其他意义，例如对空间中的直线与平面作解析几何，对多组数据的线性处理研究等。

在学习线性代数之前，笔者阅读了国内的部分教材，其中包括同济本（人民邮电出版社³），很经典地，上来就是行列式，缺少动机也缺少直观，不过好在我认为我们线性代数的老师水平不错，能够让大家在不知其所以然的时候知道行列式是什么东西，让我们放下思想包袱，不要去深究这里面的内容；不过数学学习主打一个“兴趣”，如果你深入研究你会发现数学其实还是很好玩的，在不同的领域中数学的应用，或许会让你产生闭环的感觉，发现原来这里也用到了线性代数的知识，就像一位故友一般。所以我很赞同我们大学数学老师的话，那就是学习重在一个过程而不是结果。有人曾言，若离开考试，语文就是一个绝美的化身，那么数学则是充满了理性之美的艺术。

所以写这本书的动机就是希望它是笔者亲自打造的艺术品，希望传达数学的艺术，不仅限于书本与考试的一片三亩方地，而是能够实现自我探索，发现更多知识的天空。当然我们目前学习也是为了考试，所以在后面笔者也会放上大部分课上的笔记与注意事项，具体的书目会书写大概 3 个部分：

1. 理论部分，这个部分会放在第一个讲，从数与集合开始，并产生线性代数的一个研究的容器，逐步生成线性空间与线性映射的概念。这一部分会有些难度，如果看不明白可以跳过，至于为什么放在第一个，是因为我认为这是线性代数的本质，能够帮助读者对行列式和矩阵有一个不同于“解方程”的认识。该部分包括线性空间，线性映射，内积空间。
2. 基础部分，讲解行列式，矩阵的计算，依照线性方程组的特性与理论部分辅助理解内容。同时会结合教师讲义讲解计算技巧，满足考试的需要，让这本书更贴合实际。
3. 附加部分，讲解多项式，度量空间等，给希望拓展线性代数学习的读者学习交流。

每个包含若干个章节，章节后面会有两组练习题，对于这两组练习题的难度，个人认为要求如下：

¹有区别于高等代数，只涉及计算技巧，不涉及严格证明

²如果这本书没有严格说明，那么都是描述解线性方程组

³同济大学数学系编著的共有 2 个出版社，其中最常用的是高等教育出版社，此外人民邮电出版社内容编排更合理一些，但是更为小众



-
- A 组是基础的题目，需要读者熟练掌握并了解其含义，不会涉及很难的证明题，是对该章节最基本的认知；几乎会以例题为标准举一反三，所以与该书的例题难度不相上下；
 - B 组是拓展题，对工科学生来说不应要求掌握，但是可以给想挑战自我的读者一个发挥的空间；
 - 所有题目的答案将会在本人的 GitHub 主页手写发布。

我希望这本书能够帮助所有初次接触线性代数的同学们一些参考，承载等待的时候，多了份默契，这份默契不需要过多的彩排，源于大家的支持，若能够帮助到大家的学习，我何其有幸！

最后感谢在前行路上给予我指导的老师与同学们，特别鸣谢：广西民族大学向子凤同学，南京大学余沛宸同学，江苏理工学院蒋盈峰教师，清华大学李思教授，以及江苏省江阴高级中学翟国华教师；特别致谢江苏理工学院 2023-2024 经济统计学班的全体学生和老师。

匡睿同学

July 1, 2025 于长广溪畔





目录

1

第 1 部分 * 理论部分

第 1 章 线性空间	2
1.1 标量与向量	2
1.1.1 数的运算	2
1.1.2 标量	3
1.1.3 元组	3
1.1.4 笛卡尔积	3
1.1.5 空间 \mathbb{F}^n	4
1.1.6 向量	4
1.2 向量的运算法则	5
1.2.1 加法	5
1.2.2 0 向量	6
1.2.3 加法逆元	6
1.2.4 标量乘法	6
1.3 线性空间 \mathbb{R}^2 与 \mathbb{R}^3	7
1.3.1 线性空间	7
1.3.2 直线, 平面与三维空间上的向量	8
1.3.3 其他线性空间	11
1.4 子空间	12
1.4.1 子空间 (subspace) 的定义	12
1.4.2 子空间之和	12
1.4.3 直和	13
1.5 张成空间与线性相关性	13
1.5.1 线性组合与张成空间	13
1.5.2 线性相关性	14
1.5.3 基与维数	16
1.6 章节练习	17
1.6.1 A 组	17



1.6.2 B 组	17
第 2 章 矩阵与线性映射	19
2.1 线性方程组与矩阵	19
2.1.1 向量的矩阵表示	19
2.1.2 线性方程组与平面上的线性变换	20
2.1.3 矩阵表示线性方程组	23
2.1.4 矩阵的定义	23
2.2 矩阵的运算法则	24
2.2.1 矩阵和	24
2.2.2 0 矩阵	24
2.2.3 加法逆元	25
2.2.4 标量乘法	25
2.2.5 矩阵乘法	25
2.2.6 平面上的线性变换	27
2.3 线性映射	28
2.3.1 从平面走向更高维度	28
2.3.2 线性空间中的线性映射	28
2.3.3 零空间	29
2.3.4 值域空间	30
2.3.5 维数定理	30
2.4 矩阵表示线性映射	31
2.4.1 线性映射的运算法则	31
2.4.2 矩阵描述线性映射	31
2.5 章节练习	33
2.5.1 A 组	33
2.5.2 B 组	33
第 3 章 内积空间	35
3.1 内积	35
3.1.1 向量的点积与长度表示	35
3.1.2 内积的定义	36
3.1.3 范数	37
3.1.4 内积空间	38
3.1.5 Cauchy-Schwarz 不等式	38
3.1.6 余弦定理	39
3.1.7 三角不等式	40
3.2 正交	41
3.2.1 正交的定义	41
3.2.2 正交基	42
3.2.3 Gram-Schmidt 正交化	42
3.2.4 正交投影	44
3.3 极小化问题	45
3.3.1 到子空间的最短距离	45
3.3.2 多项式与最小化	45
3.4 章节练习	47
3.4.1 A 组	47





3.4.2 B 组	48
---------------------	----

2

第 2 部分 * 基础部分

第 4 章 行列式	50
4.1 面积与体积	50
4.1.1 面积	50
4.1.2 体积	52
4.1.3 2 或 3 阶行列式的性质	52
4.2 n 阶行列式	53
4.2.1 n 阶行列式的定义	53
4.2.2 2 阶与 3 阶行列式的运算	53
4.3 行列式按行展开	53
4.3.1 代数余子式	53
4.3.2 行列式按行展开	54
4.4 行列式的性质	56
4.4.1 转置	56
4.4.2 其他性质	56
4.4.3 初等变换计算行列式	58
4.5 置换	60
4.5.1 全排列与置换	60
4.5.2 有向的 n 维测度	60
4.5.3 最少置换次数与 $\text{sign}(\sigma)$	60
4.5.4 行列式的排列定义	61
4.6 章节练习	62
4.6.1 A 组	62
4.6.2 B 组	62
第 5 章 初等矩阵	63
5.1 线性方程组的解	63
5.1.1 二元线性方程组	63
5.1.2 Cramer 法则	63
5.1.3 齐次线性方程组	65
5.2 初等矩阵	65
5.2.1 增广矩阵	65
5.2.2 矩阵的初等变换	66
5.2.2.1 交换两行	66
5.2.2.2 某一行乘以常数 k	67
5.2.2.3 某个数乘以某一行并加到另一行中去	67
5.2.2.4 最简矩阵	67
5.2.3 高斯消元法	68
5.3 矩阵的逆	69
5.3.1 单位矩阵	69
5.3.2 逆矩阵	70
5.3.3 初等变换求逆矩阵	70





5.4 矩阵的秩	72
5.4.1 极大线性无关组	72
5.4.2 向量组与矩阵的秩	73
5.4.3 秩与维度	75
5.5 线性方程组的解空间	75
5.5.1 齐次线性方程组解的结构	75
5.5.2 解空间	76
5.5.3 非齐次线性方程组的导出组与特解	78
5.6 章节练习	80
5.6.1 A 组	80
5.6.2 B 组	81
第 6 章 特征值理论	82
6.1 特特征值与特征向量	82
6.1.1 特特征值	82
6.1.2 特特征向量	84
6.2 迹与上三角矩阵	86
6.2.1 矩阵的迹	86
6.2.2 上三角矩阵	87
6.3 相似变换	89
6.3.1 相似矩阵	89
6.3.2 Hamilton-Cayley 定理	90
6.3.3 对角化矩阵	91
6.4 Jordan 标准型	94
6.4.1 Jordan 块	94
6.4.2 代数重数	94
6.4.3 几何重数与不变子空间	95
6.4.4 Jordan 标准型	96
6.4.5 Jordan 标准型的幂与幂零算子	97
6.4.6 将矩阵相似化为 Jordan 标准型	99
6.5 章节练习	103
6.5.1 A 组	103



第一部分 理论部分

第1部分目录

第1章 线性空间	2
1.1 标量与向量	2
1.2 向量的运算法则	5
1.3 线性空间 \mathbb{R}^n 与 \mathbb{R}^3	7
1.4 子空间	12
1.5 张成空间与线性相关性	13
1.6 章节练习	17
第2章 矩阵与线性映射	19
2.1 线性方程组与矩阵	19
2.2 矩阵的运算法则	24
2.3 线性映射	28
2.4 矩阵表示线性映射	31
2.5 章节练习	33
第3章 内积空间	35
3.1 内积	35
3.2 正交	41
3.3 极小化问题	45
3.4 章节练习	47



第 1 章 线性空间

「越想贴近事实，不明白的事情就越多」

——《宝石之国》

1.1 标量与向量

1.1.1 数的运算

从小学一年级开始我们就接触了整数，并且学会进行加减乘除；长大一些后，我们了解到了分数，将数系扩充到了有理数部分；再后来初中阶段我们遇到了平方根这种运算，接触到了无理数，我们把这些数都统称为实数，后来我们又了解了 -1 也能被开平方根，开创了复数领域；同时在高中阶段，我们也学习了简单的集合论，通常我们会用几个记号来表示这些数域。

定义 1.1 (数集的符号)

这些数集有着相对应的符号：

- \mathbb{N} 表示所有自然数的集合，其元素是所有非负整数。
- \mathbb{Z} 表示所有整数的集合。
- \mathbb{Q} 表示所有有理数的集合，其中有理数被定义为所有能够表示为 $\frac{m}{n}, m \in \mathbb{Z}, n \in \mathbb{Z}, n \neq 0$ 的数。
- \mathbb{R} 表示所有实数的集合。
- \mathbb{C} 表示所有复数的集合。



由于这门课不是实分析或高等代数，所以不会详细讲解这些数集怎么构造，读者只需要对此有一个简单的认识就可以了。众所周知，数的运算分别有加减乘除四种，如果读者觉得自己小学数学不错的话，一定知道下面的定理。

公理 1.1 (域公理)

对于数集 \mathbb{C} 的基本算数性质有：

1. 交换律 (communitativity): 对于所有的 $\alpha, \beta \in \mathbb{C}$ 有 $\alpha + \beta = \beta + \alpha, \alpha\beta = \beta\alpha$;
2. 结合律 (associativity): 对于所有的 $\alpha, \beta, \gamma \in \mathbb{C}$ 有 $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma), (\alpha\beta)\gamma = \alpha(\beta\gamma)$;
3. 单位元 (identities): 对于所有的 $\lambda \in \mathbb{C}$ 都有 $\lambda + 0 = \lambda, 1\lambda = \lambda$;



4. 加法逆元 (additive inverse): 对于所有的 $\alpha \in \mathbb{C}$ 都存在唯一的 $\beta \in \mathbb{C}$ 使得 $\alpha + \beta = 0$;
5. 乘法逆元 (multiplicative inverse): 对于每个 $\alpha \in \mathbb{C}$ 且 $\alpha \neq 0$, 都存在唯一的 $\beta \in \mathbb{C}$ 使得 $\alpha\beta = 1$;
6. 分配律 (distributive): 对于所有的 $\alpha, \beta, \gamma \in \mathbb{C}$ 都有对于所有的 $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.



如果一个数域 \mathbb{F} 和 \mathbb{C} 一样满足上述性质, 我们称 \mathbb{F} 中的元素为标量 (scalar), 当然 \mathbb{C} 和 \mathbb{R} 的所有元素也是标量。

1.1.2 标量

如上所述, 标量通常强调的是, 这一个元素是一个数, 而不是向量 (稍后给出定义), 实际上只是“数”的一种花里胡哨的叫法而已。德国数学家 Frobenius 已经证明, 当下数学体系内再也没有比 \mathbb{C} 更大的数系了, 所以我们可以称对于每一个 $\alpha \in \mathbb{C}$, α 是标量, 例如 $3 + 4i$ 是一个标量。

例题 1.1.

前文所述, 我们说过复数是最大的数系, 请求出 \sqrt{i} 的两个解, 其中 $i = \sqrt{-1}$ 。

解: 设 $\sqrt{i} = z = (a + bi)$, 其中 $a, b \in \mathbb{R}$, 通过两边将其平方可得

$$i = (a + bi)^2 = a^2 - b^2 + 2abi$$

观察式子联立方程组为

$$\begin{cases} a^2 - b^2 = 0 \\ 2ab = 1 \end{cases}$$

解此方程组可得

$$\begin{cases} a = b \\ a = \pm \frac{\sqrt{2}}{2} \end{cases}$$

所以 \sqrt{i} 的两个解为

$$z_1 = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, z_2 = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$$

1.1.3 元组

若 $n \in \mathbb{N}$, 我们将由 n 个数按顺序排列成为一个组称之为元组 (tuple), 通常的记法为: 有小括号包含这 n 个数, 且有序排列, 中间由逗号隔开, 例如 (a, b) 是一个长度为 2 的元组, 如果长度为 n 且由 $z_1, z_2, z_3, \dots, z_n$ 这些数组成的元组的记法可能如下所示: $(z_1, z_2, z_3, \dots, z_n)$ 。

思考

集合是元组吗?

对于这个思考题, 很显然答案是不对的, 因为元组强调以下两点:

- 有序性, $(3, 5)$ 和 $(5, 3)$ 并不相等, 而 $\{3, 5\}$ 和 $\{5, 3\}$ 是相等的集合;
- 可重复, $(3, 3)$ 和 $(3, 3, 3)$ 并不相等, 而 $\{3, 3\}$ 和 $\{3, 3, 3\}$ 是相等的集合, 它们等同于 $\{3\}$ 。

1.1.4 笛卡尔积

接下来我们慢慢引入向量的概念, 在讲向量之前, 笔者想和大家谈一谈如何生成这样的空间来容纳我们向量的概念, 这里有一个定义叫做笛卡尔积 (Cartesian product)。



定义 1.2 (笛卡尔积)

两个集合 X 和 Y 的笛卡尔积是所有可能的有序对组成的集合，其中有序对的第一个对象是 X 的成员，第二个对象是 Y 的成员，记作：

$$X \times Y := ^a \{(x, y) \mid x \in X \text{ 且 } y \in Y\}$$

^a 符号 := 表示定义为，用来明确地给某个符号或概念赋予一个新的定义，例如：妈妈 := 养育与教养子女成长的女性



举个很常见的例子，如果集合 X 是 13 个元素的点数集合 $\{A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2\}$ 而集合 Y 是四种花色 $\{\spadesuit, \heartsuit, \diamondsuit, \clubsuit\}$ ，则这两个集合的笛卡尔积是有 52 个元素的标准扑克牌的集合：

$$X \times Y = \{(A, \spadesuit), (K, \spadesuit), (Q, \spadesuit), (J, \spadesuit), (10, \spadesuit), (9, \spadesuit), (8, \spadesuit), (7, \spadesuit), (6, \spadesuit), (5, \spadesuit), (4, \spadesuit), (3, \spadesuit), (2, \spadesuit), (A, \heartsuit), (K, \heartsuit), (Q, \heartsuit), (J, \heartsuit), (10, \heartsuit), (9, \heartsuit), (8, \heartsuit), (7, \heartsuit), (6, \heartsuit), (5, \heartsuit), (4, \heartsuit), (3, \heartsuit), (2, \heartsuit), (A, \diamondsuit), (K, \diamondsuit), (Q, \diamondsuit), (J, \diamondsuit), (10, \diamondsuit), (9, \diamondsuit), (8, \diamondsuit), (7, \diamondsuit), (6, \diamondsuit), (5, \diamondsuit), (4, \diamondsuit), (3, \diamondsuit), (2, \diamondsuit), (A, \clubsuit), (K, \clubsuit), (Q, \clubsuit), (J, \clubsuit), (10, \clubsuit), (9, \clubsuit), (8, \clubsuit), (7, \clubsuit), (6, \clubsuit), (5, \clubsuit), (4, \clubsuit), (3, \clubsuit), (2, \clubsuit)\}$$

1.1.5 空间 \mathbb{F}^n

我们定义 $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ ，根据笛卡尔积的运算规则，集合 \mathbb{R}^2 为所有有序实数 2 元组构成，即

$$\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$$

在现实生活中，我们可以将它看成是一个二维平面，而 \mathbb{R}^3 在现实生活中，我们可以将它看成是一个三维空间。

我们上述提到集合 \mathbb{F} 内的元素是一个标量，那么 \mathbb{F}^n 则由无数个的 n 元组构成，例如 \mathbb{C}^6 就是含有 6 个有序元素构成的一个空间，虽然当 $n > 3$ 的时候在实际生活中想象它们绝非易事，但是数学就是一门由具体到抽象的科目，我们可以使用类比的手法，让它们和在 \mathbb{R}^2 与 \mathbb{R}^3 上运算一样有自己的实际意义。

1.1.6 向量

如之前所述，我们终于提到了向量 (vector) 这个概念，只不过可能会令读者失望的是，数学科目上的向量是一个非常抽象的东西，那么我们把目光转换到另外一门和数学紧密相关的科目上：物理。

学基础物理的学生总是很实际，他们将数学转化为解决现实生活中问题的工具，所以在他们眼中，向量一般是一个箭头，它具有方向和大小，且这两个特征不会随着向量的移动发生变化，因此该形式的向量可以在空间中的任何一个位置。

所以我们在此非严格定义一个 n 元组被看作箭头时，就是一个 n 维向量 (稍后我们会给出它的更准确的定义)；下面是一个二维向量在平面 \mathbb{R}^2 内



(a) 一个指向点 (2, 2) 的向量

(b) 这两个向量是同一个，因为它们的大小和方向都一样

图 1.1: 二维向量



1.2 向量的运算法则

1.2.1 加法

定义 1.3 (向量的加法)

若 $x, y \in \mathbb{F}^n, x = (x_1, x_2, x_3, \dots, x_n), y = (y_1, y_2, y_3, \dots, y_n)$ 则 $x + y$ 定义为

$$x + y := (x_1 + y_1, x_2 + y_2, x_3 + y_3, \dots, x_n + y_n)$$



实际上两个向量相加，就是将它们元组里面的所有元素依次相加得到一个新的元组。

思考

维度不同的向量是否可以相加？

要解答这个问题，我们需要返回到定义中，我们只定义了维度相同都为 n 的向量相加，并没有定义维度不同的向量相加的情况，所以这是一个未定义的行为；所以下面的对向量的运算均是在同一维度下进行。同理，向量与标量无法相加。

公设 1.1

只有维度相同的向量才能互相相加。



同样，向量的加法也满足交换律：

推论 1.1

若 $x, y \in \mathbb{F}^n$ ，则 $x + y = y + x$ 。



下面是简单的对此性质的证明（读者不作要求）。

证明 若 $x, y \in \mathbb{F}^n, x = (x_1, x_2, x_3, \dots, x_n), y = (y_1, y_2, y_3, \dots, y_n)$ 则

$$x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \dots, x_n + y_n)$$

根据标量加法的交换律可知

$$x + y = (y_1 + x_1, y_2 + x_2, y_3 + x_3, \dots, y_n + x_n) = y + x \quad \square$$

如果我们将其反映在二维坐标系上，那应该是这样的

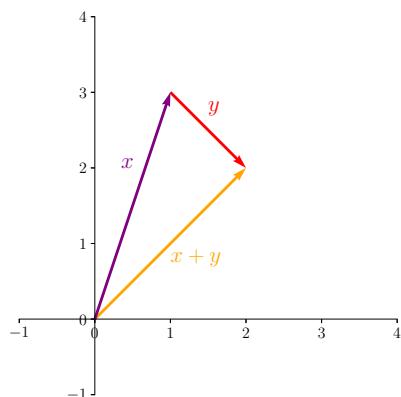


图 1.2：平面内两向量之和

1.2.2 0 向量

我们之前说过，向量与标量无法相加，所以在 \mathbb{F}^n 内我们定义 n 维 0 向量作为加法的单位元，即若 $x, y \in \mathbb{F}^n$ ， $x + y = x$ ，那么 y 就是 0 向量。

定义 1.4 (0 向量)

若 $\alpha, \beta \in \mathbb{F}^n$ 且 $\alpha + \beta = \alpha$ ，那么 β 为 \mathbb{F}^n 的单位元，我们将其称为 0 向量，显然 0 向量就是一个全是由 0 组成的 n 元组，记作 $\mathbf{0}$ ^a，那么

$$\mathbf{0} = (0, 0, 0, \dots, 0)$$

^a实际上是以粗体形式出现的 0



1.2.3 加法逆元

类似于标量加法的 $x, y \in \mathbb{C}, x + y = 0$ ，其中标量 0 表示标量加法的单位元，其中 $y = -x$ ，那么在空间 \mathbb{F}^n 中同样表示若 $\alpha, \beta \in \mathbb{F}^n, \alpha + \beta = \mathbf{0}$ ， $\beta = -\alpha$ 表示 α 的加法逆元。

定义 1.5 (加法逆元)

对于 $\alpha \in \mathbb{F}^n$ ， α 的加法逆元表示满足下面条件的向量 $-\alpha \in \mathbb{F}^n$ 有

$$\alpha + (-\alpha) = \mathbf{0}$$

换而言之当 $\alpha = (a, b, c, d, \dots)$ 时， $-\alpha = (-a, -b, -c, -d, \dots)$



1.2.4 标量乘法

定义 1.6 (向量的标量乘法)

若标量 $\lambda \in \mathbb{F}$ ，其对一个向量 $x = (x_1, x_2, x_3, \dots, x_n)$ 的乘积满足以下的运算法则

$$\lambda(x_1, x_2, x_3, \dots, x_n) := (\lambda x_1, \lambda x_2, \lambda x_3, \dots, \lambda x_n)$$



实际上如图1.3所示，向量的标量乘法从 \mathbb{R}^2 的几何意义来看表示的是向量长度的缩放倍率。例如 $x = (1, 2)$ 那么 $2x = (2, 4)$ 。

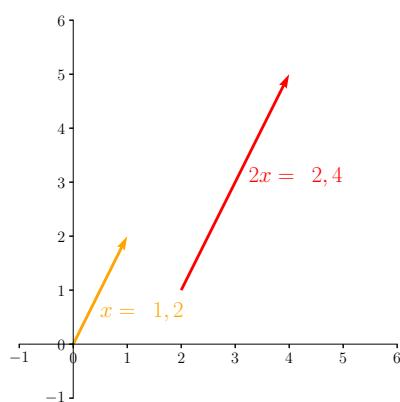


图 1.3: 向量长度的缩放倍率

同时，向量的标量乘法满足标量结合律。

推论 1.2

若 $a, b \in \mathbb{F}, x \in \mathbb{F}^n$, 满足 $(ab)x = a(bx)$ 。



至此, 我们对向量的运算法则做一个总结, 我们目前只定义了 2 种运算法则, 一种是向量和, 另一种是标量积, 下面我们会用这两种运算讲解一个特殊的集合, 使得这个集合由这两部分构成。

例题 1.2.

已知 $x \in \mathbb{R}^4$ 且满足 $(1, 1, 4, -5) + 2x = (5, 1, -4, 1)$, 求 x 。

解: 等式两边同时加上 $(1, 1, 4, -5)$ 的逆元 $(-1, -1, -4, 5)$ 可得

$$2x + (0, 0, 0, 0) = (5, 1, -4, 1) + (-1, -1, -4, 5) = (4, 0, -8, 6)$$

设 $x = (a, b, c, d)$ 可得

$$(2a, 2b, 2c, 2d) = (4, 0, -8, 6)$$

由此可以得到方程组

$$\begin{cases} 2a = 4 \\ 2b = 0 \\ 2c = -8 \\ 2d = 6 \end{cases}$$

所以可得 $x = (2, 0, -4, 3)$

读者平时在作业的时候无需大费周章地严格用定理来写题目, 如果你熟练的话可以直接自己写一个关于向量的减法法则 (实际上就是加法法则的逆运算) 来做题。

1.3 线性空间 \mathbb{R}^2 与 \mathbb{R}^3

1.3.1 线性空间

如果要刻画线性空间, 我们先从空间说起。

前文我们提到了“空间 \mathbb{F}^n ”, 那么我们可以在此断章取义地说, 空间就是一个集合, 而线性空间是一个特殊的集合, 我们给它一个符号 V 来表示。

下面给出定义来描述线性空间特殊的地方, 线性空间 V 是一个带有加法和标量乘法的集合, 下面我们将前面的向量的运算法则抽象化来定义它们。不过在此之前, 我希望和大家来聊一聊一些你可能不是很了解的名词, 例如函数 (function)。

定义 1.7 (函数 (一元))

对于任意的非空集合 A, B , 对集合 A 施加一个法则 f 使得 A 中的元素 x 可以转变为集合 B 中的元素 y , 记作 $f(x)$ 或 $f : A \rightarrow B$, 即 $y = f(x)$ 。



上面讲述了函数的一般情形, 那就是一元函数; 一元函数有一个显著的特点, 就是“输入一个元素并输出一个元素”, 当然还有多元函数, 下面请读者思考。

思考

两个可以相加的元素进行加法运算, 这是一种函数吗?



回答是肯定的，只不过这就不是一元函数了，这是二元函数，即“输入两个元素并输出一个元素”，翻译成数学的符号语言可以是：若 $a, b \in A$ 则 $f(a, b) := a + b$ 。同理，如果我们将集合内的元素特殊化为一种向量，那就是若 $a, b \in \mathbb{F}^n$ 则 $f(a, b) := a + b$ 。

那我们回到线性空间上来，如果一个集合 V 满足下面这一些条件的我们称其为线性空间¹。

定义 1.8 (线性空间上的加法与标量乘法)

若集合 V 与数集 \mathbb{F} 满足以下的性质，则称 V 为 \mathbb{F} 上的线性空间：

1. V 上的加法是一个函数使得任意一对 $u, v \in V$ 经过法则 $f(u, v) = u + v$ 转变为 V 中的另一个元素，这说明了 $u + v \in V$ 。
2. V 上的标量乘法是一个函数使得任意一个 $\alpha \in \mathbb{F}$ 和 $u \in V$ 经过法则 $f(u) = \alpha u$ 转变为 V 中的另一个元素，这说明了 $\alpha u \in V$ 。



根据的定义，线性空间 V 具有如下的性质：

推论 1.3 (线性空间的性质)

线性空间就是带有加法和标量乘法的集合 V ，满足如下的性质：

- 交换性 (commutativity): 对于所有的 $u, v \in V$ 都有 $u + v = v + u$;
- 结合性 (associativity): 对于所有的 $u, v, w \in V$ 和 $a, b \in \mathbb{F}$ 都有 $(u + v) + w = u + (v + w)$ 和 $(ab)v = a(bv)$;
- 加法单位元 (additive identity): 存在 $\mathbf{0} \in V$ 使得对于所有的 $v \in V$ 都有 $v + \mathbf{0} = v$;
- 加法逆元 (additive inverse): 对于每一个 $v \in V$ 都存在 $w \in V$ 使得 $v + w = \mathbf{0}$;
- 乘法单位元 (multiplicative identity): 对于所有的 $v \in V$ 都存在 $w \in V$ 使得 $v + w = v$;
- 分配性质 (distributive properties): 对于所有的 $a, b \in \mathbb{F}$ 和 $u, v \in V$ 都有 $a(u + v) = au + av$ 和 $(a + b)v = av + bv$ 。



思考

空集可以是线性空间吗？

答案为不是，因为空集 A 它不存在单位元 $\mathbf{0} \in A$ 所以并不能是线性空间，此外，需要注意的是我们一般会说 V 是 \mathbb{F} 上的线性空间而不会直接说 V 是线性空间，因为我们在定义线性空间的时候，依赖了 \mathbb{F} 这个标量数集。

如之前所承诺的，下面给出向量的定义。

定义 1.9 (向量)

线性空间中的元素称为向量。



读者可能会觉得这样给出会有些突然，我们不妨反过来向看看向量作为线性空间 V 的元素，是否满足定义1.3的6个条件，然后与读者高中所学的向量的特性对比，发现是否吻合。如果读者觉得这样做还是太抽象，我们不妨研究一下在线性空间 \mathbb{R}^2 与 \mathbb{R}^3 中向量的相关图形表现。

1.3.2 直线、平面与三维空间上的向量

向量同时具有长度和方向两种特性，如果我们在 \mathbb{R}^2 上绘制平面直角坐标系来图像化向量这个概念的话，向量就是一个有向箭头，由起点和终点构成，若一个有向箭头的起点为 P ，终点为 Q ，我们将此有向箭头²记为

¹有些教材会叫做向量空间

²向量与有向箭头的区别在于，长度相同且方向相同的有向箭头，可以有很多个表示，如图1.1b所示，而这两个有向箭头表示的向量是同一个。不过，通常而言，该有向线段也可直接称为向量

\overrightarrow{PQ} 。若 $a \in \mathbb{R}^2$, $a = \overrightarrow{PQ}$, 则向量 a 的长度为线段 PQ 的长度, 记作 $\|a\|$ 。

定义 1.10 (\mathbb{R}^2 上向量的长度)

若 $x \in \mathbb{R}^2$ 且 $x = (x_1, x_2)$, 则 x 的长度 $\|x\|$ 为

$$\|x\| := \sqrt{x_1^2 + x_2^2}$$



例题 1.3.

已知 $a, b \in \mathbb{R}^2$, 点 O 为平面直角坐标系的原点, P_1, P_2 分别为一条线段的两个端点, M 为该线段的重点, 其中 $a = \overrightarrow{OP_1}, b = \overrightarrow{OP_2}$, 求 \overrightarrow{OM} 。

解: 在长度上 $\|\overrightarrow{P_1M}\| = \|\overrightarrow{MP_2}\|$ 且在同一条直线上, 所以 $\overrightarrow{P_1M} = \overrightarrow{MP_2}$, 根据向量的加法法则有

$$\overrightarrow{OM} = \overrightarrow{OP_1} + \overrightarrow{P_1M}, \overrightarrow{OM} = \overrightarrow{OP_2} - \overrightarrow{P_1M}$$

前后两式相加可得

$$\overrightarrow{OM} = \frac{a+b}{2}$$



此外由于向量具有方向性, 且可以随处移动, 由此我们也可以由此性质确定 \mathbb{R}^2 和 \mathbb{R}^3 空间中直线的方向并确定直线的方程。

定义 1.11 (直线的方向向量)

空间直线的方向用一个与该直线平行的有向箭头表示, 该有向箭头所表示的向量即为空间直线的方向向量。

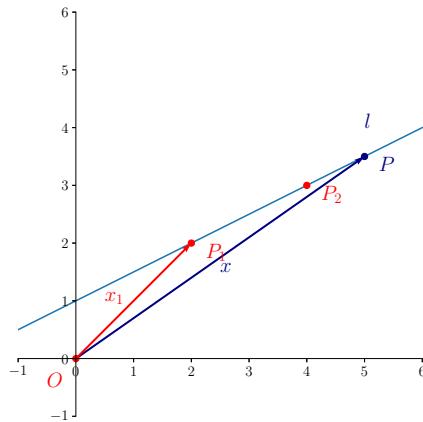


图 1.4: 方向向量与直线方程

如图1.4所示, P_1 和 P_2 都在直线 l 上, 设 $x, x_1 \in \mathbb{R}^2, x = \overrightarrow{OP}, x_1 = \overrightarrow{OP_1}$, 有向箭头 $\overrightarrow{P_1P_2}$ 在直线 l 上, 若向量 $a \in \mathbb{R}^2$ 满足 $a = \overrightarrow{P_1P_2}$, 则该向量 a 是这条直线的方向向量。

既然知道了直线的方向向量和 O 到直线上某一点的有向箭头所表示的向量, 我们就可以建立一个等式来唯一确定一条直线, 设 $t \in \mathbb{R}$, 那么对于任意处于直线上的 P 点, 总存在 t 使得 $x = \overrightarrow{OP}$ 满足

$$x = x_1 + ta$$

如果我们设直线上的 P 点坐标为 (x, y) 那么 $x = (x, y)$ 设 P_1 的坐标是 (x_1, y_1) , P_2 的坐标是 (x_2, y_2) , 于是乎



$a = (x_2 - x_1, y_2 - y_1)$, 由 $x = x_1 + ta$ 并代入坐标, 根据向量的加法与标量乘积的运算定理, 我们可以得到下面的式子

$$\begin{cases} x - x_1 = t(x_2 - x_1) \\ y - y_1 = t(y_2 - y_1) \end{cases}$$

所以我们可以得到这就是经过 P_1 和 P_2 两点的参数方程, 如果要进一步往下化简的话, 我们可以得到

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}$$

例题 1.4.

求二维空间 \mathbb{R}^2 内直线方程 $3x + 4y - 5 = 0$ 的一个方向向量。

解: 设 P_1, P_2 点坐标分别为 $(x_1, y_1), (x_2, y_2)$, 则 $a \in \mathbb{R}^2, a = \overrightarrow{P_1 P_2}$ 为其一个方向向量, 令 $x_1 = 0$ 则 $y_1 = \frac{5}{4}$, $\overrightarrow{OP_1} = \left(0, \frac{5}{4}\right)$, 令 $y_2 = 0$ 则 $x_2 = \frac{5}{3}$, $\overrightarrow{OP_2} = \left(\frac{5}{3}, 0\right)$, $\overrightarrow{P_1 P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1}$ 所以该直线的一个方向向量为 $\left(\frac{5}{3}, -\frac{5}{4}\right)$

接下来我们将维度扩展到三维空间 \mathbb{R}^3 内, 依然有

$$x = x_1 + ta$$

由此如果 P_1 与 P_2 的坐标分别为 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) , 我们将会得到下面的等式

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

这就是空间内的直线方程 (实际上就是两个一次方程)

例题 1.5.

求三维空间 \mathbb{R}^3 内直线方程

$$\begin{cases} x + y + z = 3 \\ x + 2y + 3z = 6 \end{cases}$$

的一个方向向量。

解: 设 P_1, P_2 点坐标分别为 $(x_1, y_1, z_1), (x_2, y_2, z_2)$, 则 $a \in \mathbb{R}^3, a = \overrightarrow{P_1 P_2}$ 为其一个方向向量, 令 $x_1 = 1$ 则 $y_1 = 1, z_1 = 1$, $\overrightarrow{OP_1} = (1, 1, 1)$, 令 $x_2 = 0$ 则 $y_2 = 3, z_2 = 0$, $\overrightarrow{OP_2} = (0, 3, 0)$, $\overrightarrow{P_1 P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1}$ 所以该直线的一个方向向量为 $(-1, 2, -1)$

同样的, 根据两条相交直线可以确定平面, 如果我们使用向量来考虑 \mathbb{R}^3 上的平面的话, 通过两个不互相平行的非零向量和一点也能确定一个平面, 如图1.5所示。

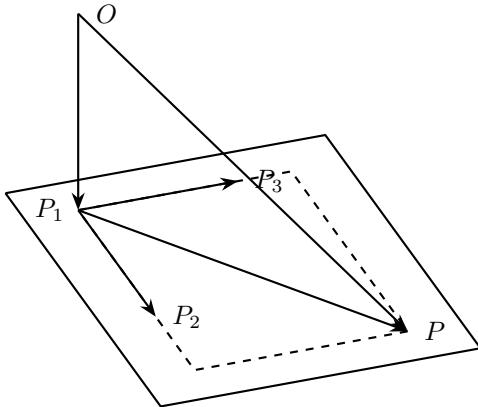
类比于直线的向量表示, 设向量 $x, x_1, a, b \in \mathbb{R}^3$, P, P_1, P_2, P_3 点在平面上, 令 $x = \overrightarrow{OP}, a = \overrightarrow{P_1 P_2}, b = \overrightarrow{P_1 P_3}, x_1 = \overrightarrow{OP_1}$, 设 $t, s \in \mathbb{R}$, 对于任意处于平面上的 P 点, 总存在 t, s 使得 $x = \overrightarrow{OP}$ 满足

$$x = x_1 + ta + sb$$

由于过空间内不在同一直线上的三个点写方程很复杂, 限于篇幅结构笔者不再详细介绍推导过程³。不过笔者在此给出平面方程的一般形式 (只有一个等式)。

$$ax + by + cz + d = 0, a, b, c, d \in \mathbb{R}$$

³可能会有读者会说平面的点法式方程, 我知道学霸读者很多, 但是笔者需要考虑给入门的选手讲清楚这些, 后面我们会慢慢接触它们, 请不要着急。

图 1.5: P 在平面上

1.3.3 其他线性空间

上面我们讲了 \mathbb{R}^2 和 \mathbb{R}^3 是一种线性空间，更一般的例子就是 \mathbb{R}^n ，再一般的例子就是 \mathbb{C}^n 。

定义 1.12 (实线性空间与复线性空间)

- \mathbb{R} 上的线性空间称作实线性空间。
- \mathbb{C} 上的线性空间称作复线性空间。



除此之外，还有多项式空间也是线性空间，下面给出定义。

定义 1.13

所有次数不超过 n 的实系数多项式的集合，记作 P_n 。



P_n 中的元素是形如 $p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ 的多项式，其中 $a_0, a_1, \dots, a_n \in \mathbb{R}$ ，接下来我们来验证 P_n 是线性空间。

推论 1.4

P_n 是 \mathbb{R} 或 \mathbb{C} 上线性空间。



下面给出证明（读者并不要求）

证明 验证加法，两个多项式相加是按系数相加，例如：

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \dots + (a_n + b_n)x^n$$

可以得到他们的和依然是多项式。其次验证标量乘法，标量 $c \in \mathbb{R}$ 与多项式 $p(x)$ 的乘积是

$$cp(x) = ca_0 + ca_1x + \dots + ca_nx^n$$

所得的结果依然是多项式，若 $P(x) = 0$ ，代表的是该线性空间的零向量，满足线性空间的所有公理。□

思考

你能不能举出其他的一些例子表示线性空间？



1.4 子空间

1.4.1 子空间 (subspace) 的定义

定义 1.14

如果 V 的子集 U 也是线性空间，则称 U 是 V 的子空间。



例如 $U = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}$ 是 \mathbb{R}^3 的子空间。需要注意的是，线性空间的子集也得是线性空间所以才能是子空间，所以请读者思考下面的一个问题。

思考

$U = \{(x_1, x_2, 1) \mid x_1, x_2 \in \mathbb{R}\}$ 是否是 \mathbb{R}^3 的子空间？

答案为不是，因为它不满足 $\mathbf{0} \in U$ ，根本就不是线性空间，所以不是 \mathbb{R}^3 的子空间。除此之外，子空间需要满足下面的情况（也就是说，也需要满足作为线性空间的条件）。

推论 1.5

当且仅当 V 的子集 U 满足以下三个条件时， U 是 V 的子空间。

- 存在加法单位元 $\mathbf{0} \in U$ ；
- 对于加法封闭，即 $v, w \in U$ 满足 $v + w \in U$ ；
- 对于标量乘法封闭，即 $\lambda \in \mathbb{F}, v \in U$ 满足 $\lambda v \in \mathbb{F}$ 。



思考

若集合 $U = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}$ ，集合 $T = \{(0, 0, x_3) \mid x_3 \in \mathbb{R}\}$ ， $T \cup U$ 是否是 \mathbb{R}^3 的子空间？

答案为不是，因为它不满足对于加法封闭，取 $u \in U, u = (1, 1, 0), t \in T, t = (0, 0, 1)$, $u+t = (1, 1, 1) \notin U \cup T$ ，所以不是 \mathbb{R}^3 的子空间。

1.4.2 子空间之和

为了更深入了解线性空间以及知识完整性，下面讲解子空间和的概念。

定义 1.15 (子空间的和)

子空间的和是有其所有元素可能的和所构成的集合，其结果仍是子空间。若 $V_1, V_2, V_3, \dots, V_m$ 是 V 的子空间，则子空间的和 V_s 记作

$$V_s = V_1 + V_2 + V_3 + \dots + V_m = \sum_{i=1}^m V_i$$

或者说来

$$V_s = \sum_{i=1}^m V_i = \left\{ \sum_{i=1}^m v_i \mid i \in \mathbb{N}^+, v_i \in V_i \right\}$$



不要被定义上的 Sigma 符号吓跑了，实际上就是分别任意两个属于不同子空间的元素相加后的新元素所构成的空间，例如子空间 $V_1 = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}$ 和 $V_2 = \{(0, x_3, x_4) \mid x_3, x_4 \in \mathbb{R}\}$ ，则 V_1, V_2 子空间的和是 \mathbb{R}^3 ，即 \mathbb{R}^3 中所有的元素 r 均存在 $t, s \in \mathbb{R}$ 使得 $v_1 \in V_1, v_2 \in V_2$ 满足

$$r = tv_1 + sv_2$$

所以可以记作 $\mathbb{R}^3 = V_1 + V_2$ 。



上面讲了子空间的和，请读者先思考子空间的和与子空间的并的区别。

思考

子空间的并是否一定是子空间？在什么情况下子空间的并是子空间？

首先回答第一个，子空间的并不一定是子空间，如上面的一个思考 $T \cup U$ 一样， U, T 均为 \mathbb{R}^3 的子空间，但是我们证明了它们的并集不是子空间。对于第二个问题，首先我们考虑 V 子空间 T 和 U ，如果 T 是 U 的非空子集（思考为什么非空？），那么 $T \cup U = U$ 是子空间显然成立，若 T 不是 U 的非空子集，设 $t \in T, u \in U$ 由子空间的性质应有 $t + u$ 属于 V ，即 $t + u$ 可能属于 T 也可能属于 U 。但是若 $t + u$ 属于 T ，而根据封闭性 $u = (t + u) - t$ 也应当属于 T ， $t + u$ 和 u 都是子空间 T 的元素，所以与“若 T 不是 U 的非空子集”的设定不符。

推论 1.6

两个子空间的并集是子空间，当且仅当其中一个包含于另一个。



1.4.3 直和

首先我们回顾一下笛卡尔积⁴，如果读者已经忘记了的话请阅读定义1.2，我们了解到 \mathbb{R}^3 空间由 $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ 通过笛卡尔积生成。而直和也是生成生成空间的一种方式，是一种特殊的子空间之和。

定义 1.16 (直和 (direct sum))

设 $V_1, V_2, V_3, \dots, V_m$ 是 V 的子空间，如果 $V_1 + V_2 + V_3 + \dots + V_m$ 中的每个元素都能用 $\sum_{i=1}^m v_i, v_i \in V_i$ 唯一地表示出来，则称子空间之和 $\sum_{i=1}^m V_i$ 为直和，可写作

$$V = V_1 \oplus V_2 \oplus V_3 \oplus \dots \oplus V_m$$



例如 $A = \{(x, y, 0) \mid x, y \in \mathbb{R}\}, B = \{(0, 0, z) \mid z \in \mathbb{R}\}$ ，则 $\mathbb{R}^3 = A \oplus B$ ；而 $A = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}, B = \{(0, 0, z) \mid z \in \mathbb{R}\}$ ，则 $\mathbb{R}^3 \neq A \oplus B$ 。

推论 1.7

U 和 W 是 V 的子空间且 $V = U \oplus W$ ，当且仅当 $W \cap U = \{0\}$ 。



接下来我们继续研究构建线性代数的基本内容，即向量与空间的关系。

1.5 张成空间与线性相关性

1.5.1 线性组合与张成空间

一组向量进行标量乘法后相加形成的一个向量，称作该向量的线性组合。

定义 1.17 (线性组合 (linear combination))

给定向量集合 $\{v_1, v_2, \dots, v_k\}$ 和标量集合 $\{c_1, c_2, \dots, c_k\}$ ，向量 v 的线性组合表示为：

$$v = c_1 v_1 + c_2 v_2 + \dots + c_k v_k$$

其中， c_1, c_2, \dots, c_k 称为组合系数。



通过一组向量的线性组合所产生的向量能表示的空间是这个向量的张成空间，下面给出定义。

⁴ 笛卡尔积的另一个名称叫做直积，和本部分的直和的运算方式有关联。

定义 1.18 (张成空间 (span))

给定向量集合 $S = \{v_1, v_2, \dots, v_k\}$ (其中 v_i 是线性空间 V 中的向量), 这些向量的张成空间是所有可能的线性组合构成的集合, 记作:

$$\text{Span}(S) = \{c_1v_1 + c_2v_2 + \dots + c_kv_k \mid c_1, c_2, \dots, c_k \in \mathbb{F}\}$$

如果 $V = \text{Span}(S)$, 则称向量集合 S 张成^a空间 V 。

^a这里的张成是动词。



例如 $(1, 2, 4) \in \text{Span}(\{(1, 0, 0), (0, 1, 2)\})$ 因为存在线性组合使得 $(1, 2, 4) = (1, 0, 0) + 2(0, 1, 2)$, 而 $(1, 2, 5) \notin \text{Span}(\{(1, 0, 0), (0, 1, 2)\})$ 因为不存在线性组合使得 $(1, 2, 5) = \lambda(1, 0, 0) + \beta(0, 1, 2)$ 。

再有例子 $\mathbb{R}^3 = \text{Span}(\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\})$, 因为任何一个 $v \in \mathbb{R}^3$ 都可以由 $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ 的线性组合构成。

推论 1.8

张成空间 $\text{Span}(S)$ 是包含向量 S 空间的子空间。



不要笑! 可能这个概念对读者来说就像 $2 + 2 = 4$ 一样简单, 这个推论是显然的, 张成空间是子空间, 但是我们还是来证明它。

证明 对于 $V = \text{Span}(S), S = \{v_1, v_2, v_3, \dots, v_n\}$, 有加法单位元 $\mathbf{0} \in V$, 在此空间 V 中

$$\mathbf{0} = \sum_{i=1}^n (0v_i)$$

在空间 V 中满足在加法封闭

$$\sum_{i=1}^n a_i v_i + \sum_{i=1}^n c_i v_i = \sum_{i=1}^n (a_i + c_i) v_i$$

也满足在标量乘法下封闭

$$\lambda \sum_{i=1}^n a_i v_i = \sum_{i=1}^n \lambda a_i v_i$$

所以张成空间 $\text{Span}(S)$ 是包含向量 S 空间的子空间。



1.5.2 线性相关性

下面我们给出线性代数的一个关键定义, 从此往后, 我们线性代数研究的内容就是有限维的线性空间。

定义 1.19 (有限维线性空间)

线性空间如果可以由其空间内的有限个向量张成, 则这个线性空间是有限维的。



同时, 我们在此基础上研究线性相关性; 设 $v_1, v_2, v_3, \dots, v_n \in V$ 且 $v \in \text{Span}(V)$, 根据张成空间的定义我们有 $a_1, a_2, a_3, \dots, a_n \in \mathbb{F}$ 使得

$$v = \sum_{i=1}^n a_i v_i = a_1 v_1 + a_2 v_2 + a_3 v_3 + \dots + a_n v_n$$

若 $v = \mathbf{0}$, 我们有一组显然的结果, 即当 $a_1 = a_2 = a_3 = \dots = a_n = 0$ 的时候, $v = \mathbf{0}$, 然而通常来说, 我们可能还会遇到其他的 a_i 组合也可以使得 $v = \mathbf{0}$, 于是我们做出如下定义:



定义 1.20 (线性无关⁵(linearly independent))

对于向量 $v_1, v_2, v_3, \dots, v_n$ 如果使得其线性组合 $a_1v_1 + a_2v_2 + a_3v_3 + \dots + a_nv_n = \mathbf{0}$ 的情况只有 $a_1 = a_2 = a_3 = \dots = a_n = 0$, 则称向量 $v_1, v_2, v_3, \dots, v_n$ 线性无关。

定义 1.21 (线性相关 (linearly dependent))

若一些向量不是线性无关的, 就称为线性相关。

例题 1.6.

向量 $(2, 3, 1), (1, -1, 2), (7, 3, c)$ 线性相关, 求 c 的值。

解: 若三个向量线性相关, 则有 $\alpha, \beta, \gamma \in \mathbb{R}, \alpha(2, 3, 1) + \beta(1, -1, 2) + \gamma(7, 3, c) = \mathbf{0}$, 可得线性方程

$$\begin{cases} 2\alpha + 1\beta + 7\gamma = 0 \\ 3\alpha + (-1)\beta + 3\gamma = 0 \\ 1\alpha + 2\beta + c\gamma = 0 \end{cases}$$

若使得线性方程组有非全为 0 解, 我们任意取 $\alpha = 1$, 则根据前两个式子可得

$$\begin{cases} 1\beta + 7\gamma = -2 \\ (-1)\beta + 3\gamma = -3 \end{cases}$$

解其可得 $\beta = 1.5, \gamma = -0.5$ 而 $\alpha = 1$, 带入第三式, $1 + 2 \times 1.5 - 0.5c = 0$ 最终结果为

$$c = 8$$

推论 1.9

如果一组向量线性相关, 那么其中一个向量均可以由其他向量的线性组合表示。



证明 一组向量 v_1, v_2, \dots, v_n 是线性相关的, 当且仅当存在不全为零的标量 c_1, c_2, \dots, c_n , 使得:

$$c_1v_1 + c_2v_2 + \dots + c_nv_n = \mathbf{0}$$

由于向量组线性相关, 至少存在一个标量 $c_k \neq 0 (1 \leq k \leq n)$ 而将上述等式中的 v_k 单独移到等式一侧:

$$c_kv_k = -c_1v_1 - c_2v_2 - \dots - c_{k-1}v_{k-1} - c_{k+1}v_{k+1} - \dots - c_nv_n$$

两边同除 c_k 可得

$$v_k = -\frac{c_1}{c_k}v_1 - \frac{c_2}{c_k}v_2 - \dots - \frac{c_{k-1}}{c_k}v_{k-1} - \frac{c_{k+1}}{c_k}v_{k+1} - \dots - \frac{c_n}{c_k}v_n$$

本题得证。□

这个例子也说明了线性相关的向量集合张成的空间与其去掉任意一个向量后张成的空间是同一个空间。

此外也有一条重要的结论, 若向量集合 V 内的元素线性无关, 则 V 中元素的个数不大于张成该空间 U 的向量的个数。举个例子, $(0, 0, 1)$ 和 $(0, 1, 0)$ 在 \mathbb{R}^3 线性无关, 而张成 \mathbb{R}^3 的向量可能由 $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ (或更多其他 3 个不共面的向量) 构成, 但是既要线性无关, 而且还要张成这个空间的向量集合(下面会给这些向量集合一个名字叫做基) 在 \mathbb{R}^3 上为 3 个⁶。

根据上述推论我们有 Steinitz 替换定理:

⁶关于这个推论可能会比较抽象, 读者可以回想在高中的学习生活中, 通过空间直角坐标系的三个基底可以表示整个空间, 缺一不可, 且 3 个不共面的向量都可以成为基底, 在后面的学习中, 我们会给它一个名词叫做极大线性无关组。



定理 1.1 (Steinitz 替换定理)

一个向量集合 A 是线性无关的，另一个向量组 B 生成了 A 所在线性空间，设 A, B 的向量个数分别为 s, r 则 $s \leq r$ 。


这里我们不作证明，但是希望读者能够理解这个定理；例如在线性空间 \mathbb{R}^3 上，我们可以选择 3 个线性无关向量作为基底，此外也有 3 个及以上能够张成此空间的向量。⁷

1.5.3 基与维数

通过线性无关和张成空间的两个概念的结合，我们可以得到基的概念。

定义 1.22 (基 (basis))

若 V 中的一组向量线性无关且能够张成 V ，则称这组向量的集合称作 V 的基。


所以按照上面的定义，基有两个方面的约束：

- 基的选取必须是线性无关的，例如 $\{(1, 0), (2, 0), (0, 1)\}$ 不能选为 \mathbb{R}^2 的基，因为 $\mathbf{0} = -2(1, 0) + (2, 0) + 0(0, 1)$ ，这个例子说明二维平面内的基不能共线。
 - 基的选取必须能张成选取的空间，例如 $\{(1, 0, 0), (0, 1, 0)\}$ 不能选为 \mathbb{R}^3 的基，因为它们只能张成 $\{(x, y, 0) \mid x, y \in \mathbb{R}\}$ ，这个例子说明两个不共线的向量只能表示整个平面。
- 集合 $\{(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), (0, 0, 1, \dots, 0), \dots, (0, 0, 0, \dots, 1)\}$ 是 \mathbb{F}^n 的基，它比较特殊，是为 \mathbb{F}^n 的标准基。

推论 1.10

若空间 V 包含向量集合 $S = \{v_1, v_2, v_3, \dots, v_n\}$ 且 S 是空间 V 的基，则仅有唯一的 $a_1, a_2, a_3, \dots \in \mathbb{F}$ 线性组合表示任意一个 V 的元素 v ，即每个 $v \in V$ 都能唯一地表示成下面的形式：

$$v = a_1v_1 + a_2v_2 + a_3v_3 + \dots + a_nv_n$$

**例题 1.7.**

求 $U = \{(a, b, c, d, e) \in \mathbb{R}^5 \mid a = 2b, c = 4d\}$ 的一个基集合。

解：子空间 U 中的向量可以表示为：

$$(a, b, c, d, e) = (2b, b, 4d, d, e)$$

将向量表示为自由变量的线性组合

$$(2b, b, 4d, d, e) = b(2, 1, 0, 0, 0) + d(0, 0, 4, 1, 0) + e(0, 0, 0, 0, 1)$$

这表明 U 中的任意向量都可以由向量组 $\{(2, 1, 0, 0, 0), (0, 0, 4, 1, 0), (0, 0, 0, 0, 1)\}$ 线性表示，因此，子空间 U 的一个基集合为

$$\{(2, 1, 0, 0, 0), (0, 0, 4, 1, 0), (0, 0, 0, 0, 1)\}$$

接下来我们引入维数的概念：

定义 1.23 (维数 (dimension))

有限维线性空间的任意一个基集合元素的个数称为该空间的维数，若 V 是有限维的，那么它的维数记作 $\dim V$ 。


⁷关于此内容我们会在节 5.4.1 详细讲述

例如上面的例题所表示，子空间 U 的基向量元素的个数为 3，即 $\dim U = 3$ ，需要注意的是它虽然在 \mathbb{R}^5 空间内，但是张成它至少只需要 3 个向量，维度为 3 而不是 5。相当于 $\{(1, 0, 0), (0, 1, 0)\}$ 只能张成一个平面，但这个平面是放在立体空间内的。

至此，关于线性空间的内容我们就讲到这里，后面给出一些练习供读者参考。

1.6 章节练习

1.6.1 A 组

练习 1 求 $\left(\frac{-1 + \sqrt{3}i}{2}\right)^3$ 。

练习 2 设子空间

$$U = \{(x, x, y, y) \mid x, y \in \mathbb{F}\}, W = \{(x, x, x, y) \mid x, y \in \mathbb{F}\}$$

求子空间的和 $U + W$ 。

练习 3 判断 $\{(x, y, z) \mid x, y, z \in \mathbb{R}, xyz = 0\}$ 是否是 \mathbb{R}^3 的子空间。

练习 4 写出在 \mathbb{R}^3 中过点 $(1, 4, 1)$ 且方向向量等于 $(2, 3, 1)$ 的直线方程。

练习 5 写出在 \mathbb{R}^3 中的平面 $2x - y + 3z = 1$ 的一个向量表示。

练习 6 求 $t \in \mathbb{R}$ 使得 $(3, 1, 4), (2, -3, 5), (5, 9, t)$ 线性相关。

练习 7 求 $\{(x, x, y) \mid x, y \in \mathbb{R}\}$ 的一个基集合。

练习 8 若子空间 V 的 $\dim V = 1$ ，求子空间 V 的一个可能解。

1.6.2 B 组

练习 9 若 $a, b \in \mathbb{R}^2$ ，设 O 是平面直角坐标系的原点， P_1, P_2 分别为平面上一点，其中 $a = \overrightarrow{OP_1} = (x_1, y_1), b = \overrightarrow{OP_2} = (x_2, y_2)$ ，若直线 OP_1 与直线 OP_2 垂直，求 $x_1y_1 + x_2y_2$ 。

练习 10 若 $\text{Span}(\{v_1, v_2, v_3, v_4\}) = \{(a, b, b, b) \mid a, b \in \mathbb{R}\}$ ，求 $\text{Span}(\{v_1 - v_2, v_2 - v_3, v_3 - v_4, v_4\})$ 。

练习 11 若 U_1, U_2 是有限维线性空间的子空间，已知 $\dim U_1 = 3, \dim U_2 = 4, \dim(U_1 \cap U_2) = 2$ ，求 $\dim(U_1 + U_2)$ 的值。

练习 12 若 U_1, U_2 是有限维线性空间的子空间，求证

$$\dim U_1 + \dim U_2 - \dim(U_1 \cap U_2) = \dim(U_1 + U_2)$$

练习 13 设 U, W 是 \mathbb{R}^n 的子空间且 $\dim U + \dim W = \mathbb{R}^n$ ，求 $U \oplus W$ 。

练习 14 若 $U_1, U_2, U_3, \dots, U_m$ 均为 V 的有限维子空间，证明

$$\dim \sum_{i=1}^m U_i \leq \sum_{i=1}^m (\dim U_i)$$



第 2 章 矩阵与线性映射

「所以生命啊，它苦涩如歌」

——《我用什么把你留住》

2.1 线性方程组与矩阵

2.1.1 向量的矩阵表示

这是一个前置知识，只需要读者记住即可，向量的矩阵表示是将向量竖起来写。

定义 2.1 (向量的矩阵表示)

设向量 $x \in \mathbb{F}^n$, $x = (x_1, x_2, x_3, \dots, x_n)$ 的矩阵表示是

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

记作 \mathbf{x} ^a 或 $\mathcal{M}(x)$; 而矩阵的向量表示则表示为 $\mathcal{V}(x)$, 因此, 满足下面的等式

$$\mathbf{x} = \mathcal{M}(x); x = \mathcal{V}(\mathbf{x})$$

^a今后我们表示矩阵使用粗体的字母



则该 $n \times 1$ 的矩阵也有相关的运算法则, 分别为矩阵和与标量积, 类比向量进行相关运算后竖着写则其表示形式, 这里不再重复赘述相关性质, 读者可根据章节1.2来自行证明, 下面给出相关定义:

定义 2.2 ($n \times 1$ 矩阵的加法)

若矩阵 \mathbf{A}, \mathbf{B} 同为 $n \times 1$ 矩阵, 其中 $\mathbf{A} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$ 则 $\mathbf{A} + \mathbf{B}$ 定义为

$$\mathbf{A} + \mathbf{B} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ \vdots \\ x_n + y_n \end{pmatrix}$$



同样的道理我们定义乘法

定义 2.3 ($n \times 1$ 矩阵的标量乘法)

若标量 $\lambda \in \mathbb{F}$, 其对一个 $n \times 1$ 矩阵 $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$ 的乘积定义为

$$\lambda \mathbf{x} := \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \\ \vdots \\ \lambda x_n \end{pmatrix}$$



2.1.2 线性方程组与平面上的线性变换

回顾初中所学的一些内容, 我们了解到了二元一次方程组, 它们形如:

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$

这就是线性方程组的一种, 通常可用于平面上的线性变换, 下面一种情况就是平面上的线性变换:

我们平常所使用的坐标系是直角坐标系, 表示向量使用标准基 $\{(1, 0), (0, 1)\}$, 所谓的线性变换是指如果将坐标系进行一个变化, 使得构成这个系的基由原来的 $\{(1, 0), (0, 1)\}$ 变成 $\{(x_1, y_1), (x_2, y_2)\}$ 后所表示的向量在原坐标系中所表示的坐标。如图2.1所示, 如果我们的基由原来的 $\{(1, 0), (0, 1)\}$ 变成 $\{(1, 2), (2, 1)\}$, 那么变换前坐标 $(4, 5)$ 的线性组合是 $(4, 5) = 4(1, 0) + 5(0, 1)$, 而变化基后的坐标则变为 $(14, 13) = 4(1, 2) + 5(2, 1)$, 即空间内的每一点 $(x, y) = x(1, 0) + y(0, 1)$ 都变化为 $(x + 2y, 2x + y) = x(1, 2) + y(2, 1)$, 设变化后的空间的每一个点在原坐标系的坐标表示为 (x', y') ; 表示为线性方程组就是

$$\begin{cases} 1x + 2y = x' \\ 2x + 1y = y' \end{cases}$$



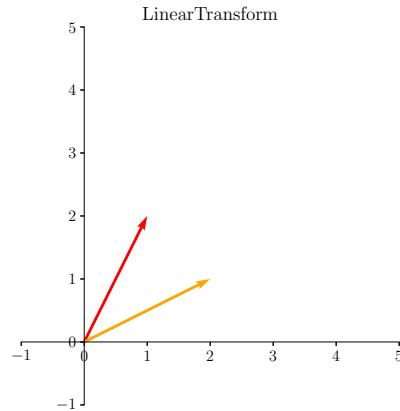


图 2.1: 线性变换

如果我们把这个式子进一步抽象，可以得到一个矩阵方程

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (2.1)$$

例题 2.1.

在平面直角坐标系中，向量 $a = \overrightarrow{OP_1} = (3, 2), b = \overrightarrow{OP_2} = (4, 7)$ ；

1. 求 $\triangle OP_1P_2$ 的面积；
2. 若将 x 轴的每一个值扩大一倍，求标准基经过变换后的向量集合；
3. 若将 y 轴的每一个值扩大一倍，求变换后的三角形面积 $\triangle OP'_1P'_2$

解：

1. 可使用割补法和叉积法，对于割补法，如图2.2b所示，则是将其填充为矩形，即 $S = 4 \times 7 - S_1 - S_2 - S_3$ 其中 $S_1 = 4 \times 7 \div 2 = 14, S_2 = (2 + 7) \times 1 \div 2 = 4.5, S_3 = 2 \times 3 \div 2 = 3$ 则 $\triangle OP_1P_2$ 的结果为 $S = 28 - 14 - 18 - 3 = 6.5$ ，若使用叉积法，则有公式

$$S = \frac{1}{2} |x_1y_2 - x_2y_1| = \frac{1}{2} = |3 \times 7 - 2 \times 4| = 6.5$$

2. $\{(2, 0), (0, 1)\}$
3. 若 y 轴扩大一倍，则该部分所有的面积都扩大一倍，所以 $6.5 \times 2 = 13$

为了方便表示，我们把公式2.1乘式中最左边的内容提取出来，称作变化矩阵，这个例子将是我们以后研究矩阵的一个特殊情况；为了方便描述，我们将该矩阵用 \mathbf{A} 表示

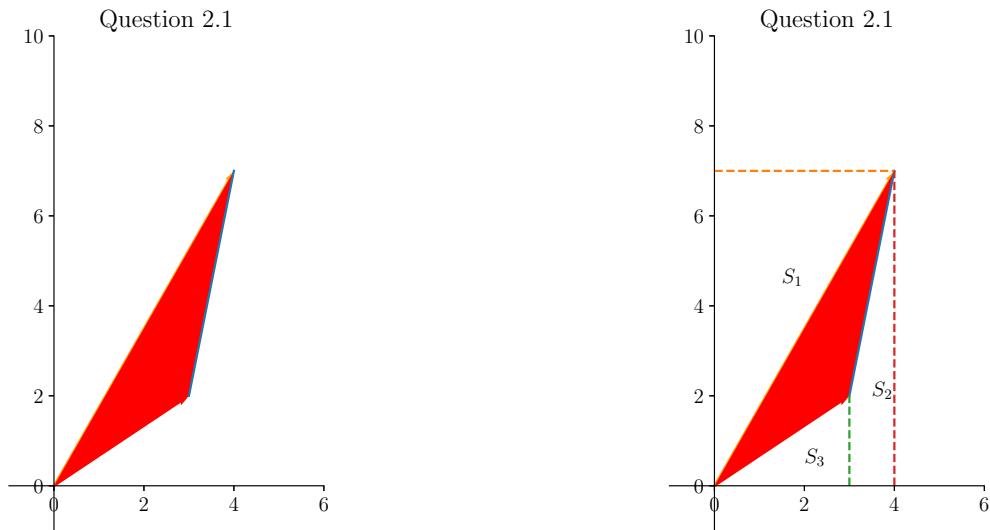
$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

若使用 \mathbf{x} 和 \mathbf{x}' 分别表示 (x, y) 和 (x', y') 的矩阵形式，那么公式2.1可表示为

$$\mathbf{x}' = \mathbf{A}\mathbf{x}$$

接下来我们将矩阵 \mathbf{A} 中的第 i 行第 j 列元素表示为 $\mathbf{A}_{i,j}$ ，例如 $A_{1,2} = 2$ 。

通过上面的描述，虽然我们并没有严格证明，但是我们也可以得知线性空间经过线性变换后，其结果仍然是线性空间。我们还是从二维平面出发来讲解矩阵的相关性质，于是我们有如下推论：

(a) 三角形 $\triangle OP_1P_2$ 图

(b) 割补法

图 2.2: 例题 2.1 图像

推论 2.1

在二维平面内, 矩阵 \mathbf{A} 表示其空间内的线性变换, 设 $x, y \in \mathbb{R}^2$ 且其矩阵表示分别为 \mathbf{x}, \mathbf{y} , 则有如下推论:

1. $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}$
2. 若 $\lambda \in \mathbb{R}$ 则有 $\lambda(\mathbf{Ax}) = \mathbf{A}(\lambda\mathbf{x})$

**证明**

1. 设 $x = (x_1, y_1), y = (x_2, y_2)$, 则对于等式左侧 $x + y = (x_1 + x_2, y_1 + y_2)$, 其矩阵表示为 $\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$, 设 2×2 矩阵 \mathbf{A} , 令

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a, b, c, d \in \mathbb{R}$$

那么类比矩阵方程 2.1, 设结果向量为 (x', y') 我们可得矩阵式

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$$

的方程组式为

$$\begin{cases} x' = a(x_1 + y_1) + b(x_2 + y_2) \\ y' = c(x_1 + y_1) + d(x_2 + y_2) \end{cases}$$

而根据标量加法分配法则, 方程组可写成

$$\begin{cases} x' = ax_1 + ay_1 + bx_2 + by_2 \\ y' = cx_1 + cy_1 + dx_2 + dy_2 \end{cases} \quad (2.2)$$

同理右侧可表示为方程组式 2.2, 等式成立。□

2. 读者自证



2.1.3 矩阵表示线性方程组

首先讲解什么是线性方程组，定义如下：

定义 2.4 (线性方程组)

线性方程组是由多个线性方程组成的集合，通常表示为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

其中， x_1, x_2, \dots, x_n 是未知数， a_{ij} 是系数， b_i 是常数项。



根据上面的矩阵式2.1所表示的二元一次方程组的几何意义，同样这里也可以看作是 n 维向量 $x = (x_1, x_2, x_3, \dots, x_n)$ 经过一个复杂的变化 \mathbf{A} 变为 $b = (b_1, b_2, b_3, \dots, b_n)$ 一样，所以同样我们可以将线性方程组的一般形式抽象成一个矩阵：

$$\mathbf{Ax} = \mathbf{b}$$

其中

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

2.1.4 矩阵的定义

我们已经对矩阵有了一个较为感性的认知，终于我们来到了系统讲解矩阵的这一部分，不过可能会让读者失望的是，矩阵的定义并非是你之前所认识的那样，是一个表示线性变换的符号；事实上，矩阵的真正定义其实就是一堆数字按照 $m \times n$ 的方式排列而已，下面我们给出矩阵的定义。

定义 2.5 (矩阵的定义)

矩阵是一个由圆括号^a包裹的按照矩形排列的 $m \times n$ 的数表，一个 $m \times n$ 的矩阵 \mathbf{A} 可以表示为：

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

其中： m 是行数， n 是列数。 a_{ij} 表示矩阵中第 i 行第 j 列的元素。

^a在一些教材中，会使用方形括号表示



所有元素都是实数的矩阵叫作实矩阵，接下来我们将上面所讲到的 $n \times 1$ 矩阵拓展到一般的矩阵计算。

2.2 矩阵的运算法则

2.2.1 矩阵和

定义 2.6 ($m \times n$ 矩阵的加法)

若两个 $m \times n$ 的同类型矩阵 \mathbf{A}, \mathbf{B} 如果

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix},$$

那么

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$



实际上，两个矩阵相加，就是把它们里面的所有元素依次相加并得到一个新的矩阵。与向量的加法相同的是，我们只定义了同类型的矩阵才能相加。

公设 2.1

只有类型相同的矩阵才能互相相加。



那么也是和向量一样，它们满足交换律。

推论 2.2

若 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ 均为 $m \times n$ 的矩阵，它们满足交换律 $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ 和结合律 $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{B} + (\mathbf{A} + \mathbf{C})$



请读者仿照向量的运算律，验证上述推论。

2.2.2 0 矩阵

根据公设2.1，我们得知相同类型的矩阵的加法才有意义，所以我们定义在矩阵 $\mathbf{A} + \mathbf{B} = \mathbf{A}$ ，那么 \mathbf{B} 就是 0 矩阵

定义 2.7 (0 矩阵)

若 \mathbf{A}, \mathbf{B} 均为 $m \times n$ 的矩阵，且满足 $\mathbf{A} + \mathbf{B} = \mathbf{A}$ ，那么 \mathbf{B} 为 $m \times n$ 矩阵的单位元，显然，0 矩阵就是全是由 0 组成的矩阵，记作 \mathbf{O} ^a，那么

$$\mathbf{O} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

^a不同于向量的记法，是以粗体形式出现的字母 \mathbf{O} 。



2.2.3 加法逆元

类似于标量加法的 $x, y \in \mathbb{C}, x + y = 0$ 与向量加法的 $x, y \in \mathbb{F}^n, x + y = \mathbf{0}$, 矩阵中同样表示若 \mathbf{A}, \mathbf{B} 为同类型的矩阵, $\mathbf{A} + \mathbf{B} = \mathbf{O}$, $\mathbf{B} = -\mathbf{A}$ 表示 \mathbf{A} 的加法逆元。

定义 2.8 (加法逆元)

对于矩阵 \mathbf{A} , \mathbf{A} 的加法逆元表示满足下面条件的矩阵 $-\mathbf{A}$ 有

$$\mathbf{A} + (-\mathbf{A}) = \mathbf{O}$$

换而言之当

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

时,

$$-\mathbf{A} = \begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix}$$

2.2.4 标量乘法

定义 2.9 (矩阵的标量乘法)

若标量 $\lambda \in \mathbb{F}$, 其对一个矩阵 $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ 的乘积满足以下的运算法则

$$\lambda \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}$$

同时, 矩阵的标量乘法满足标量结合律。

推论 2.3

若 $a, b \in \mathbb{F}$, \mathbf{A} 表示一个 $m \times n$ 的矩阵, 满足 $(ab)\mathbf{A} = a(b\mathbf{A})$



2.2.5 矩阵乘法

这里讲一个特殊的乘法叫做矩阵乘法, 矩阵乘法是线性代数中的一种基本运算, 用于将两个矩阵相乘, 得到一个新的矩阵。需要注意的是, 我们在定义矩阵乘法的时候, 只有左侧的矩阵行数等于右侧矩阵的列数, 这个矩阵乘法才有意义, 例如下面的矩阵乘法是有意义的。

$$1. \quad \begin{pmatrix} 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix}$$

$$2. \quad \begin{pmatrix} 1 & 5 \\ 8 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix}$$

下面给出矩阵乘法的定义。



定义 2.10 (矩阵乘法的定义)

假设有两个矩阵 A 和 B , 其中 A 是一个 $m \times n$ 的矩阵, B 是一个 $n \times p$ 的矩阵, 那么它们的乘积 $C = A \times B$ 将是一个 $m \times p$ 的矩阵。

矩阵乘法的计算规则如下:

对于矩阵 C 中的每一个元素 c_{ij} , 它是通过将矩阵 A 的第 i 行与矩阵 B 的第 j 列对应元素相乘后再相加得到的。具体公式为:

$$c_{ij} = \sum_{k=1}^n a_{ik} \times b_{kj}$$

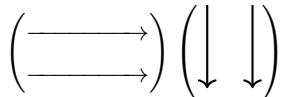
其中

- a_{ik} 是矩阵 A 的第 i 行第 k 列的元素,
- b_{kj} 是矩阵 B 的第 k 行第 j 列的元素。



如定义所示, sigma 符号还是如此的令人讨厌, 读者不妨和笔者慢下来, 在大脑内建立一个乘积的画面, 下面我们给出矩阵乘法的直观操作。请注意, 下面的内容请结合定义2.10一起阅读, 这样可以加深对矩阵加法的理解。

根据定义2.10, 直观的矩阵乘应该以下面这样的结构进行相乘运算:



最简单的如 $\begin{pmatrix} 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix}$, 乘法则为左矩阵从左往右, 右矩阵从上到下, 依次相乘的结果相加, 我们可得

$$\begin{pmatrix} 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix} = (1 \times 2 + 3 \times 4 + 7 \times 8) = (2 + 12 + 56) = (70)^1$$

再复杂一些, 我们尝试 $\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix}$, 求矩阵就是重复按照上述操作填充表的过程, 依照公式, $c_{11} = a_{11}b_{11} + a_{12}b_{21}$

$$\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 1 \times 2 + 5 \times 3 = 17 & ? \\ ? & ? \end{pmatrix}$$

以此类推, $c_{12} = a_{11}b_{12} + a_{12}b_{22}$

$$\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 17 & 1 \times 4 + 5 \times 7 = 39 \\ ? & ? \end{pmatrix}$$

$$c_{21} = a_{21}b_{11} + a_{22}b_{21}$$

$$\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 17 & 39 \\ 8 \times 2 + 6 \times 3 = 34 & ? \end{pmatrix}$$

$$c_{22} = a_{21}b_{12} + a_{22}b_{22}$$

$$\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 17 & 39 \\ 34 & 8 \times 4 + 6 \times 7 = 74 \end{pmatrix}$$

¹这里的 70 不是元组, 也不是向量, 而是 1×1 矩阵

至此，我们填充了整个矩阵表，所以 $\begin{pmatrix} 1 & 5 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} = \begin{pmatrix} 17 & 39 \\ 34 & 74 \end{pmatrix}$

例题 2.2.

求下列矩阵的乘积：

$$1. \begin{pmatrix} 1 & 4 & -1 \\ -2 & 6 & 5 \end{pmatrix} \begin{pmatrix} 0 & 2 & 5 \\ -1 & 1 & 7 \\ 3 & 2 & -3 \end{pmatrix}$$

$$2. \begin{pmatrix} 1 & 1 & -4 \\ 5 & -1 & 4 \\ -1 & 9 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ 8 \\ 1 \end{pmatrix}$$

$$3. \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

$$4. \begin{pmatrix} e & f \\ g & h \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

5. 矩阵乘法是否满足交换律？

解：

$$1. \begin{pmatrix} -7 & 4 & 36 \\ 9 & 12 & 17 \end{pmatrix}$$

$$2. \begin{pmatrix} 21 \\ 41 \\ 64 \end{pmatrix}$$

$$3. \begin{pmatrix} bg + ae & af + bh \\ cd + gh & cf + dh \end{pmatrix}$$

$$4. \begin{pmatrix} cf + ae & df + be \\ ag + ch & bg + dh \end{pmatrix}$$

5. 否

2.2.6 平面上的线性变换

欲要了解矩阵的作用，我们先来了解一下线性变换。

平面上的线性变换是指通过一个矩阵将平面上的每一个点（向量）映射到另一个点的变换。这种变换满足线性性质，即保持向量加法和标量乘法。

在平面直角坐标系内，常见的线性变换包括：

1. 缩放：沿坐标轴方向拉伸或压缩，例如例题2.1.2的第2, 3小问；
2. 旋转：坐标轴绕原点旋转一定角度 θ ；
3. 剪切：使图形沿某一方向倾斜，例如图2.2；
4. 反射：关于某条直线或点的对称变换。

由此，我们可以得到在 \mathbb{R}^2 线性变换的性质：

- 保持原点不变；
- 保持直线性：直线变换后仍为直线。

而矩阵则表示了线性变换的过程，矩阵与矩阵的乘积表示线性变换后的过程，在这里我们介绍一个术语叫做左乘，矩阵的乘法表示一般从右往左写，例如：如果我们将正交的直角坐标系中的每一个向量都顺时针旋转

α , 那么一个向量 $a = (x, y)$ 经过变化后的坐标 (x', y') 满足下面这一个式子:

$$\begin{cases} x' = \cos \alpha \cdot x - \sin \alpha \cdot y \\ y' = \sin \alpha \cdot x + \cos \alpha \cdot y \end{cases}$$

接下来我们将这个方程组抽象为矩阵, 可得

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

提取矩阵并给其一个符号

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

仔细观察矩阵 \mathbf{A} 第一列的两个数按序写成 2 元组, 即为 $(\cos \alpha, \sin \alpha)$, 如果读者观察仔细的话, 正是标准基其中的一个 $(1, 0)$ 经过该变换后的坐标, 而第二列的 $(-\sin \alpha, \cos \alpha)$ 是标准基另一个 $(0, 1)$ 变化后的坐标, 而该矩阵乘积所得的矩阵, 正是变化后在原坐标系中的向量表示。

那如果我们在旋转的基础上继续旋转 β 弧度, 我们直接可以得到矩阵 $\mathbf{C} = \begin{pmatrix} \cos(\alpha + \beta) & -\sin(\alpha + \beta) \\ \sin(\alpha + \beta) & \cos(\alpha + \beta) \end{pmatrix}$, 如果表示成两个矩阵相乘, 则是矩阵 \mathbf{A} 左乘一个矩阵 \mathbf{B} , 该矩阵 \mathbf{B} 即为只旋转 β 弧度的矩阵表示, 则

$$\mathbf{B} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix}$$

写完整些是这样的

$$\begin{pmatrix} \cos(\alpha + \beta) & -\sin(\alpha + \beta) \\ \sin(\alpha + \beta) & \cos(\alpha + \beta) \end{pmatrix} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$

请读者按照矩阵的乘法验证该等式成立。

2.3 线性映射

2.3.1 从平面走向更高维度

线性代数研究的不只是平面, 更是更高维度的内容, 如果我们把目光从平面 \mathbb{R}^2 与 \mathbb{R}^3 上移到 \mathbb{F}^n , 那么我们现在把“线性变换”这个词改一个说法, 改成“线性映射”来表示一种函数。

2.3.2 线性空间中的线性映射

如果读者忘记了什么是线性空间, 请回到章节 1.3.1 回顾一下它们的基本性质, 在线性空间中, 线性映射表示的是一个线性空间 V 的元素到另一个线性空间的元素 W , 即 $T : V \rightarrow W$

定义 2.11 (线性映射)

若 V, W 均为线性空间, $y \in W, x \in V$, 则函数 $y = T(x)$ 表示线性映射当且仅当满足下面的性质:

1. 加性 (additivity): 对所有 $u, v \in V$ 都有 $T(u + v) = T(u) + T(v)$;
2. 齐性 (homogeneity): 对所有 $\lambda \in \mathbb{F}$ 和 $v \in V$ 都有 $T(\lambda v) = \lambda T(v)$

通常满足从 V 到 W 的所有线性映射所构成的集合记为 $\mathcal{L}(V, W)$, 例如上面的函数 $T \in \mathcal{L}(V, W)$



思考

分别验证在平面上的旋转, 剪切, 缩放是一种线性映射。

对于旋转线性映射, 设 V 是映射前的空间, 映射前向量 $v = (x, y), v \in V$ 经过绕远点旋转 α 后其空间变为



W , 设 $w = (x', y')$, $w \in W$ 为映射后的向量, 那么它们满足

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

首先他们满足加性, 若有向量 $v_1 = (x_1, y_1)$, 旋转后则变为 $w_1 = (x'_1, y'_1)$ 满足

$$\begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

则一定满足

$$\begin{pmatrix} x' + x'_1 \\ y' + y'_1 \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x + x_1 \\ y + y_1 \end{pmatrix}$$

其次他们满足齐性, 设 $\lambda \in \mathbb{F}$, 则有

$$\begin{pmatrix} \lambda x' \\ \lambda y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$$

所以旋转在平面上是一种线性映射, 限于篇幅, 请读者自行验证剪切与放缩。

2.3.3 零空间

我们将线性映射从 V 到 W 中, 将原来空间中的某个映射为 $\mathbf{0}$ 的子空间叫做零空间, 下面给出其定义:

定义 2.12 (零空间 (null space))

对于映射 $T : V \rightarrow W$, T 的零空间是在 V 中经过 T 映射为 $\mathbf{0}$ 向量的子空间, 记为 $\text{null } T^{\text{a}}$, 则

$$\text{null } T = \{v \in V : T(v)^{\text{b}} = \mathbf{0}\}$$

^a有些书本会使用核空间, 并记作 $\text{Ker } T$

^b对于线性映射 T , 一些书本会将其简写为 Tv



同样用 \mathbb{R}^2 的例子, 存在一个映射 $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, 若表示这是一个将平面上的所有点压缩为一条线的一个映射, 设该法则将 $v \in \mathbb{R}^2$, $v = (x, y)$ 映射为 $w \in \mathbb{R}^2$, $w = (x, 0)$, 记为 $T(v) = w$ 或 $T((x, y)) = (x, 0)$, 那么零空间 $\text{null } T$ 则为:

$$\text{null } T = \{(0, t) \mid t \in \mathbb{R}\}$$

思考

上面的 $\dim \text{null } T$ 的值是多少?

由于 $\text{null } T$ 只需要一个基 $\{(0, 1)\}$, 使得任意一个 $v \in \text{null } T$ 存在 $\alpha \in \mathbb{R}$ 满足 $v = \alpha(0, 1)$ 表示, 所以 $\dim \text{null } T = 1$, 即使它放在了 \mathbb{R}^2 中。

例题 2.3.

若 $\varphi \in \mathcal{L}(\mathbb{C}^3, \mathbb{C})$, $v_1 \in \mathbb{C}^3$, $v_2 \in \mathbb{C}$, 定义线性映射 $\varphi(v_1) = v_2$, 其中 $v_1 = (x_1, x_2, x_3)$, $v_2 = x_1 + 2x_2 + 3x_3$ 则 $\varphi((x_1, x_2, x_3)) = x_1 + 2x_2 + 3x_3^{\text{a}}$, 求 $\text{null } \varphi$ 的一个基集合与 $\text{null } \varphi$ 的维数。

^a如果以后没有特殊说明, 在线性映射中的每个未知数前后相通

解: $\text{null } \varphi = \{(a, b, c) \mid a + 2b + 3c = 0\}$ 令 $a = 0$ 则 $b = 3, c = -2$, 若令 $b = 0$ 则 $a = -3, c = 1$, 若令 $c = 0$ 则 $a = -2, b = 1$ 而向量 c 可由 a, b 的线性组合得到, 所以 3 者向量线性相关, 故去除一个向量后二者向量线性无关, 则 $\text{null } \varphi$ 的一个基为 $\{(0, 3, -2), (-3, 0, 1)\}^{\text{a}}$, 根据基的元素个数, 我们可以推断



$\dim \text{null}\varphi = 2$ 。

^a后面我们会讲解求线性方程组并使用更加正确的方式求基

2.3.4 值域空间

作为一个函数有个映射法则 T , 那么我们给经过 T 映射后的空间取个名字叫做值域空间。

定义 2.13 (值域空间 (range space))

对于映射 $T : V \rightarrow W$, T 的值域空间是在 V 中经过 T 法则映射为 W 的所有元素的集合, 记作 $\text{range}T^{\text{a}}$, 则

$$\text{range}T = \{T(v) \mid v \in V\}$$

^a有些书本会使用像空间, 并记作 $Im T$



用之前一个 \mathbb{R}^2 的例子, 存在一个映射 $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, 若表示这是一个将平面上的所有点压缩为一条线的一个映射, 设该法则将 $v \in \mathbb{R}^2, v = (x, y)$ 映射为 $w \in \mathbb{R}^2, w = (x, 0)$, 记为 $T(v) = w$ 或 $T((x, y)) = (x, 0)$, 那么值域空间 $\text{range}T$ 则为:

$$\text{range}T = \{(t, 0) \mid t \in \mathbb{R}\}$$

很显然, 此处的 $\dim \text{range}T = 1$ 。

例题 2.4.

若 $\varphi \in \mathcal{L}(\mathbb{C}^3, \mathbb{C})$, $v_1 \in \mathbb{C}^3$, $v_2 \in \mathbb{C}$, 定义线性映射 $\varphi(v_1) = v_2$, 其中 $v_1 = (x_1, x_2, x_3)$, $v_2 = x_1 + 2x_2 + 3x_3$ 则 $\varphi((x_1, x_2, x_3)) = x_1 + 2x_2 + 3x_3$, 求 $\text{range}\varphi$ 的一个基集合与 $\text{range}\varphi$ 的维数。

解: $\text{range}\varphi = \mathbb{C}$ 那么 $\text{null}\varphi$ 的一个基为 $\{1\}$, 也就是说 \mathbb{C} 上的任意一个元素都可以用 $\alpha 1, \alpha \in \mathbb{C}$ 的线性组合表示, 所以 $\dim \text{range}\varphi = 1$ 。

2.3.5 维数定理

下面是一个非常重要的一个定理。

定理 2.1 (维数定理)

V, W 是有限维线性空间, 法则 $T \in \mathcal{L}(V, W)$, $T : V \rightarrow W$ 使得下述成立:

$$\dim V = \dim \text{null}T + \dim \text{range}T$$



证明 设 $A = \{v_1, v_2, v_3, \dots, v_m\}$ 是 $\text{null}T$, 则 $\dim \text{null}T = m$ 且集合 $\{v_1, v_2, v_3, \dots, v_m\}$ 线性无关, 那么 V 的基一定包含集合 B , 但是 V 的基的元素个数一定大于 B 元素的个数, 可设 $\{v_1, v_2, v_3, \dots, v_m, u_1, u_2, \dots, u_n\}$ 是 V 的基, 则 $\dim V = m + n$; 我们只需要证明 $\dim \text{range}T = n$ 。

对于每一个 $v \in V$ 都可以使用 $\{v_1, v_2, v_3, \dots, v_m, u_1, u_2, \dots, u_n\}$ 的线性组合表示, 即

$$v = a_1 v_1 + a_2 v_2 + a_3 v_3 + \dots + a_m v_m + b_1 u_1 + b_2 u_2 + \dots + b_n u_n, a_i \in \mathbb{F}, i \in \mathbb{N}^{+}$$

对等式两边的所有向量进行 T 法则映射, 则一部分属于 $\text{null}T$ 的基 A 将会被映射为 $\mathbf{0}$, 即

$$T(v) = \mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0} + b_1 T(u_1) + b_2 T(u_2) + \dots + b_n T(u_n)$$

² \mathbb{N}^{+} 代表正整数



所以我们可以得到 $A_T = \{T(u_1), T(u_2), T(u_3), \dots, T(u_m)\}$ 张成 $\text{range } T$, 即 $\text{Span}(A_T) = \text{range } T$ 。

下面继续证明 A_T 线性无关, 设 $\mathbf{0}$ 可以由 A_T 的线性组合得到, 即

$$c_1 T(u_1) + c_2 T(u_2) + c_3 T(u_3) + \dots + c_n T(u_n) = \mathbf{0}$$

那么根据线性映射的标量乘法的性质

$$T(c_1 u_1 + c_2 u_2 + c_3 u_3 + \dots + c_m u_m) = \mathbf{0}$$

设 $u = c_1 u_1 + c_2 u_2 + c_3 u_3 + \dots + c_m u_m$ 则有 $u \in \text{null } T$, 而 $\text{Span}(A) = \text{null } T$ 且 A 线性无关, 那么当且仅当 $c_1 = c_2 = c_3 = \dots = c_m = 0$ 时, 可由 A 的线性组合表示 u , 此时仅可表示 $\mathbf{0}$ 向量, 所以 A_T 线性无关, 故 $\text{range } T$ 的基为 $\{u_1, u_2, \dots, u_m\}$, 所以

$$\dim V = \dim \text{null } T + \dim \text{range } T \quad \square$$

拿之前的例子我们验证该定理, 例如例题 2.4 所表示的那样, $\dim \mathbb{R}^3 = 3$, $\text{null } \varphi = \{(a, b, c) \mid a + 2b + 3c = 0\}$ 则 $\dim \text{null } \varphi = 2$, 同时 $\dim \text{range } \varphi = 1$ 满足题设条件。

其实这个定理也说明了线性映射后只会丢失信息而不会增加信息, 线性映射后的空间维度只会减少不会增加, 即使你是 $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ 看似升高维度, 实际上这种映射在三维空间中还是一个平面, 只不过这个平面只能在三维空间中表示。所以它们只需要两个线性无关的向量的线性组合就可以表示映射后空间内所有的向量。

2.4 矩阵表示线性映射

2.4.1 线性映射的运算法则

线性映射也有加法和标量乘法, 它们定义如下:

定义 2.14 (线性映射的加法与标量乘法)

若 $S, T \in \mathcal{L}(V, W)$, $\alpha \in \mathbb{F}$ 则 $S + T$ 是一个从 V 到 W 的线性映射, 若 $v \in V$, 它们满足

$$(S + T)(v) = S(v) + T(v)$$

标量积 αT 满足

$$(\alpha T)(v) = \alpha(T(v))$$

此外, 线性映射也有相乘操作, 它们如同矩阵一样, 从右往左相乘, 表示先进行右边的线性映射, 再进行左边的线性映射。

定义 2.15 (线性映射的乘法)

若 $S, T \in \mathcal{L}(V, W)$, 则 ST 是一个从 V 到 W 的线性映射, 若 $v \in V$, 它们满足

$$(ST)(v) = S(T(v))^{\text{a}}$$

^a从右往左, 一定是先进行 T 再进行 S

2.4.2 矩阵描述线性映射

根据线性映射的基本定理, 我们知道经过线性映射后维度不会升高, 若 $\{v_1, v_2, v_3, \dots, v_n\}$ 为 V 的一个基, 对于每一个经过线性映射 $T \in \mathcal{L}(V, W)$ 的基为 $W_t = \{Tv_1, Tv_2, Tv_3, \dots, Tv_n\}$ 必然有 $\text{range } T = \text{Span}(W_t)$, 接下来我们使用矩阵记录 V 上的基的每一项经过映射后的 $T(v_i)$, 都可以由 W 上的基的线性组合表示。

这么说可能有些抽象且复杂, 我们给出定义并从定义入手:

定义 2.16 (线性映射的矩阵表示)

设 $T \in \mathcal{L}(V, W)$, $\{v_1, v_2, v_3, \dots, v_n\}$ 是线性空间 V 的基, $\mathcal{M}(T)$ 为 $m \times n$ 的矩阵, 其元素的第 i 行第 j 列表示为 a_{ij} , 那么根据这些基的线性映射的矩阵表示中的元素 a_{ij} 满足

$$T(v_k) = a_{1k}w_1 + a_{2k}w_2 + a_{3k}w_3 + \dots + a_{mk}w_m$$



或许读者会感觉这个定义更加抽象了, 我们还是不妨使用老朋友 \mathbb{R}^2 和 \mathbb{R}^3 来描述。

例题 2.5.

设 $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ 若向量 $a = (x, y)$ 经过变换后的向量为 $b = (x + 3y, 2x + 5y, 7x + 9y)$ 那么 $T(a) = b, T((x, y)) = (x + 3y, 2x + 5y, 7x + 9y)$, 求关于 \mathbb{R}^2 与 \mathbb{R}^3 上的标准基 $\{(1, 0), (0, 1)\}$ 和 $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ 线性映射的矩阵表示。^a

^a本题参见 Linear Algebra Done Right 的 3.33

解: $T((1, 0)) = (1, 2, 7)$ 那么 \mathbb{R}^3 的标准基 $w_1 = (1, 0, 0), w_2 = (0, 1, 0), w_3 = (0, 0, 1)$ 可表示为

$$v_1 = (1, 0), T(v_1) = (1, 2, 7) = 1w_1 + 2w_2 + 7w_3$$

所以 $(1, 2, 7)$ 依次为矩阵第一列的元素 a_{11}, a_{21}, a_{31} , 而 $T((0, 1)) = (3, 5, 9)$, 那么

$$v_2 = (0, 1), T(v_2) = (3, 5, 9) = 3w_1 + 5w_2 + 9w_3$$

所以 $(3, 5, 9)$ 依次为矩阵第二列的元素 a_{12}, a_{22}, a_{32} , 所以关于标准基的线性映射矩阵表示为

$$\begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}$$

前文我们说过线性映射也有自己的运算法则, 然后矩阵也有自己的运算法则, 那么两者之间是否存在关系, 或者说这两者可以是统一的呢? 其实答案相信对在座的学霸们来说已经能够看到这两者的关系了。

线性映射的加法可以描述为法则 S, T 将向量 v 经过这两个法则映射后的和, 反应在矩阵上便有如下的推论 (我们假设 $S, T, S + T$ 都使用同样的基):

推论 2.4

若 $S, T \in \mathcal{L}(V, W), \alpha \in \mathbb{F}$, 则 $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T), \alpha\mathcal{M}(T) = \mathcal{M}(\alpha T)$ 。



同样的道理, 线性映射的乘法同样可以抽象为矩阵的乘法, 需要注意的是, 它们都不满足交换律, 表示从右往左, 先后映射的内容。

推论 2.5

若 $T \in \mathcal{L}(U, V), S \in \mathcal{L}(V, W)$ 则 $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ 。



2.5 章节练习

2.5.1 A 组

练习 15

计算矩阵 $\mathbf{A} = \begin{pmatrix} 3 & -1 & 5 \\ \sqrt{2} & 4 & 0 \\ -2 & -7 & 6 \end{pmatrix}$ 与 $\mathbf{B} = \begin{pmatrix} 3 & 2 & \sqrt{7} \\ 4 & -2 & 1 \\ -5 & 0 & 6 \end{pmatrix}$ 的和。

练习 16

计算矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 3+i & \sqrt{5} \\ 2-i & 3 & -4 \end{pmatrix}$ 与 $\mathbf{B} = \begin{pmatrix} -1 & i & \sqrt{2} & 4 \\ 9 & 3 & -4 & 3 \\ 2-i & 4 & 8 & 1 \end{pmatrix}$ 的乘积。

练习 17 若 $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$, 有

$$T((x, y, z)) = (a + x + y, bxyz)$$

求 b, c 的值。练习 18 根据线性映射的加法和标量乘法, 试证明 $\mathcal{L}(V, W)$ 是线性空间。练习 19 三体文明有一个工具叫做二向箔, 这个工具可以将空间二维化, 设该变化为简单的线性映射 $T : V \rightarrow W$, $T \in \mathcal{L}(V, W)$ 满足

$$T((x, y, z)) = (x, y)$$

求 $\text{range } T$ 和 $\text{null } T$ 。练习 20 平面上有三点 $A(3, 2), B(-1, 4), C(2, 0)$, 请使用叉积法求 $S_{\triangle ABC}$ 。练习 21 三维空间 \mathbb{R}^3 中任意一点 (x, y, z) , 若将 \mathbb{R}^3 的标准基均向直线方向 $x = y = z$ 旋转 $\frac{\pi}{3}$, 请使用关于 (x, y, z) 的矩阵式表示 (x', y', z') 练习 22 举例线性映射 T 使得 $\dim \text{range } T = 2, \dim \text{null } T = 3$

2.5.2 B 组

练习 23 线性方程组常数项为 0

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = 0 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = 0 \end{array} \right.$$

可视为线性映射

$$T(x_1, x_2, \dots, x_n) = \left(\sum_{i=1}^n a_{1i}x_i, \sum_{i=1}^n a_{2i}x_i, \dots, \sum_{i=1}^n a_{mi}x_i \right)$$

除 $x_1 = x_2 = x_3 = \cdots = x_n = 0$ 外, $\dim \text{null } T$ 需要满足条件才有其他解?

练习 24 若 $S \in \mathcal{L}(V, W), T \in \mathcal{L}(U, V)$ 求证

$$\dim \text{null}ST \leq \dim \text{null}S + \dim \text{null}T$$

练习 25 请不要看笔者的证明过程，自己复盘一遍线性映射的基本定理，并尝试自己写证明过程。



第3章 内积空间

「花有重开日，人无再少年」

——《续侄溥赏酴醾劝酒二首·其一》

3.1 内积

3.1.1 向量的点积与长度表示

本章请读者自行考虑是否阅读，如果读者对线性代数要求不是那么高，可以跳过该章节，本章较有难度，但是看下去会很有意思。

在高中的学习生涯中，我们总是把二维空间 \mathbb{R}^2 与三维空间 \mathbb{R}^3 中的向量具体化为一个有向箭头。通常我们求两点之间的坐标距离会使用勾股定理，在前面的学习中，[定义1.10](#)就是使用此方法计算的向量长度。

根据[定义1.10](#)我们可以得到平面上向量 $x = (x_1, x_2)$ 的长度 $\|x\| = \sqrt{x_1^2 + x_2^2}$ ，将等式两边同时平方我们可得

$$\|x\|^2 = x_1^2 + x_2^2,$$

同时下面定义向量的点积 $x \cdot y$:

定义 3.1 (向量的点积 (dot product))

设 $x, y \in \mathbb{R}^n$, $x = (x_1, x_2, x_3, \dots, x_n)$, $y = (y_1, y_2, y_3, \dots, y_n)$ 则向量 x 与 y 的点积 $x \cdot y$ 定义为

$$x \cdot y := x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$



由于这个概念由实空间出发，所以满足如下公设。

公设 3.1

只有 \mathbb{R}^n 上的向量才能进行点积运算。



需要注意的是，点积是将两个向量 $x, y \in \mathbb{R}^n$ 映射为一个实数，所以它们具有如下的性质。



推论 3.1

对于所有的 $x, y, z \in \mathbb{R}^n$ 有：

1. $x \cdot y = y \cdot x$
2. $x \cdot (y \cdot z) = (x \cdot y) \cdot z$
3. $x \cdot (y + z) = x \cdot y + x \cdot z$
4. 若 $T(x, y) = x \cdot y$ 则 $T \in \mathcal{L}((\mathbb{R}^n, \mathbb{R}^n), \mathbb{R})$ (表示内积法则操作是线性映射)
5. $\|x\|^2 = x \cdot x$



下面的例题可以用于验证上述推论的其中几个。

例题 3.1.

已知点 $A(1, 3), B(4, 1)$ 在平面直角坐标系中，若向量 $x, y, z \in \mathbb{R}^2$ ，其中 $x = \overrightarrow{OA}, y = \overrightarrow{OB}, z = \overrightarrow{AB}$ ；

1. 求向量 z 的表示；
2. 求 $\|z\|^2$ ；
3. 使用勾股定理求线段 AB^a 的长度

^a如果没有加任何标记，则表示线段

解：

1. $z = y - x, z = (4, 1) - (1, 3) = (3, -2)$
2. $\|z\|^2 = z \cdot z = 9 + 4 = 13$
3. 略

3.1.2 内积的定义

我们刚刚了解了点积，并定义了在实空间上的定义，实际上内积就是点积的推广形式，使其在复空间内可用。我们知道对于一个实数 $r \in R$ ，有

$$|r| = \sqrt{r^2}$$

那么对于一个在二维平面 \mathbb{R}^2 和三维空间 \mathbb{R}^3 中的向量求它的长度，例如 $x = (r_1, r_2, r_3)$ 那么有

$$\|x\| = \sqrt{|r_1|^2 + |r_2|^2 + |r_3|^2}$$

如果继续向更高维度推广，那么对于 $r \in \mathbb{R}^n, r = (r_1, r_2, r_3, \dots, r_n)$ 就有

$$\|r\| = \sqrt{|r_1|^2 + |r_2|^2 + |r_3|^2 + \dots + |r_n|^2}$$

那如果我们将其推广到复空间，对于 $z = a + bi$ ，复空间向量的“长度”¹就可以这样表示，若 $z \in \mathbb{C}^n, z = (z_1, z_2, z_3, \dots, z_n)$ 那么

$$\|z\| = \sqrt{|z_1|^2 + |z_2|^2 + |z_3|^2 + \dots + |z_n|^2}$$

至于 $|z|^2$ 如何计算，下面给出复数的模的定义：

定义 3.2 (复数的模，共轭复数)

复数 z 可以表示为 $z = a + bi, a, b \in \mathbb{R}$ ，我们定义

$$|z|^2 := a^2 + b^2$$

¹后面我们会给它一个名字就叫做范数



同时 $|z|^2 = (a + bi)(a - bi)$, 而 $a - bi$ 我们称其为 z 的共轭复数, 记作 \bar{z} , 所以我们有

$$|z|^2 = z\bar{z}$$

所以 $z \in \mathbb{C}^n, z = (z_1, z_2, z_3, \dots, z_n)$, 根据到复空间的推广和复数的模的定义可得

$$\|z\| = \sqrt{z_1\bar{z}_1 + z_2\bar{z}_2 + z_3\bar{z}_3 + \dots + z_n\bar{z}_n}$$

推广到复空间的向量点积我们称其为内积, 如果我们将范围限定在 \mathbb{R}^n 上对两个向量进行操作, 那么实空间上的内积就是点积, 下面我们给出定义:

定义 3.3 (内积的定义)

若向量 $a, b \in \mathbb{C}^n, a = (z_1, z_2, z_3, \dots, z_n), b = (c_1, c_2, c_3, \dots, c_n)$ 那么它们的内积记作 $\langle a, b \rangle$ 表示为

$$\langle a, b \rangle := z_1\bar{c}_1 + z_2\bar{c}_2 + z_3\bar{c}_3 + \dots + z_n\bar{c}_n$$

我们知道, $c \in \mathbb{R}, c = a + 0i$, 所以 $|c|^2 = c^2$, 所以说实空间上面的内积就是点积。

推论 3.2

V 是 \mathbb{C} 上的线性空间, 向量 $u, v \in V$, 标量 $\lambda \in \mathbb{C}$ 对于其内积 $\langle u, v \rangle$ 有如下性质

1. 正性 (positivity): 对于所有的 v , $\langle v, v \rangle \geq 0$;
2. 定性 (definiteness): 当且仅当 $v = 0$ 时 $\langle v, v \rangle = 0$;
3. 对第一变量的线性 (linearly for first slot): $\lambda \langle u, v \rangle = \langle \lambda u, v \rangle, \langle u + w, v \rangle = \langle u, v \rangle + \langle w, v \rangle$;
4. 共轭对称性 (conjugate symmetry): $\langle u, v \rangle = \overline{\langle v, u \rangle}$.



例题 3.2.

求证: 对所有的 $u, v, w \in \mathbb{C}^n$ 都有 $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$

解: 根据共轭对称性, 则有

$$\langle u, v + w \rangle = \overline{\langle v + w, u \rangle}$$

根据内积的定义, 有

$$\overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle}$$

再根据共轭对称性有

$$\overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} = \langle u, v \rangle + \langle u, w \rangle$$

这证明了内积的第二个变量也满足线性。



3.1.3 范数

我们把之前所有实空间长度和复空间 $v \in \mathbb{C}^n$ 与自身的内积 $\langle v, v \rangle$ 赋予一个新名词叫做范数, 它的定义如下:

定义 3.4 (范数 (norm))

对于 $v \in V$ 那么 v 的范数 $\|v\|$ 定义为

$$\|v\| := \sqrt{\langle v, v \rangle}$$



3.1.4 内积空间

本章节主要研究内积空间，所以我们做出如下定义：

定义 3.5 (内积空间)

若线性空间 H 有 $x, y, z \in H$ 使得在 H 上存在函数 $f(x, y) = z$, 它们满足下面的内积映射条件

$$f : (H, H) \rightarrow \mathbb{C}$$

其中

$$f(x, y) = \langle x, y \rangle = z$$

那么称 H 为内积空间^a并满足推论3.2的相关性质。

^a完备的内积空间称为 Hilbert(希尔伯特) 空间



思考

内积空间是否是线性空间？

答案是一定是线性空间，留给读者自己思考并查阅相关资料证明。

至于内积空间它能够干什么，下面给出通过内积空间衍生的几个著名不等式。

3.1.5 Cauchy-Schwarz 不等式

首先我们来看 Cauchy-Schwarz² 不等式。

定理 3.1 (Cauchy-Schwarz 不等式)

设内积空间 H 中有 $v, w \in H$ 则有如下不等式成立：

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

当且仅当 $u = \lambda v, \lambda \in \mathbb{F}$ 。



证明 由正性与定性³可得，对于任意的 $x, y \in H, \lambda \in \mathbb{F}$ 有

$$\langle x + \lambda y, x + \lambda y \rangle \geq 0$$

由共轭对称性可知

$$\langle x + \lambda y, x + \lambda y \rangle = \overline{\langle x + \lambda y, x + \lambda y \rangle}$$

将其展开可得

$$\langle x + \lambda y, x + \lambda y \rangle = \langle x, x \rangle + \bar{\lambda} \langle x, y \rangle + \lambda \langle y, x \rangle + |\lambda|^2 \langle y, y \rangle \geq 0$$

首先考虑 $y = 0$ 的情况， $(\langle x, 0 \rangle)^2 = 0, \langle x, x \rangle = x, \langle 0, 0 \rangle = 0$ 等号严格成立，若 $y \neq 0$ 令 $\lambda = -\frac{\langle x, y \rangle}{\langle y, y \rangle}$ 并带入可以证明。□

下面是 Cauchy-Schwarz 不等式的两种形式，其中第一种是法国数学家 Augustin Louis Cauchy⁴在 1821 年证明，第二个由德国数学家 Hermann Amandus Schwarz⁵在 1886 年证明。

²一般译作柯西-施瓦茨

³可以合起来，叫做正定性

⁴一般译作：奥古斯丁·路易斯·柯西

⁵一般译作：赫尔曼·阿曼杜斯·施瓦茨

推论 3.3

1. 若 $x_1, x_2, x_3, \dots, x_n, y_1, y_2, y_3, \dots, y_n \in \mathbb{R}$ 则有

$$(x_1y_1 + x_2y_2 + \dots + x_ny_n)^2 \leq (x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2)(y_1^2 + y_2^2 + y_3^2 + \dots + y_n^2)$$

写成求和符号就是

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n y_i^2\right)$$

2. 设 f, g 在 $[-1, 1]$ 上连续, 则

$$\left(\int_{-1}^1 f(x)g(x)dx\right)^2 \leq \left(\int_{-1}^1 (f(x))^2 dx\right) \left(\int_{-1}^1 (g(x))^2 dx\right)$$



3.1.6 余弦定理

在二维空间 \mathbb{R}^2 和三维空间 \mathbb{R}^3 的空间中, 有一条余弦定理, 它能够帮助我们解三角形, 定理如下。

定理 3.2 (余弦定理)

在三角形内, 设 a, b 为角 C 的邻边, c 为角 C 的对边, 则它们满足

$$c^2 = a^2 + b^2 - 2ab \cos C$$



读者可使用几何法尝试验证, 留作练习。

下面我们对其进行推广, 我们可以得到内积空间上的余弦定理, 同时有内积公式。

推论 3.4

若 $a, b \in H$ 则有

$$\|c\|^2 = \|a\|^2 + \|b\|^2 - 2\langle a, b \rangle$$



为了更好的引入内积公式, 下面我们定义 \mathbb{C}^n 上的两向量的夹角的值。

定义 3.6 (两向量夹角的值)

若 $a, b \in \mathbb{C}^n$ 则 a, b 的夹角 φ 满足

$$\varphi := \cos^{-1} \frac{\langle a, b \rangle}{\|a\| \|b\|}$$



请读者举例验证在 $\mathbb{R}^2, \mathbb{R}^3$ 上的正确性, 并使用余弦定理证明在 $\mathbb{R}^2, \mathbb{R}^3$ 的正确性。

例题 3.3.

给出复变反余弦函数的公式: 若 $z \in \mathbb{C}$ 那么其反余弦的值为

$$\cos^{-1} z = -i + \ln \left(z + i\sqrt{1 - z^2} \right)$$

利用上述公式计算: 已知 $a, b \in \mathbb{C}^n, a = (i, 2i, 3), b = (1, 2, 3i)$ 求向量 a, b 的夹角的余弦值 φ 。

解: 根据定义, 计算 $\langle a, b \rangle = i + 4i - 9i = -4i, \|a\| = \|b\| = \sqrt{14}$ 所以

$$\varphi = \cos^{-1} \left(-\frac{2}{7}i \right)$$



带入上述公式，可得

$$\varphi = -i + \ln \left(-\frac{2}{7}i + i\sqrt{1 + \frac{4}{49}} \right)$$

实数部分化简后可得

$$\varphi = -i + \ln \left(\frac{\sqrt{53} - 2}{7}i \right)^a$$

^a由于这本书不讲复变，所以就化简至此，不取其主值。

下面给出内积公式。

定理 3.3 (内积公式)

若 $u, v \in \mathbb{C}^n$, φ 为 u, v 的夹角，那么它们的内积表示为

$$\langle u, v \rangle = \|u\| \|v\| \cos \varphi$$



根据两向量夹角的值代入可以轻松证明该定理成立，或使用余弦定理证明在 $\mathbb{R}^2, \mathbb{R}^3$ 上的正确性。

3.1.7 三角不等式

如图3.1，我们都应该知道，三角形两边之和大于第三边，在二维空间 \mathbb{R}^2 和三维空间 \mathbb{R}^3 的空间中⁶，这是作为一条公理存在的，即：

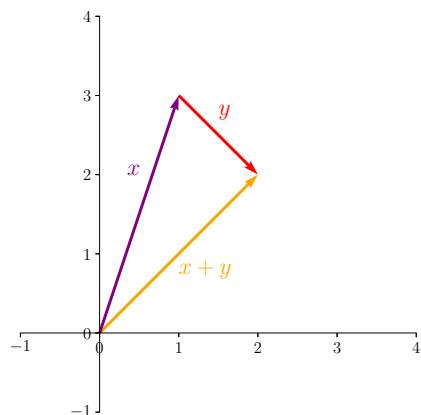


图 3.1: 三角不等式

公理 3.1 (平面长度最短公理)

在平面内，两点之间，线段最短。



如果将其推广到内积空间，则有如下不等式成立：

⁶这里的空间体系都是欧几里得体系



定理 3.4 (三角不等式)

若 $u, v \in H$, 则

$$\|u + v\| \leq \|u\| + \|v\|$$

当且仅当 $u = \lambda v, \lambda > 0$ 等号成立。



证明 首先根据范数的定义, 我们可以得到

$$\|u + v\|^2 = \langle u + v, u + v \rangle$$

根据两个变量的线性, 我们可以得到

$$\langle u + v, u + v \rangle = \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle$$

根据共轭对称性可得

$$\langle u + v, u + v \rangle = \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \overline{\langle u, v \rangle}$$

化简并合并可得

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2\operatorname{Re} \langle u, v \rangle^7$$

适当取放缩可得

$$\|u + v\|^2 \leq \|u\|^2 + \|v\|^2 + 2|\langle u, v \rangle|$$

根据 Cauchy-Schwarz 不等式可得

$$\|u\|^2 + \|v\|^2 + 2|\langle u, v \rangle| \leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|$$

最后我们合并不等式右侧, 可得

$$\|u + v\|^2 \leq (\|u\| + \|v\|)^2$$

两边开二次根即可得证。□

3.2 正交**3.2.1 正交的定义**

根据内积公式 $\langle u, v \rangle = \|u\|\|v\|\cos\varphi$, 特别地, 两个向量的内积等于 0 的时候我们有一个关键的定义: 正交。

定义 3.7 (正交 (orthogonal) 的定义)

若向量 $a, b \in H$ 有 $\langle a, b \rangle = 0$ 则称 a, b 是正交的^a。

^a有些教材会使用 a, b 是垂直的



通常来讲, 向量的范数不为 0, 当 $\cos\varphi$ 为 0 的时候两个向量正交; 也就是说, 通常两个向量的夹角为 $\frac{\pi}{2}$ 的时候两个向量正交。

对于向量 $\mathbf{0}$ 我们有如下推论:

推论 3.5

1. $\mathbf{0}$ 与 H 中的任何向量正交;
2. $\mathbf{0}$ 与自身正交。



⁷若 $x = a + bi$ 那么 $\operatorname{Re}(x)$ 表示取 x 的实数部分 a , $\operatorname{Im}(x)$ 表示取 x 的虚数部分 b



3.2.2 正交基

首先是正交基的概念，若基集合 $\{a_1, a_2, a_3 \dots, a_n\}$ 是正交基，则该基的元素两两正交。

定义 3.8 (正交基 (orthonormal basis))

若内积空间 H 上的基 $S = \{a_1, a_2, a_3 \dots, a_n\}$ 满足对于任意一个 $i, j \in \mathbb{N}^+$ 且 $i \neq j$ 使得

$$\langle a_i, a_j \rangle = 0$$

我们称 S 为 H 的正交基。



举个例子， \mathbb{R}^3 的正交基可以是 $\{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$ 。

3.2.3 Gram-Schmidt 正交化

在前面的学习中，若线性空间 V 的基为 $S = \{a_1, a_2, a_3, \dots, a_n\}$ 那么对于任意一个 $v \in V$ 都可以使用 S 的线性组合表示，即对于任何一个 $v \in V$ 存在 $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ 满足下面的式子成立。

$$v = \lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3 + \dots + \lambda_n a_n$$

由于正交基是一种特殊的基，所以它们也满足该性质，我们有如下的等式来描述它们的线性组合表示的向量，称作正交分解。

定理 3.5 (正交分解 (orthogonal decomposition))

设内积空间 V 上的基为 $S = \{a_1, a_2, a_3, \dots, a_n\}$ ，则对于任何一个 $v \in V$ 满足等式

$$v = \langle v, a_1 \rangle a_1 + \langle v, a_2 \rangle a_2 + \dots + \langle v, a_n \rangle a_n$$

成立，写成求和符号就是

$$v = \sum_{i=1}^n \langle v, a_i \rangle a_i$$

并同时满足

$$\|v\|^2 = \sum_{i=1}^n \left(\frac{|\langle v, a_i \rangle|}{\|a_i\|} \right)^2$$



我们还是从 \mathbb{R}^2 和 \mathbb{R}^3 入手考虑，通常在我们高中的物理学习中会学到力的分解，下面是在 \mathbb{R}^2 上的图形直观的解释。

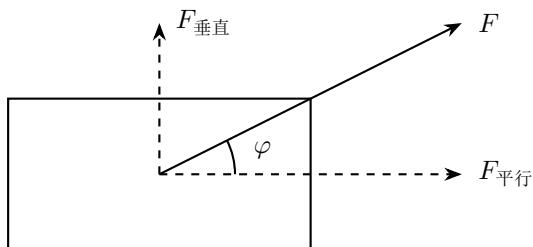


图 3.2: 力的合成与分解

如图3.2所示， $F_{\text{垂直}}$ 和 $F_{\text{平行}}$ 是正交的，我们可以将 F 分解为这两个向量，根据内积公式， $\langle F, F_{\text{平行}} \rangle = \|F\| \|F_{\text{平行}}\| \cos \varphi$ 所以分解为 $F_{\text{平行}}$ 的分量可以使用正交分解公式，同理 $F_{\text{垂直}}$ 也是如此，这样一来可以将 F 分解为两个正交基的和即

$$F = F_{\text{垂直}} + F_{\text{平行}}$$



同理，在 \mathbb{R}^n 上的向量均可使用定理3.5使其正交化；而下面要讲的一个内容表示的是，可以将一些线性无关的向量正交化，使其两两正交。且正交化后张成的空间与原向量张成的空间相同，这就是Gram-Schmidt⁸正交化过程。

定理 3.6 (Gram-Schmidt 正交化)

设向量 $S = \{v_1, v_2, v_3, \dots, v_n\}$ 线性无关，存在向量 $S' = \{a_1, a_2, a_3, \dots, a_n\}$ 为正交基，并满足

$$a_k = v_k - \sum_{i=1}^{k-1} \frac{\langle a_i, v_k \rangle}{\langle a_i, a_i \rangle} a_i$$

成立，且

$$\text{Span}(S) = \text{Span}(S')$$



下面是一个利用Gram-Schmidt正交化的一个例子。

例题 3.4.

在三维空间 \mathbb{R}^3 中，有一组基 $\{a_1, a_2, a_3\}$ ，其中 $a_1 = (1, 1, 1), a_2 = (1, 1, 0), a_3 = (1, 0, 0)$ 请使用Gram-Schmidt正交化将其转化为 $S' = \{v_1, v_2, v_3\}$ ， S' 为 \mathbb{R}^3 的正交基。

解：根据Gram-Schmidt正交化依次可得

$$v_1 = a_1 = (1, 1, 1) \quad (3.1)$$

$$v_2 = a_2 - \frac{\langle v_1, a_2 \rangle}{\langle v_1, v_1 \rangle} v_1 = (0, 1, 1) - \frac{2}{3}(1, 1, 1) = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) \quad (3.2)$$

$$v_3 = a_3 - \frac{\langle v_1, a_3 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle v_2, a_3 \rangle}{\langle v_2, v_2 \rangle} v_2 \\ v_3 = (0, 0, 1) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{2/3} \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) = \left(0, -\frac{1}{2}, \frac{1}{2}\right) \quad (3.3)$$

思考

如果向量 S 线性相关，如果对其使用Gram-Schmidt正交化会发生什么？

答案是最后一个向量正交化后会变为**0**向量，它们确实两两正交，不过**0**向量不能成为基的一个。

读者不难发现，向量在正交化的时候也相当于一个线性映射，即原向量 v_i 被映射为了 a_i ，同时也不难发现这是一个从自己空间映射到自己空间的映射，我们给它一个名称叫做算子(operator)。

定义 3.9 (算子 (operator))

线性映射从 V 到 V 的集合 $\mathcal{L}(V, V)$ 中的元素称作 V 上的算子，我们把线性映射到本身的集合简写为 $\mathcal{L}(V)$ 。



例如线性映射， $T((x, y)) = (y, x)$ ， T 就是 \mathbb{R}^2 上的算子，因为映射前后的空间都是 \mathbb{R}^2 ，可以写作 $T \in \mathcal{L}(\mathbb{R}^2)$ 。

回到刚刚所说的原向量 v_i 被映射为了 a_i ，若映射法则 T 将 v_i 被映射为了 a_i ，如果我们将此线性映射表示为矩阵积，那么它应当满足：

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & v_3 & \cdots & v_n \end{pmatrix} \mathcal{M}(T)^9$$

⁸一般译作格拉姆-施密特，由两位数学家的名字构成，分别是丹麦数学家 Jørgen Pedersen Gram 和德国数学家 Erhard Schmidt

⁹这种写法用向量的矩阵表示，即若 $a_i, v_i \in \mathbb{R}^m$ 则这两个是 $m \times n$ 的矩阵

在正交化这样的过程中，我们可以得到 $\mathcal{M}(T)$ 是上三角矩阵，即

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & v_3 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} 1 & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & * \end{pmatrix}$$

3.2.4 正交投影

想象一下在 \mathbb{R}^3 中，有一条直线和一个平面垂直，在高中的学习生涯中我们知道在平面上任意取一个有向箭头，都与这条直线垂直，换句话说，平面上所有有向箭头表示的向量都与这条直线的方向向量正交。

所以我们引入一个正交补空间，表示该空间内向量对一个子空间中的任意一个向量正交，下面我们给出定义：

定义 3.10 (正交补空间)

设子空间 U 中的向量均与子空间 V 中的向量正交，我们称 V 是 U 的正交补空间，并有一个符号 U^\perp 表示；即

$$U^\perp := \{a \in V \mid \forall^a u \in U \implies \langle u, a \rangle = 0\}$$

^a 符号 \forall 表示对于所有的，这个集合的描述为， U^\perp 集合的元素 $a \in V$ 对于所有的 $u \in U$ 的使得 $\langle u, a \rangle = 0$ 成立

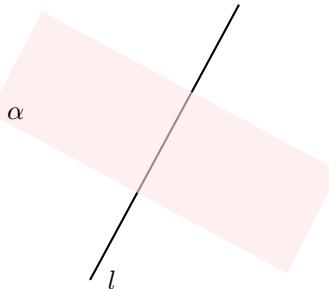


图 3.3: 正交补空间

如图3.3所示， \mathbb{R}^3 上满足 $l \perp \alpha$ 所以 l 和 α 表示的线性空间相互正交补，即 $\alpha^\perp = l, l^\perp = \alpha$ ，由此我们引入正交投影的概念，并解决一些极小化问题。

定义 3.11 (正交投影 (orthogonal projection))

设 V 的一个有限维子空间为 U ，定义 V 到 U 上的正交投影为 $P_U \in \mathcal{L}(V)$ 为对所有的 $v \in V$ 均可写成 $v = u + w$ ，其中 $u \in U$ 且 $w \in U^\perp$ ，则 $P_U(v) = w$ ，向量 w 称为 v 在 U 上的投影。



结合我们之前讲过的线性映射的零空间和值域空间，便有如下推论。

推论 3.6

若 U 是 V 的有限维子空间，则 $P_U \in \mathcal{L}(V)$ 满足

$$\text{range } P_U = (\text{null } P_U)^\perp = U$$



3.3 极小化问题

3.3.1 到子空间的最短距离

还是在高中的学习生涯中我们遇到过在空间 \mathbb{R}^3 上给定一个点，并求它到直线或平面的最短距离，为了后续抽象化的操作，我们先来定义什么是点。

定义 3.12 (点 (point) 与点的距离 (point distance))

子空间 V 上的点 $v \in V$ 是一个向量，在图形中通常表示为从 $\mathbf{0}$ 开始的有向箭头。若两个不相同的点 $v, w \in V$ 则它们的距离表示为

$$d(v, w) := \|v - w\|^a$$

^a在这一部分，我们都用欧几里得体系表示空间



例如在 \mathbb{R}^2 中点 $v = (3, 4)$ 和点 $w = (0, 1)$ 的距离为 $d(v, w) = \sqrt{3^2 + 3^2} = 3\sqrt{2}$ ，根据定义我们可以把上面的问题描述成：给定一个空间 V 的子空间 U 和点 $v \in V$ ，求点 $u \in U$ 使得 $\|v - u\|$ 最小；在解决该问题的策略中，我们知道点到直线或平面的距离最短的情况应该是垂直到点的情形，所以到子空间的最小距离应当是到该子空间正交补空间上取点，下面的一个命题表示将点通过 P_U 映射到子空间后点的距离最短。

定理 3.7 (到子空间的最短距离)

设 U 为 V 有限维子空间，有 $v \in V$ 且 $u \in U$ 那么

$$\|v - P_U(v)\| \leq \|v - u\|$$

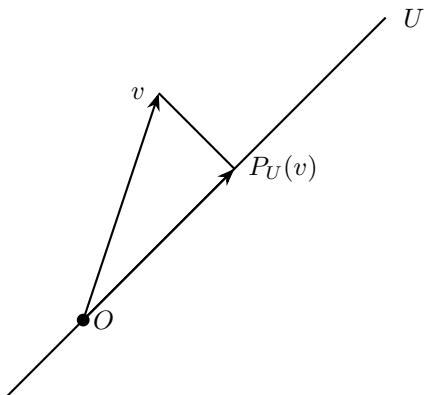


图 3.4: $P_U(v)$ 是 v 的最短距离

如图3.4所示，将 v 正交投影到 U 上的 P_U 点，它们距离最短。

3.3.2 多项式与最小化

首先我们给出两个函数内积的定义。

定义 3.13 (函数的内积)

定义区间 $S = [a, b]$ 上的连续实值函数构成的的向量可定义内积如下：

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$



下面将以此介绍一个最小化方法，求解一些函数的多项式近似。在之前微积分的学习中，我们学到了麦克劳林近似，它在 0 处对函数有着极好的逼近。

定理 3.8 (麦克劳林近似)

函数 $f(x)$ 的 n 阶导数为 $f^{(n)}(x)$ 满足公式

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1} (0 < \theta < 1)$$



下面我们新介绍一个方法叫做 Legendre 多项式，利用正交的极小化完成对函数的多项式近似，为了形成对比，我们先使用泰勒近似取最高次数项为 5 次的结果

$$\sin x \sim x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

根据 Python-Matplotlib 工具我们绘制了这样的图：

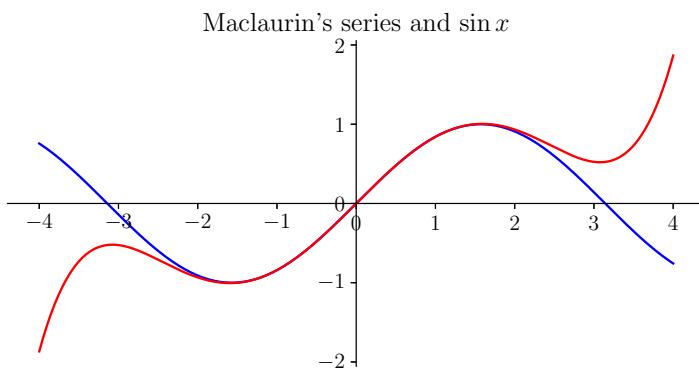


图 3.5：蓝色为 $\sin x$ 的图像，红色为麦克劳林近似多项式

接下来使用正交最小化的方式来取近似，首先我们令多项式为函数 $g(x)$ 设函数 $f(x) = \sin x$ 与函数 $g(x)$ 在 $[-\pi, \pi]$ 的内积表示为

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx$$

下面选取 $u(x)$ 的基 $\{1, x, x^2, x^3, x^4, x^5\}$ 对其使用 Gram-Schmidt 正交化，设 $\{v_0, v_1, v_2, v_3, v_4, v_5\}$ 为正交化过程后的基，我们有：

$$v_0 = 1 \quad (3.4)$$

$$v_1 = x - \frac{\langle x, v_0 \rangle}{\langle v_0, v_0 \rangle} v_0 = x \quad (3.5)$$

$$v_2 = x^2 - \frac{\langle x^2, v_0 \rangle}{\langle v_0, v_0 \rangle} v_0 - \frac{\langle x^2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \quad (3.6)$$

...

重复采用函数内积公式将对空间 $U = \{v_0, v_1, v_2, v_3, v_4, v_5\}$ 和函数空间 $f(x) = \sin x$ 计算 $P_U(f)$ ，计算 $P_U(f)$ 表示函数空间

$$g(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5$$

如果你使用的是电子介质，你可以放大并看到它们拟合的几乎严丝合缝，相较于麦克劳林近似，这种近似法在 3 处仅有微小差别。而麦克劳林近似由于采用的在 0 处展开，所以越靠近 0，拟合的效果越好，下面给出一张表来观察两种近似的对比：

如图表所示，麦克劳林近似在越靠近 0 处的拟合效果更好，而纵观全局，Legendre 多项式近似法在 $[-\pi, \pi]$

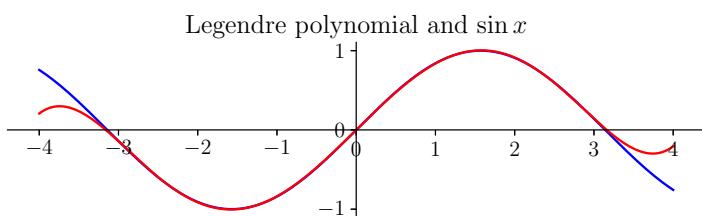
表 3.1: Legendre 多项式近似与麦克劳林近似的数值比较

	$\sin x$	Legendre 多项式近似	麦克劳林近似
$x = 0$	0.0	0.0	0.0
$x = 1/3$	0.3271946967961522	0.3235597782716049	0.3271947873799726
$x = 2/3$	0.618369803069737	0.6133115713580247	0.6183813443072702
$x = 1$	0.8414709848078965	0.8382341200000001	0.8416666666666667
$x = 4/3$	0.9719379013633127	0.9728796167901235	0.9733882030178327
$x = 5/3$	0.9954079577517649	1.0001604320987654	1.0022290809327847
$x = 2$	0.9092974268256817	0.9141358400000001	0.9333333333333333
$x = 7/3$	0.7230858817383246	0.7227987441975305	0.7924211248285326
$x = 8/3$	0.457272626635812	0.4508624039506175	0.6299039780521267
$x = 3$	0.1411200080598672	0.14254716000000012	0.5249999999999999

上拟合效果更好。正交投影最小化帮我们找到了一个方法，改进了我们在微积分学习中学到的近似。实际上，如果我们给定一个函数如下

$$\mathcal{F}(x) = \int_{-\pi}^{\pi} |\sin x - u(x)|^2 dx$$

这个函数表示在拟合过程中，多项式函数与被拟合函数相减并平方围成的面积，这个值越小，在这种情况下拟合效果越好，实际上，使用 Legendre 多项式近似法要比麦克劳林近似，该值要小得多。

图 3.6: 蓝色为 $\sin x$ 的图像，红色为 Legendre 多项式近似法

3.4 章节练习

3.4.1 A 组

练习 26 设 V 为实内积空间，尝试证明极化恒等式成立，即对 $v, w \in V$ 有

$$\langle v, w \rangle = \frac{\|v + w\|^2 - \|v - w\|^2}{4}$$

练习 27 设 $v, w \in V$ ，若 $\|v\| = 3, \|v + w\| = 4, \|v - w\| = 6$ 求 $\|w\|$ 的值。

练习 28 请自行选取一个关于 \mathbb{R}^4 的非正交基，使用 Gram-Schmidt 正交化将其转化为正交基。

练习 29 证明，菱形的对角线相互垂直，即对角线表示的有向箭头正交。

练习 30 比较大小：设数列 $\{a_n\}$ ，比较它们的平均数的平方的和与它们和的平方的平均数的大小，即

$$\frac{1}{n} \left(\sum_{i=0}^n a_i^2 \right) \quad \text{and} \quad \left(\frac{1}{n} \left(\sum_{i=1}^n a_i \right) \right)^2$$

练习 31 若基 $S = \{(1, 2, 3, 4), (9, 8, 7, 6)\}$ 其张成空间为 $\text{Span}(S) = U$ 为 \mathbb{R}^4 的子空间，求 U 的一个正交基和 U^\perp 的一个正交基。

练习 32 若基 $S = \{(1, 1, 0, 0), (1, 1, 1, 2)\}$ 其张成空间为 $\text{Span}(S) = U$ 为 \mathbb{R}^4 的子空间，求 $u \in U$ 的使得 $\|u - (1, 2, 3, 4)\|$ 最小。

3.4.2 B 组

练习 33 多项式 $P(x)$ 的最高次数为 2，在此之上考虑内积

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx$$

对基 $\{1, x, x^2\}$ 使用 Gram-Schmidt 正交化求 $P(x)$ 的一个正交基。

练习 34 请使用 Legendre 多项式与正交化，对 $\ln x$ 在 $[1, 2]$ 上进行最高次数为 5 的多项式拟合。

第二部分

基础部分

第 2 部分目录

第 4 章 行列式	50
4.1 面积与体积	50
4.2 n 阶行列式	53
4.3 行列式按行展开	53
4.4 行列式的性质	56
4.5 置换	60
4.6 章节练习	62
第 5 章 初等矩阵	63
5.1 线性方程组的解	63
5.2 初等矩阵	65
5.3 矩阵的逆	69
5.4 矩阵的秩	72
5.5 线性方程组的解空间	75
5.6 章节练习	80
第 6 章 特征值理论	82
6.1 特征值与特征向量	82
6.2 迹与上三角矩阵	86
6.3 相似变换	89
6.4 Jordan 标准型	94
6.5 章节练习	103



第 4 章 行列式

「鸟下绿芜秦苑夕，蝉鸣黄叶汉宫秋」

——《咸阳城东楼》



4.1 面积与体积

4.1.1 面积

考虑在如图4.1所示平面直角坐标系中 $u, v \in \mathbb{R}^2$, 有向箭头 $\overrightarrow{OA} = v = (v_1, v_2)$, $\overrightarrow{OB} = u = (u_1, u_2)$, $\overrightarrow{OC} = u + v$, 求平行四边形 $OACB$ 的面积。

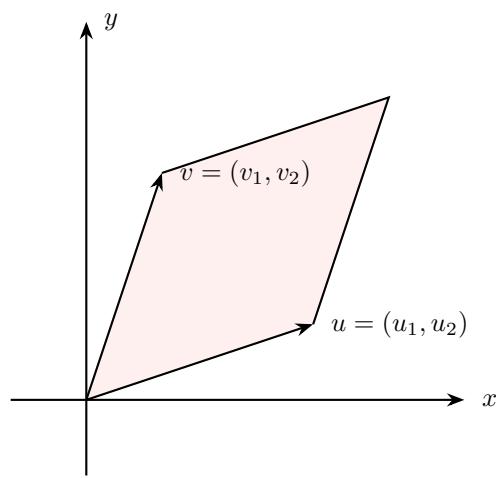


图 4.1: 平行四边形

在小时候我们没有学习过高级方法的时候, 我们通常会使用割补法来计算它们的面积, 通常会如下作图

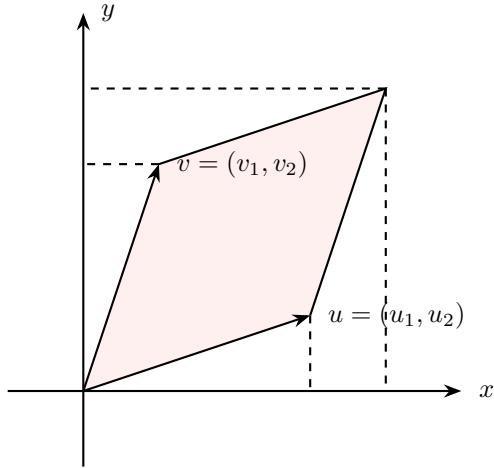


图 4.2: 割补法求面积

设平行四边形面积为 S 那么

$$S = (v_1 + u_1)(v_2 + u_2) - 2u_2v_1 - u_1u_2 - v_1v_2 = u_1v_2 - u_2v_1^1$$

所以我们可以快速地写出它们围成的平行四边形面积，即在平面直角坐标系内两个向量 $u = (u_1, u_2)$ 和向量 $v = (v_1, v_2)$ 围成的平行四边形面积 S 为

$$S = |u_1v_2 - u_2v_1|$$

例题 4.1.

求向量 $u = (3, 1)$ 和向量 $v = (1, 2)$ 围成的平行四边形面积 S 。

解： $S = 3 \times 2 - 1 \times 1 = 5$

我们引入一个记号来表达这两个向量围成的平行四边形面积，即

$$\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} := u_1v_2 - u_2v_1$$

也称为 2×2 矩阵的行列式，该行列式的值的绝对值为平行四边形面积。

我们注意到该行列式由两个向量构成，如果我们尝试交换行列式的第一行和第二行，也就是交换一下围成这个四边形的向量的顺序，我们可以得到

$$\begin{vmatrix} v_1 & v_2 \\ u_1 & u_2 \end{vmatrix} := u_2v_1 - u_1v_2 = -\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$$

读者可以将其理解为这个面积是一个带符号的面积，这个符号和两个向量顺序有关。通过观察发现，如果第二个向量在第一个向量的顺时针方向上的话，其值为正。

如果我们将某个向量放缩 λ 倍那么面积也会进行相对应的伸缩，即

$$\lambda \begin{vmatrix} v_1 & v_2 \\ u_1 & u_2 \end{vmatrix} = \begin{vmatrix} \lambda v_1 & \lambda v_2 \\ u_1 & u_2 \end{vmatrix} = \begin{vmatrix} v_1 & v_2 \\ \lambda u_1 & \lambda u_2 \end{vmatrix}$$

¹向量 u 在向量 v 的顺时针方向上时为正否则为负

4.1.2 体积

考虑在三维空间 \mathbb{R}^3 中，一个直的四棱柱的面积计算公式为 $S_{\text{底面积}} \times h$ ，考虑由图4.1作为空间直角坐标系的底面，从原点长出一个 z 轴后如图4.3所示。

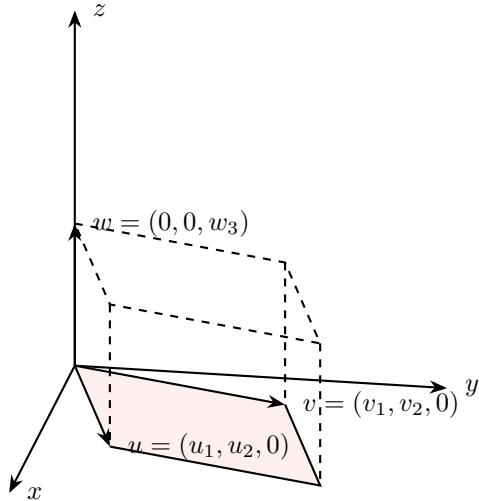


图 4.3: 三维体积行列式表示

所以如果我们在 z 上取一个向量围成一个四棱柱，那么它的体积就是 $S_{\text{底面积}} \times h$ 计算如图4.3的四棱柱体积为 $w_3(u_1v_2 - u_2v_1)$ ；下面引入三阶行列式，分别写三个向量来表示这个体积，即

$$\begin{vmatrix} u_1 & u_2 & 0 \\ v_1 & v_2 & 0 \\ 0 & 0 & w_3 \end{vmatrix} = w_3(u_1v_2 - u_2v_1)$$

经过割补法验证，对于 $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3), w = (w_1, w_2, w_3) \in \mathbb{R}^3$ 其张成的四棱柱体积计算公式为

$$S = |u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)|$$

故我们定义一个三阶行列式，表示由这三个向量构成的立方体的体积，即

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} := u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)$$

4.1.3 2 或 3 阶行列式的性质

前面我们定义了 2, 3 阶的行列式，通过几何意义我们不难发现如下性质。

推论 4.1

2 阶, 3 阶行列式有如下性质

1. 行列式某几行所表示的向量线性相关，行列式值为 0。
2. 行列式某行向量乘以标量 λ ，行列式值也乘 λ 。



我们通过图形的角度证明两个性质，第一个如果某几行向量线性相关说明三维图形会坍缩为一个平面，平面会坍缩为直线，所以它们对应的体积或面积均为 0。第二个如果某行向量乘以标量 λ 那么它们的测度²也会放缩为 λ 倍。

²这里的测度指的是，当描述的是一维空间其测度就是长度，二维测度为面积，三维测度为体积

4.2 n 阶行列式

4.2.1 n 阶行列式的定义

首先我们定义一下什么叫作 n 阶方阵。

定义 4.1 (n 阶方阵)

若矩阵 \mathbf{A} 是一个 $m \times n$ 矩阵，当 $m = n$ 时，我们称 \mathbf{A} 是 n 阶方阵

例如 $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ 表示为 n 阶方阵，我们以此定义一个数表示该方阵的行列式。

定义 4.2 (n 阶行列式)

定义 n 阶方阵 \mathbf{A} 的行列式为 n 阶行列式，记作 $\det \mathbf{A}$ ，也可以记作

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

前面说到了 2 阶，3 阶行列式，我们来讨论更一般的情况，在 n 维空间上的测度。如果我们把 \mathbb{R}^n 空间上的标准正交基 $\{(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), (0, 0, 1, \dots, 0), \dots, (0, 0, 0, \dots, 1)\}$ 放入行列式内，我们就可以得到

$$\begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = 1$$

实际上，行列式的几何意义就是 n 向量张成的一个 n 维“图形”的测度。后面我们会以行展开的方式，来讲解它的运算³。

4.2.2 2 阶与 3 阶行列式的运算

首先对于 2 阶行列式，计算方式肯定是：主对角线的乘积减去副对角线的乘积；对于三阶行列式，我们有如图4.4的计算记忆方法，这种方法称为 Sarrus⁴ 法。

需要注意的是只有二阶和三阶行列式具有 Sarrus 法则，四阶及以上的行列式不存在 Sarrus 法。

4.3 行列式按行展开

4.3.1 代数余子式

对于更高维度的行列式，我们通常使用按行展开的方式定义它们的运算；在此之前，我们先了解一下什么是代数余子式。

³注意，这本书不会使用逆序排列法详细讲行列式的计算，虽然我们定义它的运算是这样的，但是这不会让我们更加理解行列式，且前面用几何法定义了它的含义，所以我们在后面的讲解中会直接讲运算方法

⁴一般译作：萨鲁法

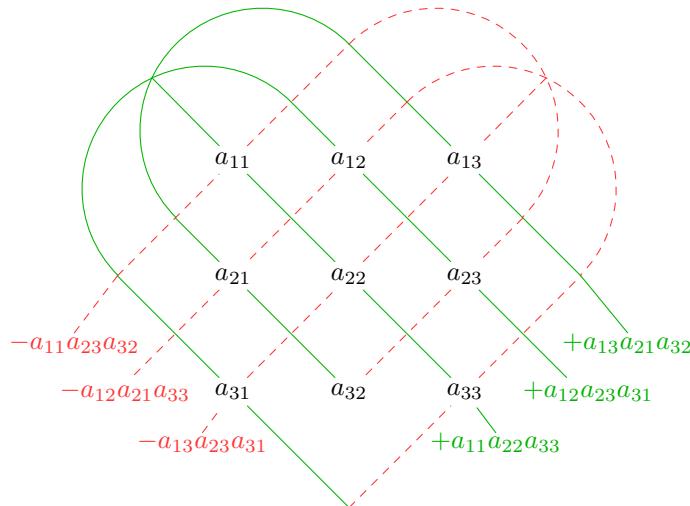


图 4.4: Sarrus 法求 3 阶行列式

定义 4.3 (余子式与代数余子式)

若 n 阶行列式的第 i 行第 j 列的余子式为 M_{ij} , 表示将该行与该列所有元素去除后剩下的元素将其合成 $n-1$ 阶行列式。而代数余子式 A_{ij} 定义为

$$A_{ij} := (-1)^{i+j} M_{ij}$$

我们用一张图来说明什么是代数余子式, 若 n 阶行列式 \mathbf{D} 有

$$\mathbf{D} = \begin{vmatrix} 1 & 3 & 4 & 10 \\ 2 & 5 & 9 & 11 \\ 6 & 8 & 12 & 15 \\ 7 & 13 & 14 & 16 \end{vmatrix}$$

则它的 a_{23} 的余子式 M_{23} 为去除第二行和第三列的所有元素, 留下来的成为一个 3 阶行列式, 即

$$\begin{vmatrix} 1 & 3 & 4 & 10 \\ 2 & 5 & 9 & 11 \\ 6 & 8 & 12 & 15 \\ 7 & 13 & 14 & 16 \end{vmatrix} \Rightarrow M_{23} = \begin{vmatrix} 1 & 3 & 10 \\ 6 & 8 & 15 \\ 7 & 13 & 16 \end{vmatrix}, A_{23} = (-1)^{2+3} M_{23} = -M_{23}$$

4.3.2 行列式按行展开

介绍完余子式后, 我们使用 Laplace⁵ 定理来计算行列式, 在此之前我们先观察几何定义的三阶行列式, 如果使用割补法来计算相关体积, 我们有一个重要的推论, 即

$$\text{体积} = \text{底面积} \times \text{高}$$

按照行列式的几何定义, n 阶行列式的值定义为在 n 维空间上这些向量围成的 n 维测度, 所以我们把式子抽象化, 即

$$n\text{维测度} = (n-1)\text{维测度} \times \text{高}$$

⁵一般译作拉普拉斯

仿照体积的算法

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} := u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)$$

我们逐步建立高维空间，有 Laplace 定理如下

定理 4.1 (Laplace 定理 (第一行))

设行列式 \mathbf{D} 取第一行^a所有元素 $a_{11}, a_{12}, a_{13}, \dots, a_{1n}$ 以及它们的代数余子式 $A_{11}, A_{12}, A_{13}, \dots, A_{1n}$ ，有

$$\mathbf{D} = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n}$$

写成求和符号为

$$\mathbf{D} = \sum_{i=1}^n a_{1i}A_{1i}$$

^a由于我们没有讲到行列式的性质，这里暂时取第一行，实际上按照后面的行列式的性质，我们可以取任意一行或任意一列，此外，Laplace 定理允许我们选择多行多列展开



下面我们来看一些例子。

例题 4.2.

使用 Laplace 定理计算行列式

$$\mathbf{D} = \begin{vmatrix} 1 & 2 & 0 & 3 \\ 2 & 0 & 1 & 1 \\ 1 & 5 & 3 & 6 \\ 0 & -1 & 2 & 4 \end{vmatrix}$$

解：按第一行展开行列式为

$$\mathbf{D} = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} + a_{14}A_{14}$$

其中 $a_{11} = 1, a_{12} = 2, a_{13} = 0, a_{14} = 3$ 代数余子式分别为

$$A_{11} = \begin{vmatrix} 0 & 1 & 1 \\ 5 & 3 & 6 \\ -1 & 2 & 4 \end{vmatrix} = -13 \quad (4.1)$$

$$A_{12} = -\begin{vmatrix} 2 & 1 & 1 \\ 1 & 3 & 6 \\ 0 & 2 & 4 \end{vmatrix} = 2 \quad (4.2)$$

$$A_{13} = \begin{vmatrix} 2 & 0 & 1 \\ 1 & 5 & 6 \\ 0 & -1 & 4 \end{vmatrix} = 51 \quad (4.3)$$

$$A_{14} = -\begin{vmatrix} 2 & 0 & 1 \\ 1 & 5 & 3 \\ 0 & -1 & 2 \end{vmatrix} = -25 \quad (4.4)$$

其中三阶行列式可使用 Sarrus 法解决，也可以继续按行展开，我们得到

$$\mathbf{D} = -13 \times 1 + 2 \times 2 + 51 \times 0 + (-25) \times 3 = -84$$

不过在实际的做题中还有比这种方法更加简洁的方法，这里仅作演示，运算技巧会在后续的行列式的性质



详细讲述。

4.4 行列式的性质

4.4.1 转置

行列式转置表示把行列进行调换，行变成列，列变成行。

$$\mathbf{D} = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} \implies \mathbf{D}^T = \begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix}$$

推论 4.2

行列式转置的转置等于其本身，即 $(\mathbf{D}^T)^T = \mathbf{D}$



这个推论是显然的。

推论 4.3

行列式转置的值等于其本身的值，即 $\mathbf{D}^T = \mathbf{D}$



转置不变性说明在行列式中行和列的地位是平等的。因此行列式关于行变换的性质对于列变换也成立，此处不作证明，读者可以自己举例验证。

4.4.2 其他性质

推论 4.4

把行列式的某一行乘以一个系数，则它的行列式的值也乘以同一个系数。



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda a_{k1} & \lambda a_{k2} & \lambda a_{k3} & \cdots & \lambda a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

推论 4.5

行列式任意两行(列)交换，行列式的值变为原来的相反数。



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$



推论 4.6

行列式中的一行是另一行的倍数，行列式值为 0；或构成行列式的向量线性相关，行列式值为 0。



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda a_{k1} & \lambda a_{k2} & \lambda a_{k3} & \cdots & \lambda a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = 0$$

推论 4.7

行列式的一行具有线性性。



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} + b_{k1} & a_{k2} + b_{k2} & a_{k3} + b_{k3} & \cdots & a_{kn} + b_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{k1} & b_{k2} & b_{k3} & \cdots & b_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

推论 4.8

把某一行乘以一个数加到另一行，行列式值不变。



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{k1} + a_{j1} & a_{k2} + a_{j2} & a_{k3} + a_{j3} & \cdots & a_{kn} + a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

推论 4.9

上三角方阵的行列式是对角线元素的乘积。



$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22} \cdots a_{nn}$$



推论 4.10

设 n 阶方阵 \mathbf{A}, \mathbf{B} 的行列式为 $\det \mathbf{A}, \det \mathbf{B}$ 有 $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$



4.4.3 初等变换计算行列式

我们可以利用上述的性质来计算行列式，上述性质的变换我们称为行列式的初等变换，下面是常用的几种初等变换：

1. 将行列式的两行交换位置
2. 将其中的一行乘一个非 0 常数
3. 将其中的一行乘一个非 0 常数后加到另一行

它们分别对应

1. 行列式的值变为原来的相反数
2. 行列式乘以同一个非零常数
3. 保持不变

例题 4.3.

计算行列式 $\mathbf{D} = \begin{vmatrix} 1 & 2 & 1 \\ 0 & -3 & 4 \\ 1 & 0 & 5 \end{vmatrix}$

解：将第 1 行乘 (-1) 加到第 3 行有

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & -3 & 4 \\ 1 & 0 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 1 \\ 0 & -3 & 4 \\ 0 & -2 & 4 \end{vmatrix}$$

交换第 2 列与第 3 列

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & -3 & 4 \\ 0 & -2 & 4 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 2 \\ 0 & 4 & -3 \\ 0 & 4 & -2 \end{vmatrix}$$

第 2 行乘 (-1) 加到第 3 行

$$-\begin{vmatrix} 1 & 1 & 2 \\ 0 & 4 & -3 \\ 0 & 4 & -2 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 2 \\ 0 & 4 & -3 \\ 0 & 0 & 1 \end{vmatrix} = -(1 \times 4 \times 1) = -4$$

再来看一个 4 阶行列式的例子。

例题 4.4.

计算行列式 $\mathbf{D} = \begin{vmatrix} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 3 & -1 & 5 & 0 \\ -2 & 0 & 1 & 1 \end{vmatrix}$



解：将第1行乘 (-3) 加到第3行有

$$\left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 3 & -1 & 5 & 0 \\ -2 & 0 & 1 & 1 \end{array} \right| = \left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -7 & 5 & -12 \\ -2 & 0 & 1 & 1 \end{array} \right|$$

将第1行乘2加到第4行有

$$\left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -7 & 5 & -12 \\ -2 & 0 & 1 & 1 \end{array} \right| = \left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -7 & 5 & -12 \\ 0 & 4 & 1 & 9 \end{array} \right|$$

将第2行乘2加到第3行；第2行乘 (-1) 加到第4行有

$$\left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -7 & 5 & -12 \\ 0 & 4 & 1 & 9 \end{array} \right| = \left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -1 & 3 & -10 \\ 0 & 1 & 2 & 8 \end{array} \right|$$

将第4行加到第3行有

$$\left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & -1 & 3 & -10 \\ 0 & 1 & 2 & 8 \end{array} \right| = \left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 5 & -2 \\ 0 & 1 & 2 & 8 \end{array} \right|$$

运算到这里，为了计算不出错，我们最好遵循一个原则，就是行列式内的元素尽量不要出现分数，否则计算会比较麻烦，为了转化为上三角行列式我们先对第1行第2列的元素凑为1，即：将第4行乘 (-1) 加到第一行，并提取系数4

$$\left| \begin{array}{cccc} 1 & 2 & 0 & 4 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 5 & -6 \\ 0 & 1 & 2 & 8 \end{array} \right| = \left| \begin{array}{cccc} 1 & 1 & -2 & -4 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 5 & -2 \\ 0 & 1 & 2 & 8 \end{array} \right|$$

再将第4行乘 (-3) 加到第2行后交换2, 4行有

$$\left| \begin{array}{cccc} 1 & 1 & -2 & -4 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 5 & -2 \\ 0 & 1 & 2 & 8 \end{array} \right| = \left| \begin{array}{cccc} 1 & 1 & -2 & -4 \\ 0 & 0 & -7 & -23 \\ 0 & 0 & 5 & -2 \\ 0 & 1 & 2 & 8 \end{array} \right| = - \left| \begin{array}{cccc} 1 & 1 & -2 & -4 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 5 & -2 \\ 0 & 0 & -7 & -23 \end{array} \right|$$

算到这里同学们可以继续将其化为上三角，但是我们选择按列展开两次便捷，即

$$-\left| \begin{array}{cccc} 1 & 1 & -2 & -4 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 5 & -2 \\ 0 & 0 & -7 & -23 \end{array} \right| = -1 \times 1 \times \left| \begin{array}{cc} 5 & -2 \\ -7 & -23 \end{array} \right|$$

最后求得行列式值为

$$D = - \left| \begin{array}{cc} 5 & -2 \\ -7 & -23 \end{array} \right| = 129$$

4.5 置换⁶

4.5.1 全排列与置换

本节内容选学。

如果有 n 个元素进行排列，那么将其全部有序排列的方案一共有 $n!$ 种排列方式。如果我们将 n 个元素的初始排列集合称作初始集合，那么我们可以看作一个函数映射，将这个集合映射为 $n! - 1$ 中的一个，我们称该函数为置换函数；

定义 4.4 (置换函数)

若集合 $V = \{1, 2, 3, \dots, n\}$ ，设映射法则 $\sigma : V \rightarrow V$ 的一一映射称为一个 n 元置换。函数 $f(x)$ 称为置换函数。



例如置换映射 $\sigma : \{1, 2, 3, 4, 5\} \rightarrow \{3, 1, 4, 2, 5\}$ ，那么函数 $\sigma(1) = 3, \sigma(2) = 1, \dots$ 。

4.5.2 有向的 n 维测度

首先我们还是看二维平面，其二维测度为面积，我们不难发现当向量 $u = (u_1, u_2), v = (v_1, v_2)$ ，当 v 在 u 的顺时针方向上的时候面积为正，否则为负。

由此我们不妨以 n 维空间的标准基定义 n 维上向量的“旋转”方向，以此为基准形成排列，则如果有两个向量进行一次交换，那么它们进行一次置换后符号方向改变。如果上面的话有些抽象，我们不妨再拿二维空间举例；在二维空间中的标准基为 $\{(1, 0), (0, 1)\}$ ，其中如果有两个向量符合标准基的第一个和标准基的第二个的位置关系（第二个向量在第一个的顺时针方向），那么行列式值为正，否则为负。推广到高维空间中也是如此。

那么接下来我们继续探讨置换这些向量对行列式的影响，在此之前我们先对 σ 函数引入一个记号，表示 σ 置换了多少次。

4.5.3 最少置换次数与 $\text{sign}(\sigma)$

首先定义最少置换次数函数。

定义 4.5 (最少置换次数)

最少置换次数对置换映射法则 σ 的函数 $R : \sigma \rightarrow \mathbb{N}$ ，表示对使用映射法则 σ 之前的集合至少需要多少次置换以达到使用映射法则 σ 之后的集合^a。

^a需要注意的是，这集合无序，但是映射后的集合却是一一对应，这里的集合可以看作是一个有序的数组，虽然集合讲究无序性



下面我们来看一个例子，置换映射 $\sigma : \{1, 2, 3, 4, 5\} \rightarrow \{3, 1, 4, 2, 5\}$ ，那么它的最少置换次数为 3 次，置换如下：首先初始状态为

$$3, 1, 4, 2, 5$$

第一次选择 3, 1 调换顺序

$$1, 3, 4, 2, 5$$

接着第二次选择 3, 4 调换

$$1, 4, 3, 2, 5$$

最后第三次选择 2, 4 调换顺序

$$1, 2, 3, 4, 5$$

⁶本节选学



所以

$$\mathcal{R}(\sigma) = 3$$

事实上如果我们去掉最少二字的限定，置换次数可以是 $3 + 2k, k \in \mathbb{N}$ 。

接下来我们继续定义函数 $\text{sign}(\sigma)$ ，它表示的是映射 σ 的奇偶性；

定义 4.6 (奇偶置换)

若式子 $\mathcal{R}(\sigma)$ 的值如果为偶数，那么该置换为偶置换，否则为奇置换，另有函数 $\text{sign}(\sigma)$ 定义为

$$\text{sign}(\sigma) := (-1)^{\mathcal{R}(\sigma)}$$



4.5.4 行列式的排列定义

还是这个例子：置换映射 $\sigma : \{1, 2, 3, 4, 5\} \rightarrow \{3, 1, 4, 2, 5\}$ ，如果我们把它们的空间的标准基标号作为映射集合有 $\sigma' : \{e_1, e_2, e_3, e_4, e_5\} \rightarrow \{e_3, e_1, e_4, e_2, e_5\}$ ，相当于在 \mathbb{R}^5 空间中定义标准基的方向是这 5 维测度的方向，每个向量如果按这样的顺序（相当于二维空间的顺时针）排列的话，其测度为正；不过，我们对其进行至少 3 次的置换，所以其测度应当为 $\text{sign}(\sigma) = -1$ 。

如果我们类比 2, 3 阶行列式的推导过程，即

$$n \text{维测度} = (n-1) \text{维测度} \times \text{高}$$

通过更高维的割补法，我们可以如下定义一个行列式：

定义 4.7 (行列式的排列定义)

设 \mathbf{A} 是 n 阶方阵，设其第 i 行第 j 列的元素为 a_{ij} 那么有

$$\det \mathbf{A} = \sum_{\sigma \in S_n} \text{sign}(\sigma) \prod_{i=1}^n a_{(\sigma(i))i}$$

其中 S_n 表示所有 n 元置换的集合，即共 $n!$ 个的 $\{1, 2, 3, \dots, n\}$ 的全排列集合。



下面举个例子，3 阶行列式 $\mathbf{D} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ ，它们的全排列共有 $3! = 6$ 种，分别映射为

$$\{\{1, 2, 3\}, \{1, 3, 2\}, \{2, 1, 3\}, \{2, 3, 1\}, \{3, 1, 2\}, \{3, 2, 1\}\}$$

那么对于每一个从集合 $\{1, 2, 3\}$ 置换开始求和，首先是 $\sigma : \{1, 2, 3\} \rightarrow \{1, 2, 3\}$ ，则

$$\text{sign}(\sigma)a_{11}a_{22}a_{33} = +a_{11}a_{22}a_{33}$$

接下来是 $\sigma : \{1, 2, 3\} \rightarrow \{1, 3, 2\}$ 则

$$\text{sign}(\sigma)a_{11}a_{32}a_{23} = -a_{11}a_{32}a_{23}$$

...

最终进行 6 次这样的操作后，我们得到多项式

$$a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} \dots$$

当然也可以按顺序从每行找不同列的元素相乘，结果都是一样的。



4.6 章节练习

4.6.1 A组

练习 35

$$\text{计算行列式 } \mathbf{D} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & -2 & 1 \\ 6 & 1 & 0 \end{vmatrix}$$

练习 36

$$\text{计算行列式 } \mathbf{D} = \begin{vmatrix} 1 & 1 & 2 & 2 & 0 \\ 0 & 2 & -3 & -1 & 1 \\ 1 & 1 & 3 & 1 & 1 \\ 2 & -4 & 0 & 0 & 3 \\ 1 & 7 & -1 & 0 & 2 \end{vmatrix}$$

练习 37

$$\text{计算行列式 } \mathbf{D} = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{vmatrix}$$

练习 38

$$\text{矩阵 } \mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 1 & 1 \\ 0 & 2 & 0 \end{pmatrix}, \text{ 求 } \det(3\mathbf{A}) \text{ 和 } \det(\mathbf{A}^2)$$

4.6.2 B组

练习 39 设数列 $\{F_n\}$ 为 Fibonacci 数列, 其中 $F_1 = 0, F_2 = 1, F_{n-1} + F_{n-2} = F_n, n \geq 2$ 证明

$$F_{n+1}F_{n-1} - F_n^2 = (-1)^n$$



第 5 章 初等矩阵

「只愿君心似我心，定不负相思意」

——《卜算子·我住长江头》

5.1 线性方程组的解

5.1.1 二元线性方程组

考虑一般的二元线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \quad (5.1)$$

(5.2)

按照正常的方法，我们将式5.1乘系数 a_{21} ，将5.2乘系数 a_{11} 后，用变换过的5.1减去5.2后我们可以得到

$$(a_{11}a_{22} - a_{12}a_{21})x_1 = b_1a_{22} - b_2a_{12}$$

我们可以得到方程的解

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

通过上述方法，我们可以得到

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

5.1.2 Cramer 法则

以上面的二元线性方程组为例我们抽象出的两个解的表示方法，其中我们有系数矩阵行列式

$$\mathbf{D} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

其中常数矩阵 $\mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, 我们将常数矩阵替换系数矩阵的第 i 列, 得到 \mathbf{D}_i 行列式, 例如

$$\mathbf{D}_1 = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}, \mathbf{D}_2 = \begin{vmatrix} a_{11} & b_1 \\ a_{12} & b_2 \end{vmatrix}$$

最后方程组的第 i 个解为

$$x_i = \frac{\mathbf{D}_i}{\mathbf{D}}$$

例题 5.1.

使用 Cramer 法则计算线性方程组

$$\begin{cases} 2x + 3y + 4z = 3 \\ 5x + 4z = 2 \\ x + 3y + z = 1 \end{cases}$$

解: 考虑系数矩阵行列式

$$\mathbf{D} = \begin{vmatrix} 2 & 3 & 4 \\ 5 & 0 & 4 \\ 1 & 3 & 1 \end{vmatrix} = 33$$

常数矩阵

$$\mathbf{B} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

使用常数矩阵替换行列式的第 i 列, 我们可以得到

$$\mathbf{D}_1 = \begin{vmatrix} 3 & 3 & 4 \\ 2 & 0 & 4 \\ 1 & 3 & 1 \end{vmatrix} = -6, \mathbf{D}_2 = \begin{vmatrix} 2 & 3 & 4 \\ 5 & 2 & 4 \\ 1 & 1 & 1 \end{vmatrix} = 5, \mathbf{D}_3 = \begin{vmatrix} 2 & 3 & 3 \\ 5 & 0 & 2 \\ 1 & 3 & 1 \end{vmatrix} = 24$$

那么该方程组的三个解分别为

$$x_1 = \frac{\mathbf{D}_1}{\mathbf{D}} = -\frac{2}{11}, x_2 = \frac{\mathbf{D}_2}{\mathbf{D}} = \frac{5}{33}, x_3 = \frac{\mathbf{D}_3}{\mathbf{D}} = -\frac{8}{11}$$

思考

当系数矩阵的行列式 $\mathbf{D} = 0$ 时是什么情况?

Cramer 法则的适用前提是系数矩阵的行列式值不能为 0, 读者可以发现若系数矩阵的行列式为 0 的时候会有一组方程组线性相关, 其所面临的情况为有无穷多解或无解, 例如

$$\begin{cases} 2x_1 + 3x_2 = 4 \\ 4x_1 + 6x_2 = 8 \end{cases}$$

这种情况就是无穷多解, 而

$$\begin{cases} 2x_1 + 3x_2 = 4 \\ 4x_1 + 6x_2 = 9 \end{cases}$$

则是无解。



5.1.3 齐次线性方程组

齐次线性方程组是线性方程组中的一个特例，定义如下

定义 5.1 (齐次线性方程组 (homogeneous linear equations))

齐次线性方程组指的是常数项全部为零的线性方程组，它们一般记作

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = 0 \end{array} \right.$$



我们在第 2 章的 B 组练习中有提到这种方程组，先说结论，当 $n > m$ 的时候齐次线性方程组必有非零解，在练习中，我们使用线性映射

$$T(x_1, x_2, \dots, x_n) = \left(\sum_{i=1}^n a_{1i}x_i, \sum_{i=1}^n a_{2i}x_i, \dots, \sum_{i=1}^n a_{mi}x_i \right)$$

这里的线性映射 T 将 \mathbb{F}^n 映射为 \mathbb{F}^m ，根据线性映射的基本定理有

$$\dim \mathbb{F}^n = \dim \text{null } T + \dim \text{range } T$$

使得齐次线性方程组有无穷多解的充要条件是 T 不是单射；所以 $\text{null } T \neq \{0\}$ （否则对应映射到零空间的向量只有 $\mathbf{0}$ ）。

接下来我们考虑特殊的情况，当 $m = n$ 的时候，系数矩阵的行列式 $\mathbf{D} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{vmatrix}$ 如果我们

对其使用 Cramer 法则，当 $\mathbf{D} \neq 0$ 时， $x_i = \frac{\mathbf{D}_i}{\mathbf{D}} = 0$ 有且仅有全为 0 解，即 $x_1 = x_2 = \cdots = x_n = 0$ 。那如果 $\mathbf{D} = 0$ 则系数矩阵的几行向量线性相关，根据线性相关的向量可以通过去除 i 个向量可以得到线性无关的向量，由于 $m = n, m - i < n$ ，所以我们得到齐次线性方程组有无穷多解。

5.2 初等矩阵

5.2.1 增广矩阵

首先需要注意的是，按照初等矩阵的定义，增广矩阵不算是初等矩阵的一种，但是因为它可以使用矩阵的初等行变化，所以笔者将其纳入初等矩阵这一节。

考虑线性方程组

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{array} \right.$$

在系数矩阵的右边添上一列，这一列是线性方程组的等号右边的值，构成增广矩阵，记作 $\tilde{\mathbf{A}}$



定义 5.2 (增广矩阵 (augmented matrix))

线性方程组

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{array} \right.$$

其系数构成一个矩阵后在最后一列添加系数矩阵, 记作 $\tilde{\mathbf{A}}$, 即

$$\tilde{\mathbf{A}} := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$



通过增广矩阵的初等变换, 我们可以求解线性方程组, 通常这个方法叫做高斯消元法, 接下里我们看一个例子:

例题 5.2.

写出线性方程组

$$\left\{ \begin{array}{l} x + y + z = 4 \\ 2x + 4y + 5z = 16 \\ x + 5y - 3z = 0 \end{array} \right.$$

的增广矩阵

解: 见下讲解

我们提取该方程组的增广矩阵, 即

$$\tilde{A} = \begin{pmatrix} 1 & 1 & 1 & 4 \\ 2 & 4 & 5 & 16 \\ 1 & 5 & -3 & 0 \end{pmatrix}$$

应用矩阵的初等行变换就可以求解该线性方程组, 下面先讲解一下什么是矩阵的初等变换。

5.2.2 矩阵的初等变换

首先为了说明变换后的矩阵和原矩阵的关系, 我们使用“等价”来描述它们, 需要注意的是, 它们通常并不相等。设矩阵 \mathbf{A} 经过初等变换后变为 \mathbf{A}' , 则 \mathbf{A} 与 \mathbf{A}' 等价, 记作

$$\mathbf{A} \Rightarrow \mathbf{A}'$$

如无特殊说明, 我们这里都使用初等行变换, 实际上对矩阵的列进行变换操作在某些情况下也是等价的。

5.2.2.1 交换两行

交换两行是矩阵的初等变换, 例如

$$\begin{pmatrix} 1 & 3 & 3 & 3 \\ 2 & 2 & 2 & 1 \\ 4 & 2 & 1 & 3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 3 & 3 & 3 \\ 4 & 2 & 1 & 3 \\ 2 & 2 & 2 & 1 \end{pmatrix}$$



5.2.2.2 某一行乘以常数 k

某一行乘以常数 k 是矩阵的初等变换，例如

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 4 & 2 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 3 & 1 \\ 4k & 2k & k & 2k \\ 2 & 2 & 2 & 1 \end{pmatrix}, k \in \mathbb{C}$$

5.2.2.3 某个数乘以某一行并加到另一行中去

某个数乘以某一行并加到另一行中去是矩阵的初等变换，例如

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 2 & 4 \\ 2 & 3 & 4 & 5 \\ 4 & 5 & 6 & 7 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 2 & 4 \\ 2+2 & 3+4 & 4+6 & 5+8 \\ 4 & 5 & 6 & 7 \end{pmatrix}$$

5.2.2.4 最简矩阵

矩阵经过一定的初等变换，可化简为最简矩阵，在给出最简矩阵的定义之前，先给出一个术语来描述一个特定行。

定义 5.3 (零行)

矩阵的某一行均为 0，则称这一行为零行，否则为非零行。



下面给出行阶梯矩阵与最简矩阵的定义。

定义 5.4 (行阶梯形矩阵与最简矩阵)

非零矩阵若满足，非零行在零行的上面，且非零行的首个非零元素在列的上一行（如果存在的话）的首个非零元素所在列的后面，例如下面一个就是行阶梯形矩阵^a

$$A = \begin{pmatrix} 1 & 0 & 4 & 1 & 2 & 3 \\ 0 & 1 & 3 & 5 & 0 & 3 \\ 0 & 0 & 0 & 2 & 3 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

特别地，当行阶梯形矩阵的每一行第一个非零元素，例如

$$\begin{pmatrix} 1 & 0 & 4 & 1 & 2 & 3 \\ 0 & 1 & 3 & 5 & 0 & 3 \\ 0 & 0 & 0 & 1 & 3 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

我们称这个矩阵为最简矩阵^b。

^a如红色标出的 0 元素，像一个阶梯一样从左往右下降

^b最简矩阵一般指的是最简行矩阵，如果是列最简，我们会明确说明是“列最简矩阵”



推论 5.1

任何矩阵经过有限次的行变换都可以转化为最简矩阵。



5.2.3 高斯消元法

线性方程组的增广矩阵可以由行变换变为最简行矩阵，通常可用于解无法使用 Cramer 法则的线性方程组¹，这种方法叫做高斯消元法。

首先看线性方程组有唯一解的情况，即在 Cramer 法则下系数矩阵的行列式值不为 0，下面是一个例子

$$\begin{cases} x + y + z = 4 \\ 2x + 4y + 5z = 16 \\ x + 5y - 3z = 0 \end{cases}$$

我们提取该方程组的增广矩阵为

$$\tilde{\mathbf{A}} = \begin{pmatrix} 1 & 1 & 1 & 4 \\ 2 & 4 & 5 & 16 \\ 1 & 5 & -3 & 0 \end{pmatrix}$$

将其变化为行阶梯矩阵，首先将第 1 行乘以 -2 加到第 2 行，我们可得等价矩阵

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 2 & 4 & 5 & 16 \\ 1 & 5 & -3 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 3 & 8 \\ 1 & 5 & -3 & 0 \end{pmatrix}$$

将第 1 行乘以 -1 加到第 3 行，我们继续可得等价矩阵

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 3 & 8 \\ 1 & 5 & -3 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 3 & 8 \\ 0 & 4 & -4 & -4 \end{pmatrix}$$

将第 3 行除以 -2 加到第 2 行后，交换 2, 3 两行我们继续可得等价矩阵

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 3 & 8 \\ 0 & 4 & -4 & -4 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & -2 & 2 & 2 \\ 0 & 0 & 5 & 10 \end{pmatrix}$$

最后将第 2 行除以 -2 ，第 3 行除以 5 可得矩阵

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & -2 & 2 & 2 \\ 0 & 0 & 5 & 10 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

如果我们重写为方程组即

$$\begin{cases} x + y + z = 4 \\ y - z = -1 \\ z = 2 \end{cases}$$

将结果逐个带入线性方程组，可得 $x = 1, y = 1, z = 2$ 。下面我们给出线性方程组有无穷个解的情况。

例题 5.3.

使用高斯消元法求解线性方程组

$$\begin{cases} x + 2y + 4z = 2 \\ 2x + y - z = 1 \\ x + y + z = 1 \end{cases}$$

¹斋藤正彦在它的著作《线性代数入门》这本书中，讲到了有位计算机学家告诉他，Cramer 法则不适合用于大量行列式的数值计算，相比之下高斯消元法更好

解: $x = 3z, y = 1 - 3z, z = z$ 见下讲解

我们提取该方程组的增广矩阵为

$$\tilde{\mathbf{A}} = \begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

首先将第 3 行乘以 3, 我们可得

$$\begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 3 & 3 & 3 & 3 \end{pmatrix}$$

其次将第 1, 2 行乘以 -1 后依次加到第三行可得

$$\begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 3 & 3 & 3 & 3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

最后将第 1 行乘以 -2 加到第 2 行, 第 2 行除以 -3 得到

$$\begin{pmatrix} 1 & 2 & 4 & 2 \\ 2 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 & 4 & 2 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

如果我们重写为方程组即

$$\begin{cases} x + 2y + 4z = 2 \\ y + 3z = 1 \end{cases}$$

由此我们可以知道该方程组有无穷个解, 我们令线性无关组 $z = z$ 其他的量均可使用 z 来表示, 那么回代后可得 $y = 1 - 3z, x = 2z$; 至于更多复杂的方程组, 读者可使用 <https://matrixcalc.org/zh-CN/slu.html> 页面求解。

5.3 矩阵的逆

5.3.1 单位矩阵

首先我们来看单位矩阵的定义:

定义 5.5 (单位矩阵)

设 $n \times n$ 矩阵的主对角线的元素均为 1, 则称该矩阵为 n 阶单位矩阵, 简称单位矩阵, 写作 \mathbf{I} , 即

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

那么对于 n 阶方阵 \mathbf{A} , 我们有下述推论成立:

推论 5.2

若 \mathbf{A} 为 n 阶方阵, \mathbf{I} 是单位矩阵, 则有

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

成立



请读者举例验证上述式子成立，不作严格证明；单位矩阵由于其相乘不变性，通常用于矩阵乘法逆元的计算，即求解逆矩阵。

5.3.2 逆矩阵

首先看逆矩阵的定义。

定义 5.6 (逆矩阵)

设 \mathbf{A} 是一个 n 阶方阵，若存在另一个 n 阶方阵 \mathbf{B} 使得

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

成立，则称 \mathbf{B} 为 \mathbf{A} 的逆矩阵，记作 \mathbf{A}^{-1} 所以

$$\mathbf{AA}^{-1} = \mathbf{I}$$



关于逆矩阵存在下述推论

推论 5.3

1. 若 \mathbf{A} 可逆，则 \mathbf{A}^{-1} 也可逆且 $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ 。
 2. 若矩阵 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_s$ 均可逆，则它们乘积 $\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_s$ 也可逆，并且满足
- $$(\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_s)^{-1} = \mathbf{A}_s^{-1} \cdots \mathbf{A}_2^{-1} \mathbf{A}_1^{-1}$$
3. 若 \mathbf{A} 可逆，则 \mathbf{A}^T 也可逆，并且 $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$ 。
 4. 若 \mathbf{A} 可逆且 $k \neq 0$ 那么有 $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1}$ 。
 5. 若 \mathbf{A} 可逆则 $\det \mathbf{A} \neq 0$ 。



上述性质我们在此不作严格证明，我们仅描述使用相关方法来求逆矩阵；通常来说，求解逆矩阵可以帮助我们求解关于矩阵的方程。

5.3.3 初等变换求逆矩阵

由单位矩阵经过一次初等变换得到的矩阵称为初等矩阵，所以初等矩阵一般为 n 阶方阵。而初等行变换则可以表示为有一个矩阵左乘另一个矩阵变化得来，例如交换两行

$$\begin{pmatrix} 3 & 2 & 1 \\ 1 & -2 & 1 \\ 3 & 5 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & -2 & 1 \\ 3 & 2 & 1 \\ 3 & 5 & 0 \end{pmatrix}$$

右边的矩阵可以由左边的矩阵左乘一个矩阵 \mathbf{P} 得到，其中 $\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 那么根据这个性质， \mathbf{A} 可以通过 n 次初等行变换，转化为单位矩阵，即存在矩阵 $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_s$ 使得 $\mathbf{P}_s, \dots, \mathbf{P}_2 \mathbf{P}_1 \mathbf{A} = \mathbf{I}$ 所以 $\mathbf{A}^{-1} = \mathbf{P}_s \cdots \mathbf{P}_2 \mathbf{P}_1$ ，然后可以转化为单位矩阵，单位矩阵继续进行该过程，可以得到逆矩阵。



接下来我们讲解逆矩阵的求法，考虑 n 阶方阵 $\mathbf{A} = \begin{pmatrix} 2 & 3 & -2 \\ 0 & 3 & 4 \\ 1 & 1 & 1 \end{pmatrix}$ ，我们使用分块矩阵

$$\left(\begin{array}{c|cc} \mathbf{A} & \mathbf{I} \end{array} \right) = \left(\begin{array}{ccc|ccc} 2 & 3 & -2 & 1 & 0 & 0 \\ 0 & 3 & 4 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right)$$

并对其施加初等行变化将左侧矩阵化为单位矩阵时，右侧矩阵即为逆矩阵，下面是相关过程：首先第一行乘以 $\frac{1}{2}$ 可得

$$\left(\begin{array}{ccc|ccc} 2 & 3 & -2 & 1 & 0 & 0 \\ 0 & 3 & 4 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 3 & 4 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right)$$

再将第 1 行乘以 -1 加到第三行可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 3 & 4 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 3 & 4 & 0 & 1 & 0 \\ 0 & -\frac{1}{2} & 2 & -\frac{1}{2} & 0 & 1 \end{array} \right)$$

将第 2 行乘以 $\frac{1}{3}$ 可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{1}{2} & 2 & -\frac{1}{2} & 0 & 1 \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{1}{2} & 2 & -\frac{1}{2} & 0 & 1 \end{array} \right)$$

将第 2 行乘以 $\frac{1}{2}$ 后加到第三行可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{1}{2} & 2 & -\frac{1}{2} & 0 & 1 \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{8}{3} & -\frac{1}{2} & \frac{1}{6} & 1 \end{array} \right)$$

将第 3 行乘以 $\frac{3}{8}$ 可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{8}{3} & -\frac{1}{2} & \frac{1}{6} & 1 \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right)$$

再将第 3 行乘以 $-\frac{4}{3}$ 后加到第 2 行可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & \frac{4}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right)$$

将第 3 行加到第 1 行可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & -1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & 0 & \frac{5}{16} & \frac{1}{16} & \frac{3}{8} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right)$$

最后将第 3 行乘以 $-\frac{3}{2}$ 加到第 1 行可得

$$\left(\begin{array}{ccc|ccc} 1 & \frac{3}{2} & 0 & \frac{5}{16} & \frac{1}{16} & \frac{3}{8} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{16} & -\frac{5}{16} & \frac{9}{8} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{array} \right)$$

所以矩阵变更为 $(I | A^{-1})$, 得到 $A^{-1} = \begin{pmatrix} -\frac{1}{16} & -\frac{5}{16} & \frac{9}{8} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ -\frac{3}{16} & \frac{1}{16} & \frac{3}{8} \end{pmatrix}$

5.4 矩阵的秩

5.4.1 极大线性无关组

接下来请读者回顾一下1.5.2相关的内容, 如果向量集合 $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 线性无关, 其中向量 β 可以由 S 的线性组合得到, 那么存在唯一的 $k_1, k_2, \dots, k_n \in \mathbb{F}$ 使得 $\beta = k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n$ 成立, 如果 S 线性相关, 则需要去除一些向量使得其线性无关这样就可以做到 $k_1, k_2, \dots, k_n \in \mathbb{F}$ 唯一。

由此我们引入极大线性无关组的概念:

定义 5.7 (极大线性无关组)

设向量组^a $S = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_s\}$ 中选取 r 个向量, 构成集合 $S_r = \{\beta_1, \beta_2, \beta_3, \dots, \beta_r\}$ 满足:

1. S_r 线性无关;
2. 若从 S 中再选取一个 α_i 到 S'_r , 则 S'_r 一定线性相关。

我们称 S_r 为 S 的极大线性无关组。

^a这里的组更偏向于组合而不是元组的意思, 可以将其理解为向量集合的另一个称谓, 因为从一个集合内选取的组合同时满足 1. 无序性, 2. 唯一性, 3. 确定性



例如前面讲到的在三维空间 \mathbb{R}^3 中, 极大线性无关组中向量个数最多为 3, 例如

$S = \{(1, 1, 1), (1, 1, 2), (1, 5, 3), (2, 3, 3)\}$ 可以选取 3 个线性无关的向量作为基底表示第 4 个向量, 而极端情况 $S = \{(1, 1, 1), (2, 2, 2), (3, 3, 3), (4, 4, 4)\}$ 则极大线性无关组中向量个数 1。

此外我们有如下推论:

推论 5.4

向量组 S 张成空间 $\text{Span}(S)$ 等于其极大线性无关组 S_r 张成空间 $\text{Span}(S_r)$ 。



证明 首先由于 S_r 是 S 的一个子集, 任何 S_r 中向量的线性组合显然也是 S 中向量的线性组合。因此, $\text{Span}(S_r)$ 是 $\text{Span}(S)$ 的一个子空间, 即 $\text{Span}(S_r) \subseteq \text{Span}(S)$ 。

其次任取向量 $v \in \text{Span}(S)$, 则 v 可以表示为 S 中某些向量的线性组合, 即存在向量 $v_1, v_2, \dots, v_k \in S$ 和标量 a_1, a_2, \dots, a_k 使得 $v = a_1v_1 + a_2v_2 + \dots + a_kv_k$ 。

将每个 v_i 代入 v 的表达式中, 得到:

$$v = a_1(b_1u_1 + \dots + b_mu_m) + \dots + a_k(b_1^{(k)}u_1 + \dots + b_m^{(k)}u_m)^2$$

展开后, 这是 S_r 中向量的线性组合, 因此 $v \in \text{Span}(S_r)$ 。从而 $\text{Span}(S) \subseteq \text{Span}(S_r)$ 。

综上所述, $\text{Span}(S) = \text{Span}(S_r)$ 。



推论 5.5

向量组 S 的所有极大线性无关组 S_r 所含的向量个数相等。



证明 设向量组 S 中有两个极大线性无关子集, 分别是:

- $S_1 = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r\}$

² $b_i^{(k)}$ 表示第 k 个向量 v_k 对应 S_r 中第 i 个向量的系数



- $S_2 = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_s\}$

我们要证明：

$$r = s$$

S_1 是线性无关组，并且在 S 中无法再添加其他向量使其仍然线性无关；所以 $S_1 \subseteq S$ ，并且 $S_2 \subseteq S$ 。根据定理 1.1，由于 $S_1 \subseteq \text{Span}(S)$ ，而 $S_2 \subseteq \text{Span}(S)$ ，并且二者都是 S 中的极大无关组 \Rightarrow 它们生成的是相同的空间（因为如果不是，就不是极大的了）。因此：

$$r \leq s, s \leq r$$

所以： $r = s$

□

5.4.2 向量组与矩阵的秩

上面我们讲到了极大线性无关组的所含向量个数是确定的，由此我们引出向量组的秩的概念。

定义 5.8 (向量组的秩 (rank))

向量组 S 的极大线性无关组所含的向量个数称为这个向量组的秩 (rank)，记作 $\text{rank}(S)$ 。



例如向量组 $S = \{(1, 1, 1), (1, 2, 1), (2, 2, 2)\}$ 的秩 $\text{rank}(S) = 2$ 因为它的极大线性无关组向量的数量为 2。

如果我们把向量组的每一个向量竖着写成列向量，从左向右分别表示 $\alpha_1, \alpha_2, \alpha_3$ ，那么上述的向量组 S 表示为矩阵 \mathbf{A} 即

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

定义 5.9 (列秩, 行秩, 矩阵的秩)

矩阵 \mathbf{A} 的列向量组的秩称为列秩，矩阵的行向量组的秩称为行秩，矩阵的秩一般指的是矩阵的列秩，记作 $r(\mathbf{A})$ 。



例如上面的矩阵 \mathbf{A} 的秩 $r(\mathbf{A}) = 2$ ，下面的推论说明了矩阵的列秩等于行秩。

推论 5.6

矩阵的列秩等于行秩。



证明 我们从线性映射的角度出发，将一个矩阵看作一个线性变换。

$$T_A : \mathcal{V}(\mathbb{F}^n, \mathbb{F}^m), \quad T_A(x) = Ax$$

根据定理 2.1，线性变换 $T_A : \mathbb{F}^n \rightarrow \mathbb{F}^m$ 满足：

$$\text{rank}(A) = \dim(\text{null}(T_A)) = n - \dim(\text{range } A)$$

同样地，对于 $A^T \in \mathbb{F}^{n \times m}$ ：

$$\text{rank}(A^T) = \dim(\text{null}(A^T)) = m - \dim(\text{range } A^T)$$

反求出 $\text{rank}(A^T) = \text{rank}(A)$

□



推论 5.7

矩阵经过初等变换，其秩不变。



此推论不作证明，下面我们会利用此推论来求解矩阵的秩：

例题 5.4.

求矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 3 & 2 & 1 & 1 \\ 3 & 2 & 6 & 3 \\ 8 & 6 & 9 & 8 \end{pmatrix}$ 的秩。

解：将矩阵第 1 行乘以 -3 加到第 2 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 3 & 2 & 1 & 2 \\ 3 & 2 & 6 & 2 \\ 8 & 6 & 9 & 8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 3 & 2 & 6 & 2 \\ 8 & 6 & 9 & 8 \end{pmatrix}$$

将矩阵第 1 行乘以 -3 加到第 3 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 3 & 2 & 6 & 2 \\ 8 & 6 & 9 & 8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & -1 & 3 & 4 \\ 8 & 6 & 9 & 8 \end{pmatrix}$$

将矩阵第 1 行乘以 -8 加到第 4 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & -1 & 3 & 4 \\ 8 & 6 & 9 & 8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & -1 & 3 & 4 \\ 0 & -2 & 1 & -8 \end{pmatrix}$$

将矩阵第 2 行乘以 -1 加到第 3 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & -1 & 3 & 4 \\ 0 & -2 & 1 & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & 5 & 0 \\ 0 & -2 & 1 & -8 \end{pmatrix}$$

将矩阵第 2 行乘以 -2 加到第 4 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & 5 & 0 \\ 0 & -2 & 1 & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 5 & 0 \end{pmatrix}$$

将矩阵第 3 行乘以 -1 加到第 4 行可得

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 5 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

由此得到， $r(\mathbf{A}) = 3$



上述题例我们把原来的矩阵经过初等变换得到一个新的矩阵，这个矩阵的非零行的数量为原矩阵的秩。

5.4.3 秩与维度

根据推论5.4我们可以得知，空间 $V = \text{Span}(S_r)$ 中的向量组 S_r 满足线性无关且张成 V ，所以向量组 S_r 可以看作是空间 V 的一个基。而刻画维度的一个重要特征就是其基中向量的个数，我们不难发现，向量组的秩等于其张成空间的维度。

所以请读者区分 2 个不同范畴的概念，一是极大线性无关组和基，二是秩和维度，它们的前者，极大线性无关组与秩均是用来刻画一个向量组的属性，而后者，基与维度用来刻画线性空间。

现在我们回顾这个重要的定理2.1，若使用矩阵 \mathbf{A} 表示线性映射 T ，那么该矩阵的秩 $r(\mathbf{A})$ 可以表示映射到非零空间的维度，即 $\dim \text{range}T$ ，它们在数量上两者相等

$$r(\mathbf{A}) = \dim \text{range}T$$

由此，我们可以将秩替换掉 $\dim \text{range}T$ 获得一个更加正式的定理，该名称为秩——零化度定理。

定理 5.1 (秩——零化度定理 (Rank-Nullity Theorem))

若 $m \times n$ 矩阵 \mathbf{A} 表示线性映射 $T : \mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ 满足

$$r(\mathbf{A}) + \dim \text{null}T = n$$



证明过程同2.1，由于 $m \times n$ 矩阵 \mathbf{A} 的行秩等于列秩，其秩通常可以通过 m, n 中较小的行或列进行初等变换处理，所以其秩 $r(\mathbf{A}) \leq \min(m, n)$ 。

5.5 线性方程组的解空间

5.5.1 齐次线性方程组解的结构

在前面的节5.1.3中我们了解到齐次线性方程组中使用 Cramer 法则研究了它解的结构，共有两种情况，分别为“有且仅有全为 0 解”和“有非零解”，本节我们从其系数矩阵的秩的角度来研究其解的结构。

接下来我们使用 $r(\mathbf{A})$ 表示系数矩阵 \mathbf{A} 的秩，未知数的个数为 n ，考虑齐次线性方程组

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = 0 \end{array} \right.$$

根据矩阵乘的定义它们可写作 $\mathbf{Ax} = \mathbf{0}$ ，其中

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

那么根据我们研究的第一种情况，我们有如下推论：

推论 5.8

若 $r(\mathbf{A}) \geq n$ ，齐次线性方程组“有且仅有全为 0 解”。



证明 设 \mathbf{A} 是一个 $m \times n$ 的系数矩阵，对应的齐次线性方程组为 $\mathbf{Ax} = \mathbf{0}$ 。根据矩阵 \mathbf{A} 的秩 $r(\mathbf{A})$ 不超过其行



数 m 和列数 n 中的较小值, 即 $r(\mathbf{A}) \leq \min\{m, n\}$ 。推论中给定 $r(\mathbf{A}) \geq n$, 因此必须满足 $\min\{m, n\} \geq n$, 即 $m \geq n$ 。结合 $r(\mathbf{A}) \leq n$, 可得 $r(\mathbf{A}) = n$ 。其次根据秩——零化度定理: 向量 x 张成的空间的维数为 $n - r(\mathbf{A})$ 。当 $r(\mathbf{A}) = n$ 时, 解空间的维数为 $n - n = 0$, 即解仅包含零向量。向量 x 张成的空间的维数为 0, 意味着齐次线性方程组 $\mathbf{Ax} = 0$ 只有零解。因此, 当系数矩阵 \mathbf{A} 的秩 $r(\mathbf{A}) \geq n$ 时, 齐次线性方程组有且仅有零解。

□

推论 5.9

若 $r(\mathbf{A}) < n$, 齐次线性方程组“有非零解”。



证明 证明同上

□

5.5.2 解空间

如上所示, x 表示一个向量, 若齐次线性方程组有非零解, 其解一般有多组, 这些解向量 x 可以张成一个解空间 $\text{Span}(x)$, 下面举个例子: 齐次线性方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ x_1 + 3x_2 + 2x_3 = 0 \end{cases}$$

其系数矩阵为

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

将其简化为

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \xrightarrow{r_2-r_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow{r_1-2r_2} \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & -1 \end{pmatrix}$$

最后还原成方程组

$$\begin{cases} 1x_1 + 0x_2 + 5x_3 = 0 \\ 0x_1 + 1x_2 - 1x_3 = 0 \end{cases}$$

我们令 $x_3 = k, k \in \mathbb{R}$ 其解为 $x = \begin{pmatrix} -5k \\ k \\ k \end{pmatrix}$ 。

如果我们把向量 x 的 k 提出, 那就是 $x = k(-5, 1, 1)$, 而解向量集合构成的空间的一个基为 $\{(-5, 1, 1)\}$, 根据该线性方程组的解, 我们有如下定义:

定义 5.10 (解空间 (solution space) 与基础解系 (fundamental system of solutions))

若齐次线性方程组的系数矩阵 \mathbf{A} 和解向量 x 满足 $\mathbf{Ax} = \mathbf{0}$, 那么其线性方程组所有解向量构成的集合称为解空间, 基础解系为其解空间的一组基。



那么根据上述的定义, 我们可以得到

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ x_1 + 3x_2 + 2x_3 = 0 \end{cases}$$

的解空间为 $S = k(-5, 1, 1), k \in \mathbb{R}$ 其一个基础解系 $\mathcal{B} = \{(-5, 1, 1)\}$ 。

推论 5.10

基础解系的一组向量的个数等于 $n - r(\mathbf{A})$, 其中 n 为未知数个数, \mathbf{A} 为系数矩阵。



证明 我们这里使用非严格证明, 若 $n - r(\mathbf{A}) > 0$ 还原为方程组的个数会比未知数个数要少, 所以我们需要 $n - r(\mathbf{A})$ 个自由变量来表示所有解。因此基础解系的一组向量的个数等于 $n - r(\mathbf{A})$ 。

□

推论 5.11

解空间是线性空间。



证明 要证明齐次线性方程组 $\mathbf{Ax} = \mathbf{0}$ 的解空间是一个线性空间, 我们需要验证解空间满足线性空间的三个核心条件: 包含零向量、加法封闭性和标量乘法封闭性。

1. 零向量存在: 零向量 $\mathbf{0}$ 显然满足 $\mathbf{A}\mathbf{0} = \mathbf{0}$, 因此 $\mathbf{0}$ 属于解空间。
2. 加法封闭性: 对于解空间中的任意两个向量 \mathbf{x}_1 和 \mathbf{x}_2 , 有 $\mathbf{Ax}_1 = \mathbf{0}$ 和 $\mathbf{Ax}_2 = \mathbf{0}$ 。根据矩阵乘法的分配律,

$$\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{Ax}_1 + \mathbf{Ax}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

因此 $\mathbf{x}_1 + \mathbf{x}_2$ 属于解空间。

3. 标量乘法封闭性: 对于任意标量 $k \in \mathbb{R}$ 和解空间中的任意向量 \mathbf{x} , 有 $\mathbf{Ax} = \mathbf{0}$ 。根据标量乘法与矩阵乘法的结合律,

$$\mathbf{A}(k\mathbf{x}) = k\mathbf{Ax} = k\mathbf{0} = \mathbf{0}$$

因此 $k\mathbf{x}$ 属于解空间。

由于解空间满足上述三个条件, 因此它是原向量空间的一个子空间, 从而自身也是一个线性空间。

□

例题 5.5.

求下列线性方程组的基础解系与解空间:

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ 2x_1 + 4x_3 = 0 \\ x_1 + x_2 + x_4 = 0 \end{cases}$$

解: 系数矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 4 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$ 经过初等变换可得

$$\begin{array}{c} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 4 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \xrightarrow{r_3 - r_1} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & -1 & -3 & 1 \end{pmatrix} \xrightarrow{r_2 / 2} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -1 & -3 & 1 \end{pmatrix} \xrightarrow{r_3 + r_2} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \\ \xrightarrow{r_3 \times (-1)} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \xrightarrow{r_2 - 2r_3} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \end{pmatrix} \xrightarrow{r_1 - 3r_3} \begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \end{pmatrix} \xrightarrow{r_1 - 2r_2} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array}$$

将其还原为线性方程组即为

$$\begin{cases} x_1 - x_4 = 0 \\ x_2 + 2x_4 = 0 \\ x_3 - x_4 = 0 \end{cases}$$

其中 $n = 4, r(\mathbf{A}) = 3$ 所以自由变量的个数为 $n - r(\mathbf{A}) = 4 - 3 = 1$, 其基础解系的个数也为 1, 故令



$x_4 = 1$ 可解得 $x = (1, -2, 1, 1)$, 基础解系为 $\mathcal{B} = \{(1, -2, 1, 1)\}$ 令 $k, k \in \mathbb{R}$ 为自由变量, 所以解空间 $S = k(1, -2, 1, 1)$

思考

非齐次线性方程组的解向量的集合是解空间吗?

上述思考题答案为不是, 具体原因见章节5.5.3。

5.5.3 非齐次线性方程组的导出组与特解

上面我们描述了齐次线性方程组的解向量, 以及其集合是一个解空间; 那么本节我们来讲一讲非齐次线性方程组解的结构。

现在我们来看若为非齐次的线性方程组, 我们把它抽象化为一个矩阵乘法 $\mathbf{A}x = b$ 其中 $x, b \in \mathbb{R}^n$, 由此我们先定义什么是导出组。

定义 5.11 (导出组)

非齐次线性方程组 $\mathbf{A}x = b, x, b \in \mathbb{R}^n$ 的导出组为 $\mathbf{A}x = \mathbf{0}$ 。



实际上导出组就是将非齐次线性方程组的常数项改为零向量, 变成齐次线性方程组, 为了研究非齐次线性方程组解的集合, 下面给出非齐次线性方程组的特解的定义:

定义 5.12 (特解)

非齐次线性方程组 $\mathbf{A}x = b$ 若存在 x_p 使得 $\mathbf{A}x_p = b$ 成立, 则 x_p 为该非齐次线性方程组的特解。



例如下面有一个非齐次线性方程组

$$\begin{cases} x_1 + x_2 + 2x_3 = 2 \\ x_1 + 2x_2 = 0 \end{cases}$$

的特解 x_p 可以为 $(0, 0, 1)$ 也可以为 $(4, -2, 0)$, 接下来我们根据导出组与特解的定义有如下的推论。

推论 5.12

非齐次线性方程组 $\mathbf{A}x = b$ 的所有解向量所构成的集合 S' 为其导出组 $\mathbf{A}x = \mathbf{0}$ 的解空间 S 加上其一个特解向量 x_p , 即

$$S' = S + x_p$$



证明 首先, 我们考虑非齐次线性方程组 $\mathbf{A}x = b$ 的解集合 S' 和其导出组 $\mathbf{A}x = \mathbf{0}$ 的解空间 S 。假设存在一个特解 x_p 使得 $\mathbf{A}x_p = b$ 。

1. 验证 $S + x_p$ 中的每个元素都是 $\mathbf{A}x = b$ 的解: 对于任意的 $s \in S$, 有 $\mathbf{A}s = \mathbf{0}$ 。令 $x = s + x_p$, 则 $\mathbf{A}(s + x_p) = \mathbf{A}s + \mathbf{A}x_p = \mathbf{0} + b = b$, 因此 $x = s + x_p$ 是 $\mathbf{A}x = b$ 的解。这说明 $S + x_p$ 是 S' 的一个子集。
2. 验证 S' 是 $S + x_p$ 的子集: 假设 x' 是 $\mathbf{A}x = b$ 的任意一个解, 则 $\mathbf{A}x' = b$ 。考虑 $x' - x_p$, 有 $\mathbf{A}(x' - x_p) = \mathbf{A}x' - \mathbf{A}x_p = b - b = 0$, 因此 $x' - x_p \in S$ 。这说明 $x' = (x' - x_p) + x_p$, 其中 $(x' - x_p) \in S$, 所以 $x' \in S + x_p$ 。因此, S' 是 $S + x_p$ 的子集。

综上所述, S' 既是 $S + x_p$ 的子集, 又是它的超集, 因此 $S' = S + x_p$ 。



由此我们可以根据上述推论来解决上面没有解决的思考题: 非齐次线性方程组的解向量的集合不是解空间;



证明 为了证明非齐次线性方程组的解向量的集合不是解空间(线性空间), 我们可以从线性空间的定义出发, 验证其是否满足线性空间的条件³。

1. 零向量的存在性: 非齐次线性方程组的形式为 $Ax = b$, 其中 $b \neq 0$ 。零向量 0 满足 $A0 = 0$, 但由于 $b \neq 0$, 显然 0 不是非齐次方程组的解。因此, 解集中不包含零向量, 而零向量是线性空间的必要条件之一。
2. 加法封闭性: 假设 x_1 和 x_2 是非齐次方程组的解, 即 $Ax_1 = b$ 和 $Ax_2 = b$ 。考虑它们的和 $x_1 + x_2$, 有:

$$A(x_1 + x_2) = Ax_1 + Ax_2 = b + b = 2b$$

由于 $b \neq 0$, 当 $2b \neq b$ 时, $x_1 + x_2$ 不是非齐次方程组的解, 因此加法不封闭。

3. 标量乘法封闭性: 假设 x 是非齐次方程组的解, 即 $Ax = b$ 。考虑标量乘法 $c \cdot x$ (其中 c 是标量), 有:

$$A(c \cdot x) = c \cdot Ax = c \cdot b$$

当 $c \neq 1$ 时, $c \cdot b \neq b$, 因此 $c \cdot x$ 不是非齐次方程组的解, 标量乘法不封闭。

由于非齐次线性方程组的解集不包含零向量, 并且不满足加法和标量乘法的封闭性, 因此它不构成线性空间。

□

根据上面的推论与定义, 我们可以使用解空间加特解的方式来表示所有解, 下面是一个例题:

例题 5.6.

求解下列的线性方程组并使用基础解系与特解的形式表示

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 = 10 \\ 2x_1 + x_2 + x_3 + x_4 = 5 \end{cases}$$

解: 提取线性方程组的增广矩阵

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 & 10 \\ 2 & 1 & 1 & 1 & 5 \end{pmatrix} \xrightarrow{r_1 \times 2} \begin{pmatrix} 2 & 4 & 6 & 8 & 20 \\ 2 & 1 & 1 & 1 & 5 \end{pmatrix} \xrightarrow{r_2 - r_1} \begin{pmatrix} 2 & 4 & 6 & 8 & 20 \\ 0 & -3 & -5 & -7 & -15 \end{pmatrix}$$

矩阵的秩 $r(\mathbf{A}) = 2$ 方程组未知数的个数为 $n = 4$ 其自由变量个数为 $n - r(\mathbf{A}) = 2$, 那么设 $k_1, k_2 \in \mathbb{R}$, 对于原方程组, 由于自由变量有 2 个, 我们可以固定 2 个未知变量的值, 例如 $x_1 = 1, x_2 = 1$ 得到一个特解为 $x_p = (1, 1, 1, 1)$ 。此外其导出组的增广矩阵为

$$\begin{pmatrix} 2 & 4 & 6 & 8 & 0 \\ 0 & -3 & -5 & -7 & 0 \end{pmatrix} \xrightarrow{r_1/2} \begin{pmatrix} 1 & 2 & 3 & 4 & 0 \\ 0 & -3 & -5 & -7 & 0 \end{pmatrix}$$

根据自由变量个数为 2, 则线性无关组会有 2 个向量, 可设 $x_1 = 0, x_2 = 1$ 并求出 $x_3 = -2, x_4 = 1$ 作为第一个向量, $x_1 = 1, x_2 = 0$ 则 $x_3 = -7, x_4 = 5$ 作为第二个向量, 由于 $(1, 0), (0, 1)$ 线性无关则它们的延长组 $(0, 1, -2, 1)$ 和 $(1, 0, -7, 5)$ 必定线性无关所以其导出组的一个基础解系 $\mathcal{B} = \{(0, 1, -2, 1), (1, 0, -7, 5)\}$, 其解空间为 $k_1(0, 1, -2, 1) + k_2(1, 0, -7, 5)$ 再加上特解的结果, 最终结果为

$$x = k_1(0, 1, -2, 1) + k_2(1, 0, -7, 5) + (1, 1, 1, 1), k_1, k_2 \in \mathbb{R}$$

³实际上我们只需证明一项不符合线性空间的要求即可

5.6 章节练习

5.6.1 A 组

练习 40 使用 Cramer 法则计算线性方程组

$$\begin{cases} 2x + 3y + 4z = 1 \\ 4x + 3y - z = -4 \\ x - 2y + 3z = 5 \end{cases}$$

练习 41

若矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 3 & 4 & 2 \\ 2 & 0 & 0 & 2 \\ -1 & \sqrt{3} & 1 & -2 \end{pmatrix}$, 将其化为最简矩阵。

练习 42

若矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 1 \\ -2 & 0 & 1 & 0 \\ 1 & 1 & 3 & 4 \\ 2 & 3 & -2 & 0 \end{pmatrix}$, 求 \mathbf{A}^{-1} 。

练习 43

若矩阵 $\mathbf{A} = \begin{pmatrix} 4 & -2 & 1 \\ 1 & 2 & -2 \\ -1 & 8 & 7 \\ 2 & 14 & -13 \end{pmatrix}$, 求 $r(\mathbf{A})$ 。

练习 44

若矩阵 $\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 & 2 \\ 3 & \lambda & -1 & 2 \\ 5 & 3 & \mu & 6 \end{pmatrix}$ 且 $r(\mathbf{A}) = 2$, 求 $\lambda^2 + \mu^2$ 的值。

练习 45 求下列齐次线性方程组的基础解系 \mathcal{B} 和解空间 S

$$\begin{cases} x_1 - 2x_3 + x_4 = 0 \\ x_2 + x_3 - x_4 = 0 \\ x_1 + 2\sqrt{3}x_2 + x_3 = 0 \end{cases}$$

练习 46 求下列线性方程组的解并使用解空间与特解的和的形式表示解向量

$$\begin{cases} x_1 - 2x_3 + x_4 = -2\sqrt{5} + 1 \\ x_2 + \sqrt{5}x_3 - x_4 = 4 \\ x_1 + 2x_2 + x_3 = \sqrt{5} \end{cases}$$

5.6.2 B 组

练习 47 若 $\lambda \in \mathbb{R}$, 讨论该线性方程组解的结构

$$\begin{cases} \lambda x_1 + x_2 + x_4 = 1 \\ x_1 + \lambda x_2 + x_3 = 1 \\ x_2 + \lambda x_3 + x_4 = 1 \\ x_1 + x_3 + \lambda x_4 = 1 \end{cases}$$

练习 48 证明定理5.1。



第 6 章 特征值理论

「人生若只如初见，何事秋风悲画扇」

——《木兰花·令拟古决绝词》

6.1 特征值与特征向量

6.1.1 特征值

首先我们来定义特征值：

定义 6.1 (特征值 (Eigenvalue))

若存在向量 $v \in V$ ，线性映射算子 $T \in \mathcal{L}(V)$ 与标量 $\lambda \in \mathbb{F}$ 满足

$$T(v) = \lambda v$$

则称 λ 为 T 的特征值。

下面是一个例子：

例题 6.1.

若算子 $T \in \mathcal{L}(\mathbb{F}^2)$ 定义为 $T((x, y)) = (-y, x)$ 求；

- (1) 算子 T 以标准基进行变换的矩阵表示 $\mathcal{M}(T)$ ；
- (2) 当 $\mathbb{F} = \mathbb{R}$ 时，验证 T 是否有特征值，如果有则求之，否则说明理由；
- (3) 当 $\mathbb{F} = \mathbb{C}$ 时，验证 T 是否有特征值，如果有则求之，否则说明理由。

解：

- (1) 算子 T 在标准基下的矩阵表示：

标准基为 $e_1 = (1, 0)$ 和 $e_2 = (0, 1)$ 。计算 T 作用在基向量上的结果：

- $T(e_1) = T((1, 0)) = (0, 1) = e_2$
- $T(e_2) = T((0, 1)) = (-1, 0) = -e_1$

因此，矩阵表示为：

$$\mathcal{M}(T) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

所以答案是：

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

(2) 假设存在实数 λ 和非零向量 $(x, y) \in \mathbb{R}^2$ ，使得：

$$T(x, y) = \lambda(x, y) \implies (-y, x) = (\lambda x, \lambda y)$$

这等价于方程组：

$$\begin{cases} -y = \lambda x \\ x = \lambda y \end{cases}$$

1. 从第一式得 $y = -\lambda x$ ，代入第二式：

$$x = \lambda(-\lambda x) = -\lambda^2 x$$

2. 整理得 $x(1 + \lambda^2) = 0$ 。

- 若 $x \neq 0$ ，则 $1 + \lambda^2 = 0$ 。但在 \mathbb{R} 中无解。
- 若 $x = 0$ ，则第一式给出 $y = 0$ ，与 v 非零矛盾。

所以在实数域 \mathbb{R} 中， T 没有特征值

3. 假设存在复数 λ 和非零向量 $(x, y) \in \mathbb{C}^2$ ，使得：

$$T(x, y) = \lambda(x, y) \implies (-y, x) = (\lambda x, \lambda y)$$

方程组仍为：

$$\begin{cases} -y = \lambda x \\ x = \lambda y \end{cases}$$

(a). 从第一式得 $y = -\lambda x$ ，代入第二式：

$$x = \lambda(-\lambda x) = -\lambda^2 x$$

(b). 整理得 $x(1 + \lambda^2) = 0$ 。

- 若 $x \neq 0$ ，则 $1 + \lambda^2 = 0 \implies \lambda = \pm i$ (复数解)。
- 若 $x = 0$ ，则 $y = -\lambda \cdot 0 = 0$ ，与 v 非零矛盾。

(c). 验证 $\lambda = i$ 和 $\lambda = -i$ 对应的非零向量：

- 对 $\lambda = i$ ，由 $y = -ix$ ，特征向量为 $(x, -ix)$ ($x \neq 0$)。
- 对 $\lambda = -i$ ，由 $y = ix$ ，特征向量为 (x, ix) ($x \neq 0$)。

所以在复数域 \mathbb{C} 中， T 的特征值为 i 和 $-i$ 。

根据上面的定义和例子，我们了解到特征值并不是唯一且一定拥有的，并且我们要求 $v \neq \mathbf{0}$ 这是因为每个标量 $\lambda \in \mathbb{F}$ 都满足 $T(\mathbf{0}) = \lambda\mathbf{0}$ 。

根据一般的算子矩阵 T ，下面推导矩阵的特征方程：设 \mathbf{A} 是一个 $n \times n$ 方阵， $\mathbf{A} = \mathcal{M}(T)$ ， λ 为 \mathbf{A} 的特征值，对应的非零特征向量 v 满足：

$$\mathbf{A}v = \lambda v$$

将等式改写为：

$$\mathbf{A}v - \lambda v = \mathbf{0}$$



引入单位矩阵 \mathbf{I} 以保持维度一致:

$$(\mathbf{A} - \lambda\mathbf{I})v = \mathbf{0}$$

则非零解 v 存在的条件是系数矩阵的行列式为零:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

上述行列式方程即为矩阵 \mathbf{A} 的特征方程, 展开后得到一个关于 λ 的多项式。

推论 6.1

关于线性映射算子 T 的矩阵表示 $\mathcal{M}(T)$ 的特征值 λ 满足

$$\det(\mathcal{M}(T) - \lambda\mathbf{I}) = 0$$



例题 6.2.

对于 2×2 矩阵 $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, 求其特征方程, 并求 $a + b = 3$, $ad - bc = 2$ 时, 矩阵 \mathbf{A} 的特征值.

解: 对于 2×2 矩阵 $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 特征方程为

$$\det \left(\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix} \right) = (a - \lambda)(d - \lambda) - bc = 0$$

展开后得到:

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0$$

带入 $a + b = 3$, $ad - bc = 2$ 可得方程

$$\lambda^2 - 3\lambda + 2 = 0$$

因式分解可得 $(\lambda - 1)(\lambda - 2) = 0$ 解得 $\lambda = 1, 2$ 故矩阵 \mathbf{A} 的特征值为 1 和 2.

对于等式 $T(v) = \lambda v$ 接下来我们逐步引出特征向量的概念, 实际上这里的 v 就是特征向量。

6.1.2 特征向量

定义 6.2 (特征向量 (Eigenvector))

设线性映射算子 $T \in \mathcal{L}(V)$ 其中 V 为一个线性子空间, $\lambda \in \mathbb{F}$ 是算子 T 的特征值, 若能满足等式 $T(v) = \lambda v$ 且 $v \neq 0, v \in V$ 则称 v 为 T 相对于 λ 的特征向量。



下面我们介绍如何求解特征值所对应的特征向量, 实际上它们与解线性方程组的方式一致, 接下来请大家带着一个问题去看下面的例题, 即:

思考

特征向量是只有一个向量构成集合还是有很多个向量构成的集合?



例题 6.3.

算子 $T \in \mathcal{L}(\mathbb{R}^2)$ 满足 $T((x, y, z)) = (y + z, x + z, x + y)$ 求;

- (1) 算子 T 以标准基进行变换的矩阵表示 $\mathcal{M}(T)$;
- (2) 求算子 T 的所有特征值, 并求出每个特征值对应的特征向量。

解: (1) 算子 T 在标准基下的矩阵表示:

标准基为 $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ 。计算每个基向量在 T 作用下的结果:

- $T(e_1) = (0, 1, 1)$
- $T(e_2) = (1, 0, 1)$
- $T(e_3) = (1, 1, 0)$

因此, 矩阵 $\mathcal{M}(T)$ 为:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

(2) 算子 T 的特征值及对应的特征向量:

首先求解特征方程 $\det(\mathcal{M}(T) - \lambda I) = 0$, 其中矩阵 $\mathcal{M}(T) - \lambda I$ 为:

$$\begin{pmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{pmatrix}$$

计算行列式:

$$\det(\mathcal{M}(T) - \lambda I) = -\lambda^3 + 3\lambda + 2 = 0$$

解得特征值为 $\lambda = 2$ 和 $\lambda = -1$ (二重根)。

- 对于 $\lambda = 2$, 解方程组 $(T - 2I)v = 0$, 得到特征向量为所有非零标量倍的 $(1, 1, 1)$ 即
- 对于 $\lambda = -1$, 解方程组 $(T + I)v = 0$, 得到特征向量为满足 $x + y + z = 0$ 的所有非零向量, 例如 $(1, -1, 0)$ 和 $(1, 0, -1)$ 。

计算行列式:

$$\det(\mathcal{M}(T) - \lambda I) = -\lambda^3 + 3\lambda + 2 = 0$$

解得特征值为 $\lambda = 2$ 和 $\lambda = -1$ (二重根)。

1. 当特征值 $\lambda = 2$ 时

$$\mathcal{M}(T) - 2I = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$$

对其进行使用初等变换

$$\begin{array}{c} \left(\begin{array}{ccc} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{array} \right) \xrightarrow{r_1 \leftrightarrow r_2} \left(\begin{array}{ccc} 1 & -2 & 1 \\ -2 & 1 & 1 \\ 1 & 1 & -2 \end{array} \right) \xrightarrow{r_2 + 2 \times r_1} \left(\begin{array}{ccc} 1 & -2 & 1 \\ 0 & -3 & 3 \\ 1 & 1 & -2 \end{array} \right) \xrightarrow{r_3 - r_1} \left(\begin{array}{ccc} 1 & -2 & 1 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{array} \right) \\ \xrightarrow{r_3 + r_2} \left(\begin{array}{ccc} 1 & -2 & 1 \\ 0 & -3 & 3 \\ 0 & 0 & 0 \end{array} \right) \xrightarrow{r_2 / 3} \left(\begin{array}{ccc} 1 & -2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{array} \right) \end{array}$$

还原为齐次线性方程组

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ -x_2 + x_3 = 0 \end{cases}$$

确定基础解系，设 $x_1 = 1$ 则 $x_2 = 1, x_3 = 1$ 由于 $r(\mathcal{M}(T)) = 2$ 所以 $n - r = 1$ 故基础解系中一个 \mathcal{B} 共一个向量即 $\mathcal{B} = \{(1, 1, 1)\}$ 所以当特征值 $\lambda = 2$ 时，特征向量为 $v = k(1, 1, 1), k \in \mathbb{R}$

2. 当特征值 $\lambda = -1$ 时

$$\mathcal{M}(T) + I = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

所有行相同，化为：

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

其对应的方程为

$$x + y + z = 0$$

最后令 $x = 0, y = 1$ 则 $z = -1$ 令 $x = 1, y = 0$ 则 $z = -1$ 那么基础解系 $\mathcal{B} = \{(0, 1, -1), (1, 0, -1)\}$ 特征向量为 $v = s(-1, 1, 0) + t(-1, 0, 1)$

接下来回答上述的思考题：特征向量是只有一个向量构成集合还是有很多个向量构成的集合？答案是很多个向量构成的集合，因为最后此类问题求解特征向量都会转化为求齐次线性方程组，然而齐次线性方程组要么全为 0 解，要么有无数个非零解，全为 0 解的时，特征值为 0 或不存在故不存在特征向量。

6.2 迹与上三角矩阵

6.2.1 矩阵的迹

首先我们先定义矩阵的迹。

定义 6.3 (矩阵的迹 (trace))

$n \times n$ 矩阵的迹 (trace) 是矩阵对角元素的和，通常表示为 $\text{Tr}(\mathbf{A})$ ；若方阵 $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ ，则

$$\text{Tr}(\mathbf{A}) := \sum_{i=1}^n a_{ii}$$

即为 \mathbf{A} 的所有对角元的和。

下面是有关迹的一些推论：



推论 6.2

设 \mathbf{A} 是 $m \times n$ 矩阵, \mathbf{B} 是 $n \times m$ 矩阵。证明:

$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$$



证明 考虑 \mathbf{AB} 的迹, 其定义为:

$$\text{Tr}(\mathbf{AB}) = \sum_{i=1}^m (\mathbf{AB})_{ii} = \sum_{i=1}^m \sum_{k=1}^n a_{ik} b_{ki}$$

同样地, 考虑 \mathbf{BA} 的迹, 其定义为:

$$\text{Tr}(\mathbf{BA}) = \sum_{k=1}^n (\mathbf{BA})_{kk} = \sum_{k=1}^n \sum_{i=1}^m b_{ki} a_{ik}$$

接下来, 我们交换双重求和的顺序。对于 $\text{Tr}(\mathbf{AB})$, 我们可以将求和顺序交换为:

$$\sum_{i=1}^m \sum_{k=1}^n a_{ik} b_{ki} = \sum_{k=1}^n \sum_{i=1}^m a_{ik} b_{ki}$$

对于 $\text{Tr}(\mathbf{BA})$, 我们注意到 $b_{ki} a_{ik} = a_{ik} b_{ki}$, 因此:

$$\sum_{k=1}^n \sum_{i=1}^m b_{ki} a_{ik} = \sum_{k=1}^n \sum_{i=1}^m a_{ik} b_{ki}$$

显然, $\text{Tr}(\mathbf{AB})$ 和 $\text{Tr}(\mathbf{BA})$ 的表达式相同, 只是求和顺序不同, 因此它们的值相等。

通过具体例子验证, 例如当 \mathbf{A} 是行向量, \mathbf{B} 是列向量时, 或者当 \mathbf{A} 和 \mathbf{B} 分别为不同维度的矩阵时, 它们的迹仍然相等。

综上所述, 我们证明了:

$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$$



推论 6.3

设 n 阶方阵 \mathbf{A} 的特征方程为多项式 $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n$, 则

$$a_1 = -\text{Tr}\mathbf{A}$$



证明 特征方程多项式可以表示为 $P(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$, 其中 \mathbf{I} 是单位矩阵。设 \mathbf{A} 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则特征方程多项式可以分解为: 根据韦达定理, λ^{n-1} 项的系数由所有特征值的和决定, 即:

$$a_1 = -(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$



6.2.2 上三角矩阵

首先我们先定义上三角矩阵。

定义 6.4 (上三角矩阵 (upper-triangular matrix))

若方阵 \mathbf{A} 主对角线下方的元素全为 0 则 \mathbf{A} 为上三角矩阵。例如

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}$$

该矩阵蓝色标注的为矩阵对角线。



在此我们引入上三角矩阵，是为了研究基于算子的某一个基的表示为上三角矩阵可以快速确定一些特征值的性质，以及为了后面的一些定理做一些铺垫。

线性映射算子 $T \in \mathcal{L}(V)$ 的矩阵 $\mathcal{M}(T)$ 依赖于其基的选取，我们设空间 V 的一个基 $S = \{v_1, v_2, v_3, \dots, v_n\}$ ，则 $\dim V = n$ ，不难看出选取基 S 的矩阵 $\mathcal{M}(T)$ 为一个 $n \times n$ 矩阵，下面我们看一个例子：

若 $T \in \mathcal{L}(\mathbb{F}^3)$ 表示为 $T((x, y, z)) = (x + y + 3z, 3y + 2z, 4z)$ 则选取 \mathbb{F}^3 的标准基 $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ 的矩阵表示为

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 4 \end{pmatrix}$$

事实上，在本章节我们会强调不是以标准基的线性映射矩阵，接下来我们约定一个记号，上述的例子我们用 $\mathcal{M}(T)$ 来默认表示它们是基于标准基的变换矩阵。下面给出基于非标准基的线性映射的矩阵相关记号与定义。

定义 6.5 (基于非标准基的线性映射的矩阵)

若线性映射算子 $T \in \mathcal{V}$ ，若 S 为线性空间 V 的一个基，那么基于 S 的线性映射 T 的矩阵矩阵表示为 $\mathcal{M}(T, S)$ ，若 S 为线性空间 V 的标准基，则简写为 $\mathcal{M}(T)$



接下来根据上述知识我们可以得到一些推论：

推论 6.4

若算子 $T \in \mathcal{L}(\mathbb{C}^n)$ 则一定存在基 S 使得算子关于该基的矩阵 $\mathcal{M}(T, S)$ 为上三角矩阵。



证明 为了证明任何线性算子 $T \in \mathcal{L}(\mathbb{C}^n)$ 存在一个基 S ，使得其矩阵表示为上三角矩阵，我们可以使用数学归纳法：当 $n = 1$ 时，任何 1×1 矩阵本身就是上三角矩阵，结论显然成立。

那么假设对于所有 $n - 1$ 维复向量空间上的线性算子，存在一个基使得其矩阵为上三角矩阵。设 V 为 n 维复向量空间， $T \in \mathcal{L}(V)$ 。由于 V 是复空间， T 必有特征值 λ_1 ，对应非零特征向量 v_1 。令 $U = \text{Span}(v_1)$ ，则 U 是 T -不变的一维子空间。考虑线性映射 $T - \lambda_1 I$ ，根据秩零化度定理：

$$\dim V = \dim \text{null}(T - \lambda_1 I) + \dim \text{range}(T - \lambda_1 I)$$

由于 $\dim \text{null}(T - \lambda_1 I) \geq 1$ ，得 $\dim \text{range}(T - \lambda_1 I) \leq n - 1$ 。选取 V 的补空间 W 使得 $V = U \oplus W$ 。这里 W 是 $n - 1$ 维子空间。定义投影 $\pi : V \rightarrow W$ ，将任意 $v \in V$ 分解为 $v = u + w$ （其中 $u \in U, w \in W$ ），并令 $\pi(v) = w$ 。定义线性算子 $S : W \rightarrow W$ 为 $S(w) = \pi(T(w))$ 。由线性性保证， S 是 W 上的线性算子。根据归纳假设，存在 W 的一组基 v_2, \dots, v_n ，使得 S 在此基下的矩阵为上三角形式。即对任意 $2 \leq j \leq n$ ，有：

$$S(v_j) \in \text{Span}(\{v_2, \dots, v_j\}).$$

合并基 v_1, v_2, \dots, v_n 。对每个 v_j ($j \geq 2$)，有：

$$T(v_j) = \pi(T(v_j)) + u_j \quad (u_j \in U).$$

由于 $\pi(T(v_j)) = S(v_j) \in \text{Span}(\{v_2, \dots, v_j\})$ ，且 $u_j \in \text{Span}(\{v_1\})$ ，因此：

$$T(v_j) \in \text{Span}(\{v_1, v_2, \dots, v_j\}).$$

对于 v_1 ，显然 $T(v_1) = \lambda_1 v_1 \in \text{Span}(\{v_1\})$ 。因此，基 $\{v_1, v_2, \dots, v_n\}$ 使得 $\mathcal{M}(T, S)$ 为上三角矩阵。

由数学归纳法，复向量空间上的任意线性算子均存在一组基，使其矩阵表示为上三角矩阵。即，存在基 S 使得 $\mathcal{M}(T, S)$ 为上三角矩阵。



6.3 相似变换

6.3.1 相似矩阵

首先给出矩阵相似的定义。

定义 6.6 (相似矩阵)

若 $n \times n$ 矩阵 \mathbf{A} 与矩阵 \mathbf{B} 相似，则存在 $n \times n$ 可逆矩阵 \mathbf{P} 使得

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

记作 $\mathbf{A} \sim \mathbf{B}$



根据上述定义，我们不难发现如下的推论：

推论 6.5

相似矩阵的性质：

1. 自反性： \mathbf{A} 与 \mathbf{A} 相似；
2. 对称性：若 \mathbf{A} 与 \mathbf{B} 相似，则 \mathbf{B} 与 \mathbf{A} 相似；
3. 传递性：若 \mathbf{A} 与 \mathbf{B} 相似， \mathbf{B} 与 \mathbf{C} 相似，则 \mathbf{A} 与 \mathbf{C} 相似



证明 证明留给读者，参见课后练习 B 组。

□

推论 6.6

如果矩阵 $\mathbf{A} \sim \mathbf{B}$ ，则它们的特征方程 $p(\lambda)$ 相同，并满足 $\text{Tr}\mathbf{A} = \text{Tr}\mathbf{B}$ 与 $\det \mathbf{A} = \det \mathbf{B}$



证明 假设矩阵 \mathbf{A} 和 \mathbf{B} 相似，则存在可逆矩阵 \mathbf{P} 使得 $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 。

首先，证明它们的特征方程相同：

计算矩阵 \mathbf{B} 的特征方程多项式：

$$\det(\lambda\mathbf{I} - \mathbf{B}) = \det(\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P})$$

将 $\lambda\mathbf{I}$ 改写为 $\mathbf{P}^{-1}(\lambda\mathbf{I})\mathbf{P}$ ，因为 $\mathbf{P}^{-1}(\lambda\mathbf{I})\mathbf{P} = \lambda\mathbf{I}$ 。于是：

$$\det(\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}) = \det(\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P})$$

利用行列式的乘法性质：

$$\det(\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}) = \det(\mathbf{P}^{-1}) \det(\lambda\mathbf{I} - \mathbf{A}) \det(\mathbf{P})$$

由于 $\det(\mathbf{P}^{-1}) \det(\mathbf{P}) = \det(\mathbf{P}^{-1}\mathbf{P}) = \det(\mathbf{I}) = 1$ ，因此：

$$\det(\lambda\mathbf{I} - \mathbf{B}) = \det(\lambda\mathbf{I} - \mathbf{A})$$

即 \mathbf{A} 和 \mathbf{B} 的特征方程相同。

接下来，由于特征方程相同，它们的特征方程多项式相同。特征方程多项式的形式为：

$$\lambda^n - (\text{Tr}\mathbf{A})\lambda^{n-1} + \cdots + (-1)^n \det \mathbf{A}^1$$

比较多项式两边的对应系数：

第二项系数²： $-\text{Tr}\mathbf{A} = -\text{Tr}\mathbf{B}$ ，因此 $\text{Tr}\mathbf{A} = \text{Tr}\mathbf{B}$ 。常数项： $(-1)^n \det \mathbf{A} = (-1)^n \det \mathbf{B}$ ，因此 $\det \mathbf{A} = \det \mathbf{B}$ 。

因此，矩阵 \mathbf{A} 和 \mathbf{B} 的迹和行列式相等。

¹为什么最后一项是 $(-1)^n \det \mathbf{A}$ ？

²根据推论6.3

□

推论 6.7

任意的 n 阶方阵 \mathbf{A} 都能相似于一个上三角矩阵。



证明 设 \mathbf{A} 是任意 $n \times n$ 复矩阵。 \mathbf{A} 可视为线性算子 $T_A : \mathbb{C}^n \rightarrow \mathbb{C}^n$ 在标准基 B 下的矩阵表示, 即 $\mathcal{M}(T_A, B) = \mathbf{A}$ 。由推论6.4证明, 存在基 S 使得 $\mathcal{M}(T_A, S)$ 是上三角矩阵。基变换矩阵 \mathbf{P} 满足 \mathbf{P} 的列是 S 在基 B 下的坐标向量, 且

$$\mathcal{M}(T_A, S) = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}.$$

因此, $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ 是上三角矩阵, 即 \mathbf{A} 相似于一个上三角矩阵。

□

6.3.2 Hamilton-Cayley 定理

根据上述的铺垫我们引入并证明 Hamilton-Cayley 定理

定理 6.1 (Hamilton-Cayley 定理)

多项式 $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n$ 为 n 阶方阵 \mathbf{A} 的特征方程, 那么矩阵运算

$$P(\mathbf{A}) := \mathbf{A}^n + a_1 \mathbf{A}^{n-1} + a_2 \mathbf{A}^{n-2} + \cdots + a_n \mathbf{I} = \mathbf{O}$$

成立, $P(\mathbf{A})$ 为零矩阵 \mathbf{O} 。



这个定理挺奇妙, 我们下面尝试一步步证明它。

证明 特征方程多项式将其可以改写为

$$P(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

根据推论6.7, 存在可逆矩阵 \mathbf{P} , 使得

$$\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \begin{pmatrix} \lambda_1 & * & * & \cdots & * \\ & \lambda_2 & * & \cdots & * \\ & & \lambda_3 & \cdots & * \\ & & & \ddots & \vdots \\ & & & & \lambda_n \end{pmatrix}$$

将 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ 代入矩阵运算即

$$P(\mathbf{P}^{-1} \mathbf{A} \mathbf{P}) = (\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_1 \mathbf{I})(\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_2 \mathbf{I}) \cdots (\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_n \mathbf{I})$$

我们将其写成矩阵乘积的形式即

$$\begin{pmatrix} 0 & * & * & \cdots & * \\ \lambda_2 - \lambda_1 & * & * & \cdots & * \\ \lambda_3 - \lambda_1 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_n - \lambda_1 & & & & \end{pmatrix} \begin{pmatrix} \lambda_1 - \lambda_2 & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \lambda_3 - \lambda_2 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_n - \lambda_2 & & & & \end{pmatrix} \cdots \begin{pmatrix} \lambda_1 - \lambda_n & * & * & \cdots & * \\ \lambda_2 - \lambda_n & * & * & \cdots & * \\ \lambda_3 - \lambda_n & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & & & & \end{pmatrix} = \mathbf{O}$$

由于单位矩阵 $\mathbf{I} = \mathbf{P} \mathbf{P}^{-1}$ 可得

$$P(\mathbf{P}^{-1} \mathbf{A} \mathbf{P}) = (\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_1 \mathbf{P} \mathbf{P}^{-1})(\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_2 \mathbf{P} \mathbf{P}^{-1}) \cdots (\mathbf{P}^{-1} \mathbf{A} \mathbf{P} - \lambda_n \mathbf{P} \mathbf{P}^{-1})$$

提出 $\mathbf{P} \mathbf{P}^{-1}$ 部分可得 $\mathbf{P} \mathbf{P}(\mathbf{A}) \mathbf{P}^{-1} = \mathbf{O}$ 即 $P(\mathbf{A}) = \mathbf{O}$

□

Hamilton-Cayley 定理在很多方面也有应用, 例如求矩阵多项式, 求逆矩阵, 伴随矩阵; 下面我们就求逆矩阵讲讲如何利用 Hamilton-Cayley 定理求解。



设 $n \times n$ 矩阵 \mathbf{A} 可逆，首先我们根据 $n \times n$ 可逆矩阵的秩等于 n ，可得其特征方程多项式为

$$P(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n$$

且 $a_n = (-1)^n \det \mathbf{A} \neq 0$ 由 Hamilton-Cayley 定理，我们有等式

$$\mathbf{A}^n + a_1\mathbf{A}^{n-1} + a_2\mathbf{A}^{n-2} + \cdots + a_n\mathbf{I} = \mathbf{O}$$

化简这个式子我们可以得到

$$-\frac{1}{a_n} (\mathbf{A}^n + a_1\mathbf{A}^{n-1} + a_2\mathbf{A}^{n-2} + \cdots + a_{n-1}\mathbf{A}) = \mathbf{I}$$

由此可得

$$\mathbf{A}^{-1} = -\frac{1}{a_n} (\mathbf{A}^{n-1} + a_1\mathbf{A}^{n-2} + a_2\mathbf{A}^{n-3} + \cdots + a_{n-1})$$

例题 6.4.

使用 Hamilton-Cayley 定理计算方阵 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix}$ 的逆 \mathbf{A}^{-1}

解： \mathbf{A} 的特征方程多项式为

$$P(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = -\lambda^3 + 3\lambda^2 + 5\lambda - 11$$

所以可得其逆矩阵

$$\mathbf{A}^{-1} = \frac{1}{11} (-\mathbf{A}^2 + 3\mathbf{A} + 5\mathbf{I}) = \frac{1}{11} \begin{pmatrix} 1 & 2 & 6 \\ 2 & 7 & -1 \\ 2 & -4 & -1 \end{pmatrix}$$

6.3.3 对角化矩阵

首先给出对角矩阵的定义：

定义 6.7 (对角矩阵 (Diagonal Matrix) 与矩阵的对角化)

若 $n \times n$ 方阵 \mathbf{A} 的元素 $a_{ij} = c_i \delta_{ij}$ ，其中 $c_i \in \mathbb{F}$ ， δ_{ij} 为函数表示为

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

则该矩阵为对角矩阵，即该矩阵的元素除了其对角线上的元素以外均为 0，并记作 $\text{diag}(a_1, a_2, \dots, a_n)$ ，即

$$\text{diag}(a_1, a_2, \dots, a_n) := \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}$$

如果方阵 \mathbf{B} 与一个对角矩阵相似，则称 \mathbf{B} 称为可对角化。

^aKronecker Delta, Kronecker function, 克罗内克函数, <https://mathworld.wolfram.com/KroneckerDelta.html>



若方阵 $\mathbf{A} \sim \text{diag}(a_1, a_2, \dots, a_n)$ ，则 \mathbf{A} 的特征方程 $p(\lambda) = (\lambda - a_1)(\lambda - a_2) \cdots (\lambda - a_n)$ 因此这个对角矩阵的元素即为 \mathbf{A} 的特征值。下面我们来描述矩阵可对角化的充分必要条件，若 \mathbf{A} 可对角化，那么则有式子成立

$$\mathbf{A} = \mathbf{P} \text{diag}(a_1, a_2, \dots, a_n) \mathbf{P}^{-1}$$



设 \mathbf{P} 矩阵由列向量 $\{b_1, b_2, b_3, \dots, b_n\}$ 构成, 则 $\mathbf{P} = (b_1 \ b_2 \ b_3 \ \dots \ b_n)$ 由于 \mathbf{P} 可逆, 所以其秩 $r(\mathbf{P}) = n$, 因此 $\{b_1, b_2, b_3, \dots, b_n\}$ 为其极大线性无关组, 那么就可以写作下面的形式

$$\mathbf{A} (b_1 \ b_2 \ b_3 \ \dots \ b_n) = (b_1 \ b_2 \ b_3 \ \dots \ b_n) \operatorname{diag}(a_1, a_2, \dots, a_n)$$

设矩阵 $\mathbf{B} = (b_1 \ b_2 \ \dots \ b_n)$, 其列由向量 b_1, b_2, \dots, b_n 组成。对角矩阵 $\mathbf{D} = \operatorname{diag}(a_1, a_2, \dots, a_n)$ 。则等式可写为:

$$\mathbf{AB} = \mathbf{BD}$$

- 左边 \mathbf{AB} 的结果是一个矩阵, 其第 j 列为 \mathbf{Ab}_j , 即:

$$\mathbf{AB} = (\mathbf{Ab}_1 \ \mathbf{Ab}_2 \ \dots \ \mathbf{Ab}_n)$$

- 右边 \mathbf{BD} 的结果是一个矩阵, 其第 j 列为 \mathbf{B} 的第 j 列乘以 \mathbf{D} 的第 j 个对角元素 a_j , 即:

$$\mathbf{BD} = (a_1 b_1 \ a_2 b_2 \ \dots \ a_n b_n)$$

因此, 等式等价于:

$$(\mathbf{Ab}_1 \ \mathbf{Ab}_2 \ \dots \ \mathbf{Ab}_n) = (a_1 b_1 \ a_2 b_2 \ \dots \ a_n b_n)$$

这表示对应列相等, 即对于每个 $j = 1, 2, \dots, n$:

$$\mathbf{Ab}_j = a_j b_j$$

该等式化简后表示: 每个列向量 b_i 是矩阵 \mathbf{A} 的特征向量, 对应的特征值为 a_i 。即:

$$\mathbf{Ab}_i = a_i b_i, \quad \forall i = 1, 2, \dots, n$$

因此我们可以得到矩阵 $\mathbf{P} = (b_1 \ b_2 \ b_3 \ \dots \ b_n)$ 的所有元素均为 \mathbf{A} 的特征向量排列形成的矩阵。

例题 6.5.

矩阵 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 2 & 1 \end{pmatrix}$, 存在 $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$ 且 \mathbf{B} 为对角矩阵, 求矩阵 \mathbf{P} 的一个可能值。

解: \mathbf{A} 特征方程为

$$\det(\lambda\mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda - 2 & -1 & 0 \\ 0 & \lambda - 3 & 0 \\ 0 & -2 & \lambda - 1 \end{vmatrix} = (\lambda - 1)(\lambda - 2)(\lambda - 3)$$

令其等于 0 可得 \mathbf{A} 有 3 个不同的特征值, 分别为 1, 2, 3。接下来分别代入 3 个特征值求方阵的特征向量:

1. 当 $\lambda = 1$ 时

$$\lambda\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -2 & 0 \\ 0 & -2 & 0 \end{pmatrix}$$

由此 $(\lambda\mathbf{I} - \mathbf{A})v = \mathbf{0}$ 通过高斯消元法解线性方程组

$$\begin{pmatrix} -1 & -1 & 0 \\ 0 & -2 & 0 \\ 0 & -2 & 0 \end{pmatrix} \xrightarrow{\frac{r_2/2}{r_3/2}} \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \end{pmatrix} \xrightarrow{\frac{r_1-r_2}{r_3-r_2}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

还原为齐次方程为

$$\begin{cases} -x_1 = 0 \\ -x_2 = 0 \end{cases}$$

其中 x_3 为自由变量, 所以 $\lambda = 1$ 其一个特征向量为 $(0, 0, 1)$

2. 同理我们算出 $\lambda = 2$ 时其一个特征向量为 $(1, 0, 0)$

3. 同理我们算出 $\lambda = 3$ 时其一个特征向量为 $(1, 1, 1)$

以特征向量为列构造 \mathbf{P} , 按特征值顺序排列:

$$\mathbf{P} = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

计算行列式:

$$\det(\mathbf{P}) = \det \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = 1 \cdot \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1 \cdot (1 \cdot 1 - 1 \cdot 0) = 1 \neq 0$$

行列式非零, 故 \mathbf{P} 可逆。因此, 矩阵 \mathbf{P} 的一个可能值为:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

事实上, 不一定每一个方阵都可以被对角化, 下面给出可对角化方阵的充分条件:

定理 6.2 (方阵可对角化的充分条件)

如果 \mathbb{F} 上 n 阶方阵 \mathbf{A} 具有 n 个互不相同的特征值, 则 \mathbf{A} 可以对角化。



证明 对每个特征值 λ_i , 选取对应的特征向量 v_i , 即 $\mathbf{A}v_i = \lambda_i v_i$, 且 $v_i \neq \mathbf{0}$ 。考虑向量组 $\{v_1, v_2, \dots, v_n\}$ 。使用数学归纳法:

1. 当 $k = 1$ 时, 单个特征向量 $v_1 \neq \mathbf{0}$, 故线性无关。

2. 假设对 $k - 1$ 个不同特征值的特征向量成立。考虑 k 个不同特征值 $\lambda_1, \dots, \lambda_k$ 的特征向量 v_1, \dots, v_k 。设线性组合:

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0} \quad (1)$$

作用 \mathbf{A} 得:

$$c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_k \lambda_k v_k = \mathbf{0} \quad (2)$$

将 (1) 乘以 λ_k :

$$\lambda_k c_1 v_1 + \lambda_k c_2 v_2 + \cdots + \lambda_k c_k v_k = \mathbf{0} \quad (3)$$

(2) 减 (3):

$$c_1 (\lambda_1 - \lambda_k) v_1 + c_2 (\lambda_2 - \lambda_k) v_2 + \cdots + c_{k-1} (\lambda_{k-1} - \lambda_k) v_{k-1} = \mathbf{0}$$

由归纳假设, v_1, \dots, v_{k-1} 线性无关, 且 $\lambda_i \neq \lambda_k$ (特征值互异), 故 $c_i (\lambda_i - \lambda_k) = 0$ 推出 $c_i = 0$ ($i = 1, 2, \dots, k - 1$)。代入 (1) 得 $c_k v_k = \mathbf{0}$, 因 $v_k \neq \mathbf{0}$, 故 $c_k = 0$ 。因此, v_1, \dots, v_k 线性无关。

由归纳假设, v_1, \dots, v_{k-1} 线性无关, 且 $\lambda_i \neq \lambda_k$ (特征值互异), 故 $c_i (\lambda_i - \lambda_k) = 0$ 推出 $c_i = 0$ ($i = 1, 2, \dots, k - 1$)。代入 (1) 得 $c_k v_k = \mathbf{0}$, 因 $v_k \neq \mathbf{0}$, 故 $c_k = 0$ 。因此, v_1, \dots, v_k 线性无关。

在本问题中, $k = n$, 故 $\{v_1, v_2, \dots, v_n\}$ 线性无关。由于 \mathbf{A} 是 n 阶方阵, 特征向量组构成 \mathbb{F}^n 的一组基。令 $\mathbf{P} = (v_1, v_2, \dots, v_n)$, 则 \mathbf{P} 可逆。计算:

$$\mathbf{AP} = \mathbf{A}(v_1, v_2, \dots, v_n) = (\mathbf{Av}_1, \mathbf{Av}_2, \dots, \mathbf{Av}_n) = (\lambda_1 v_1, \lambda_2 v_2, \dots, \lambda_n v_n)$$

令 $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 则:

$$\mathbf{P}\mathbf{D} = (v_1, v_2, \dots, v_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = (\lambda_1 v_1, \lambda_2 v_2, \dots, \lambda_n v_n)$$

故 $\mathbf{AP} = \mathbf{PD}$, 即 $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$, 其中 \mathbf{D} 为对角矩阵。

因此, \mathbf{A} 可对角化。

□

6.4 Jordan 标准型

6.4.1 Jordan 块

在这里我们引入 Jordan³ 块与 Jordan 标准型的概念, 由于不是每一个方阵都可以相似变化为对角矩阵, 但是可以相似为一个与之很近的形式。

定义 6.8 (Jordan 块 (Jordan Block))

除对角线和超对角线外, 其他位置均为零, 对角线的每个元素均由单个数字 λ 组成, 对角线上方的每个元素均由 1 组成的 m 阶方阵为 Jordan 块, 记作 $J_m(\lambda)$ 例如:

$$J_m(\lambda) := \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & \ddots & & \\ & & \ddots & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

该矩阵的空余元素均为 0。此外 1×1 矩阵的退化情况被视为 Jordan 块, 即使它缺少可填充 1 的上对角线。



6.4.2 代数重数

首先我们举一个最简单的 Jordan 块:

$$J_3(2) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

我们看看是否可以将其对角化, 按照我们之前的过程, 首先是计算该方阵的特征值, 即计算:

$$J_3(2) - \lambda \mathbf{I} = \begin{pmatrix} 2 - \lambda & 1 & 0 \\ 0 & 2 - \lambda & 1 \\ 0 & 0 & 2 - \lambda \end{pmatrix}$$

该矩阵为上三角矩阵, 其行列式为对角线元素的乘积:

$$\det(J_3(2) - \lambda \mathbf{I}) = (2 - \lambda)^3$$

³Jordan 法语读音 /ʒɔ̃dã/ 音译为约旦, 若尔当, 英语读音 /dʒɔ:dən/ 音译为乔丹, 焦尔丹。实际上两种读法都有, 但是由于 Jordan Marie Ennemond Camille 是法国人, 所以我们中文常读作: “若尔当块”而非“乔丹块”

最后解出 $\lambda_{1,2,3} = 2$; 其次计算特征向量我们可以得到所有的特征向量均为 $(1, 0, 0)$ 构成矩阵为

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

由此可得 \mathbf{P} 不可逆, 我们最后得到 $J_3(2)$ 不可以被对角化。那这个时候读者可能会问, 矩阵可对角化的充要条件是什么, 我们上面的定理6.2只写到了充分条件, 接下来我们引入代数重数和几何重数的例子。

首先作为特征方程多项式, 展开后一定可以得到一个代数多项式并将其分解为

$$P(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

例如我们刚刚举的例子 $J_3(2)$ 的特征方程多项式为

$$\det(J_3(2) - \lambda \mathbf{I}) = (2 - \lambda)^3$$

那么我们称特征值 2 的代数重数为 3。

定义 6.9 (代数重数)

n 阶方阵特征方程多项式最终可表示为

$$P(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{t_i}, \sum_{i=1}^k t_i = n$$

且满足 $\forall a \neq b \implies \lambda_a \neq \lambda_b$ 则特征值 λ_i 的代数重数为 t_i , 记为

$$m(\lambda_i) := t_i$$



6.4.3 几何重数与不变子空间

那么几何重数则表示该特征值的几个特征向量的极大线性无关组的个数, 也可以指的是这几个特征向量张成的空间的维数。接下里我们把“几个特征向量张成的空间”称作不变子空间。

定义 6.10 (不变子空间; 几何重数)

设线性映射算子 $T \in \mathcal{L}(V)$, λ_0 为 T 的一个特征值, 定义集合 V_{λ_0} 表示所有特征向量 x 可能值的集合, 即:

$$V_{\lambda_0} := \{x \mid T(x) = \lambda_0 x\}$$

称为特征值 λ_0 的不变子空间; 其几何重数为 $\dim(V_{\lambda_0})$, 记为

$$g(\lambda_i) := \dim(V_{\lambda_0})$$



关于代数重数与几何重数, 有以下推论:

推论 6.8

特征值的几何重数 \leq 代数重数.



证明 我们不作严格证明, 在计算特征方程的时候我们总是会得到一个未知数数量与式子数量均不超过 n 的齐次线性方程组, 所以其基础解系的维度不会超过 n , 所以几何重数小于其代数重数。



接下来是一个重要的定理, 它揭示了方阵可对角化的充要条件。

定理 6.3 (方阵可对角化的充要条件)

矩阵可对角化等价于其所有特征值的代数重数等于几何重数。



证明 本定理不作严格证明, 只做一个简洁的说明; 首先可对角化 \Leftrightarrow 存在 n 个线性无关的特征向量, 它们是构成 V 的一组基, 若所有 $m(\lambda) = g(\lambda)$, 则特征向量总数 $\sum g(\lambda) = \sum m(\lambda) = n$, 且不同特征值的特征向量线性无关, 故它们构成基。并且若存在 $m(\lambda) > g(\lambda)$, 则特征向量总数 $\sum g(\lambda) < n$, 无法构成基, 反映在 $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$ 中的矩阵 \mathbf{P} 不可逆, 矩阵不可对角化。

□

6.4.4 Jordan 标准型⁴

首先我们给出 Jordan 标准型的定义。

定义 6.11 (Jordan 标准型 (Jordan canonical form))

Jordan 标准型是一种特殊类型的分块矩阵, 其中每个分块由 Jordan 块组成, 且每个 Jordan 分块的常数 λ_i 可能不同, 我们将其记作 \mathcal{J} 。具体而言, 它是如下形式的分块矩阵:

$$\mathcal{J} := \begin{pmatrix} J_{m_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{m_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{m_i}(\lambda_i) \end{pmatrix}$$



例如下面一个矩阵就是 Jordan 标准型 (空白处均为 0)

$$\mathcal{J} = \left(\begin{array}{cc|c|c} 3 & 1 & & \\ 3 & 1 & & \\ 3 & 1 & & \\ 3 & & & \\ \hline & & 4 & 1 \\ & & 4 & 1 \\ & & 4 & \\ \hline & & & 5 \end{array} \right) = \begin{pmatrix} J_4(3) & 0 & 0 \\ 0 & J_3(4) & 0 \\ 0 & 0 & J_1(5) \end{pmatrix}$$

通过 Jordan 标准型我们可以清晰地看出其各个特征值, 以及它们的代数重数和几何重数, 考虑 Jordan 标准型

$$\mathcal{J} = \begin{pmatrix} J_{m_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{m_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{m_i}(\lambda_i) \end{pmatrix}$$

- \mathcal{J} 的特征值由 Jordan 块的对角线元素给出, 即 $\lambda_1, \lambda_2, \dots, \lambda_i$ (这些值可能重复); 特征值的总数 (考虑代数重数) 为 $n = \sum_{k=1}^i m_k$, 不同特征值的数量至多为 i (当所有 λ_k 互异时取等号)。
- 对于每个特征值 λ , 其代数重数为所有满足 $\lambda_k = \lambda$ 的 Jordan 块的大小之和, 即 $\sum_{\{k|\lambda_k=\lambda\}} m_k$ 。
- 对于每个特征值 λ , 其几何重数为所有满足 $\lambda_k = \lambda$ 的 Jordan 块的数量 (即索引 k 的个数)。

说的通俗一些就是对角元素的特征值 λ , 对角线上有几个相同的 λ 对应的特征值 λ 就是几重根, 即为代数重数; 对角元素是 λ 的 Jordan 块有几个, 对应的线性无关的特征向量就有几个, 即为几何重数; 例如 $\mathcal{J} = \begin{pmatrix} J_4(3) & 0 & 0 \\ 0 & J_3(4) & 0 \\ 0 & 0 & J_1(5) \end{pmatrix}$, 其特征值分别为 3, 4, 5 其中特征值 3 的几何重数为 1, 代数重数为 4; 特征值 4 的几

⁴本节选学

何重数为 1, 代数重数为 3; 特征值 5 的几何重数为 1, 代数重数为 1。根据定理6.3我们可以得出该标准型不能被对角化。

6.4.5 Jordan 标准型的幂与幂零算子

首先我们看 Jordan 标准型的幂, 由于相似变换不改变矩阵的幂 (即若 $A = P\mathcal{J}P^{-1}$, 则 $A^k = P\mathcal{J}^kP^{-1}$), 研究 Jordan 标准型 \mathcal{J} 的幂 \mathcal{J}^k 等价于研究原矩阵 A 的幂。而 J 是由 Jordan 块 $J_m(\lambda)$ 沿对角线排列的分块对角矩阵:

$$\mathcal{J} = \begin{pmatrix} J_{m_1}(\lambda_1) & & & \\ & J_{m_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{m_k}(\lambda_k) \end{pmatrix}$$

因此, \mathcal{J}^k 也是分块对角矩阵, 其每个对角块即为对应 Jordan 块的幂:

$$\mathcal{J}^k = \begin{pmatrix} J_{m_1}^k(\lambda_1) & & & \\ & J_{m_2}^k(\lambda_2) & & \\ & & \ddots & \\ & & & J_{m_k}^k(\lambda_k) \end{pmatrix}$$

问题归结为: 如何计算一个 Jordan 块 $J_m(\lambda)$ 的 k 次幂。

一个 m 阶 Jordan 块 $J_m(\lambda)$ 可以分解为两部分之和:

$$J_m(\lambda) = \lambda I_m + N_m$$

其中⁵

$$N_m = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

利用二项式定理 (因为 λI_m 与 N_m 是可交换 (见课后习题) 的: $(\lambda I_m)N_m = N_m(\lambda I_m)$), 我们可以计算 $J_m^k(\lambda)$:

$$(J_m(\lambda))^k = (\lambda I_m + N_m)^k = \sum_{r=0}^k C_k^r (\lambda I_m)^{k-r} (N_m)^r = \sum_{r=0}^k C_k^r \lambda^{k-r} (N_m)^r$$

其中 $C_k^r = \frac{k!}{r!(k-r)!}$ 是二项式系数。接下来我们先来研究 N_m 的特性, 在此之前我们先讲一下幂零算子以及它的矩阵表示。

首先我们明确一个概念, 那就是线性映射算子 $T \in \mathcal{L}(\mathbb{F}^n)$ 可以表示为矩阵 $\mathcal{M}(T)$, 令 $\mathbf{A} = \mathcal{M}(T)$ 我们可以得到向量 $x \in \mathbb{F}^n$ 可得 $T(x) = \mathbf{A}x$, 每一次变换相当于在等式中左乘一个 \mathbf{A} , 由此我们可以推断:

公理 6.1 (矩阵乘公理)

线性映射算子 $T \in \mathcal{L}(\mathbb{F}^n)$ 与向量 $x \in \mathbb{F}^n$ 存在正整数 $m \in \mathbb{N}^+$ 使得 T^m 表示 T 的 m 次复合, 则

$$T^m(x) = A^m x$$

接下来我们引入幂零算子的概念。

⁵实际上下面的 N_m 就是幂零矩阵

定义 6.12 (幂零算子 (Nilpotent operator))

线性映射算子 $T \in \mathcal{L}(\mathbb{F}^n)$ 存在 m 次复合使得 $\forall x \in \mathbb{F}^n, x \neq \mathbf{0}$ 转化为零向量即 $T^m(x) = \mathbf{0}$, 则 T 为幂零算子; 而其对应的矩阵 $\mathcal{M}(T)$ 为幂零矩阵。



根据上面的定义我们可以得出 N_m 是幂零矩阵, 具体来说 (请读者计算):

- $(N_m)^1$: 次对角线 (紧邻主对角线上方的那条对角线) 上的元素为 1, 其余位置为 0。
- $(N_m)^2$: 次次对角线 (主对角线向上数第二条) 上的元素为 1, 其余位置为 0;
- ...
- $(N_m)^{m-1}$: 仅右上角元素为 1, 其余位置为 0。
- $(N_m)^m = \mathbf{0}$: 零矩阵。

由于 N_m 是幂零指数为 m 的幂零算子, 当 $r \geq m$ 时, $(N_m)^r = \mathbf{0}$ 。因此, 上述求和实际上只需从 $r = 0$ 到 $r = \min(k, m - 1)$:

$$(J_m(\lambda))^k = \sum_{r=0}^{\min(k, m-1)} C_k^r \lambda^{k-r} (N_m)^r$$

例题 6.6.

3 阶 Jordan 块 $J_3(\lambda)$ 与其幂零矩阵 N_3 为

$$J_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad N_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

计算 $(J_3(\lambda))^k$:

解:

$$(J_3(\lambda))^k = \sum_{r=0}^{\min(k, 2)} C_k^r \lambda^{k-r} (N_3)^r$$

其中:

- $(N_3)^0 = I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
- $(N_3)^1 = N_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$
- $(N_3)^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
- $(N_3)^r = \mathbf{0}, \forall r \geq 3$.

因此:

$$\begin{aligned} (J_3(\lambda))^k &= C_k^0 \lambda^k (N_3)^0 + C_k^1 \lambda^{k-1} (N_3)^1 + C_k^2 \lambda^{k-2} (N_3)^2 \\ &= \lambda^k \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + k \lambda^{k-1} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \frac{k(k-1)}{2} \lambda^{k-2} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \frac{k(k-1)}{2}\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{pmatrix}$$

下面做一个总结：主对角线全为 λ^k ，第一条上对角线全为 $k\lambda^{k-1}$ ，第二条上对角线全为 $\frac{k(k-1)}{2}\lambda^{k-2}$ 。当 $k \geq 2$ 时，结构稳定；当 $k = 1$ 时，最后一项消失，退回到 $J_3(\lambda)$ ；当 $k = 0$ 时，为单位阵。

6.4.6 将矩阵相似化为 Jordan 标准型

虽然有些矩阵很特殊，不能将其化为对角矩阵，但是我们可以将其化为 Jordan 标准型。首先我们先不讲到底哪些矩阵可以化为 Jordan 标准型⁶，先讲如何把矩阵化为 Jordan 标准型。

考虑从如下的 Jordan 标准形

$$\mathcal{J} = \left(\begin{array}{ccc|c|c} 3 & 1 & & & \\ & 3 & 1 & & \\ & & 3 & 1 & \\ & & & 3 & \\ \hline & & & 4 & 1 \\ & & & & 4 & 1 \\ & & & & & 4 \\ \hline & & & & & 5 \end{array} \right)$$

出发，即存在矩阵 \mathbf{A} 满足 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathcal{J}$ ，通过等式两侧左乘一个矩阵 \mathbf{P} 得到

$$\mathbf{AP} = \mathbf{P}\mathcal{J}$$

设矩阵的列向量构成为 $(p_1 \ p_2 \ p_3 \ \cdots \ p_8)$ ，根据每列的值，我们依次得出：

$$\begin{aligned} \mathbf{Ap}_1 &= 3p_1 & (\mathbf{A} - 3\mathbf{I})p_1 &= \mathbf{0} \\ \mathbf{Ap}_2 &= p_1 + 3p_2 & (\mathbf{A} - 3\mathbf{I})p_2 &= p_1 \\ \mathbf{Ap}_3 &= p_2 + 3p_3 & (\mathbf{A} - 3\mathbf{I})p_3 &= p_2 \\ \frac{\mathbf{Ap}_4 = p_3 + 3p_4}{\mathbf{Ap}_5 = 4p_5} &\implies \frac{(\mathbf{A} - 3\mathbf{I})p_4 = p_3}{(\mathbf{A} - 4\mathbf{I})p_5 = \mathbf{0}} & & (6.1) \\ \mathbf{Ap}_6 &= p_5 + 4p_6 & (\mathbf{A} - 4\mathbf{I})p_6 &= p_5 \\ \frac{\mathbf{Ap}_7 = p_6 + 4p_7}{\mathbf{Ap}_8 = 5p_8} &\implies \frac{(\mathbf{A} - 4\mathbf{I})p_7 = p_6}{(\mathbf{A} - 5\mathbf{I})p_8 = \mathbf{0}} & & \end{aligned}$$

矩阵的列向量 p_1, p_5, p_8 都是矩阵 \mathbf{A} 的特征向量，对应的特征值分别为 3, 4, 5，对应的代数重数分别为 4, 3, 1，实际上这些关系揭示了 Jordan 标准型的核心结构：每个 Jordan 块对应一个由特征向量和广义特征向量组成的链（称为 Jordan 链）。具体来说就是如下定义：

定义 6.13 (广义特征向量 (Generalized Eigenvector))

$n \times n$ 矩阵 \mathbf{A} 的广义特征向量是向量 v 满足

$$(\mathbf{A} - \lambda\mathbf{I})^k v = \mathbf{0}$$

- 对于特征值 $\lambda = 3$ ，存在长度为 4 的 Jordan 链：

⁶事实上后面会提及所有方阵均可化为 Jordan 标准型

- p_1 是特征向量（满足 $(\mathbf{A} - 3\mathbf{I})p_1 = \mathbf{0}$ ）
- p_2 是一阶广义特征向量（满足 $(\mathbf{A} - 3\mathbf{I})p_2 = p_1$ ）
- p_3 是二阶广义特征向量（满足 $(\mathbf{A} - 3\mathbf{I})p_3 = p_2$ ）
- p_4 是三阶广义特征向量（满足 $(\mathbf{A} - 3\mathbf{I})p_4 = p_3$ ）
- 类似地， $\lambda = 4$ 对应长度为 3 的链： p_5 （特征向量） $\rightarrow p_6$ （一阶广义） $\rightarrow p_7$ （二阶广义）
- $\lambda = 5$ 对应长度为 1 的链：仅特征向量 p_8 。

给定矩阵 \mathbf{A} ，求其 Jordan 标准型的步骤如下：

1. 求解特征方程 $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ ，得到特征值 λ_i 及其代数重数 m_i 。
2. 对每个特征值 λ_i 计算广义特征空间，即
 - 定义广义特征空间： $W_k(\lambda_i) = \text{null}((\mathbf{A} - \lambda_i\mathbf{I})^k)$
 - 计算幂零指数 ρ_i ：满足 $W_{\rho_i}(\lambda_i) = W_{\rho_i+1}(\lambda_i)$ 的最小整数（即空间稳定化的最小幂次）
3. 构造 Jordan 链，从最高阶广义特征向量开始：选 $v_{\rho_i} \in W_{\rho_i}(\lambda_i) \setminus W_{\rho_i-1}(\lambda_i)$ ⁷
4. 逆向生成 Jordan 链

$$\begin{aligned} v_{\rho_i-1} &= (\mathbf{A} - \lambda_i\mathbf{I})v_{\rho_i} \\ v_{\rho_i-2} &= (\mathbf{A} - \lambda_i\mathbf{I})v_{\rho_i-1} \\ &\vdots \\ v_1 &= (\mathbf{A} - \lambda_i\mathbf{I})v_2 \end{aligned}$$

其中 v_1 是特征向量，整个链 $\{v_1, v_2, \dots, v_{\rho_i}\}$ 线性无关。

5. 若广义特征空间未满（即 $\dim W_{\rho_i}(\lambda_i) > \rho_i$ ），重复选取新向量生成新链。
6. 组合变换矩阵，将所有 Jordan 链的向量按顺序排列为 \mathbf{P} 的列向量。
7. 得到 Jordan 标准型 $\mathcal{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 。

我们来详细看一个例子：

例题 6.7.

矩阵 $\mathbf{A} = \begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix}$ 存在矩阵 \mathbf{P} 与 Jordan 标准形 \mathcal{J} 使得 $\mathcal{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 成立，求 \mathbf{P} 的一个可能值。

解：

为了保证有序性，使用红色文本代表选取此 Jordan 链作为 \mathbf{P} 的列向量，且保证列向量线性无关。蓝色文本表示当前条件下可选择另一个基础解系生成的 Jordan 链，两者取其一即可。

首先根据特征向量 v 对应的特征值 λ 的定义，我们有

$$(\mathbf{A} - \lambda\mathbf{I})v = \mathbf{0}$$

仅有行列式 $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ 特征多项式

$$P(\lambda) = \begin{vmatrix} -12 - \lambda & 1 & -7 & 8 \\ -27 & 3 - \lambda & -11 & 14 \\ -41 & 2 & -16 - \lambda & 22 \\ -56 & 3 & -26 & 33 - \lambda \end{vmatrix} = \lambda^4 - 8\lambda^3 + 23\lambda^2 - 28\lambda + 12$$

⁷ 符号 $A \setminus B$ 表示减集合，即 $\{x \mid x \in A \text{ and } x \notin B\}$ ，事实上线性代数还有一种空间集合运算，称为商空间，用的是 A/B 表示。

化简为

$$P(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 2)(\lambda - 3) = 0$$

我们得到三个特征值为 $\lambda_1 = 1, \lambda_{2,3} = 2, \lambda_4 = 3$, 接下来为每个特征值求线性独立的广义特征向量, 因为 $\mathbf{A} - \lambda \mathbf{I}$ 是幂零矩阵我们需要先确定其最大的秩。

首先考虑 $\lambda = 1$ 时其代数重数为 1;

1. 确定广义特征向量的最大秩

$$r \left(\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 1\mathbf{I} \right)^1 \right) = 3$$

2. 求其 Jordan 链, 即求每一个矩阵幂下的关于向量 x 的基础解系; 当幂为 1 时, 其秩为 3, 关于 x 矩阵乘

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 1\mathbf{I} \right)^1 x = \mathbf{0}$$

的一个基础解系 \mathcal{B} 为 $\mathcal{B} = \{(1, 2, 3, 4)\}$ ^a, 带入此基础解系 $\mathcal{B} = \{(1, 2, 3, 4)\}$

(a). 令 $x_1 = (1, 2, 3, 4)$ 可得

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 1\mathbf{I} \right)^{(1-1)} x_1 = \mathbf{I}x_1 \neq \mathbf{0}$$

所以该向量仅为特征向量, 为此广义特征向量生成 Jordan 链为 $v_1 = (1, 2, 3, 4)$

接下来考虑 $\lambda = 2$ 时其代数重数为 2;

1. 确定广义特征向量的最大秩

$$r \left(\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^1 \right) = 3, r \left(\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^2 \right) = 2$$

2. 求其 Jordan 链, 当幂为 1 时其秩为 3, 当幂为 2 时其秩为 2, 从最高次幂构建 Jordan 链, 关于 x 的矩阵乘法

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^2 x = \mathbf{0}$$

的一个基础解系 \mathcal{B} 为 $\mathcal{B} = \{(-1, -1, 1, 0), (1, 4, 0, 3)\}$

(a). 令 $x_1 = (-1, -1, 1, 0)$,

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^{(2-1)} \quad x_1 \neq \mathbf{0}$$

所以该向量仅为一阶的广义特征向量，为此广义特征向量生成 Jordan 链为

$$v_1 = (-1, -1, 1, 0) \leftarrow v_2 = \left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^1 \quad v_1 = (6, 15, 21, 27)$$

(b). 令 $x_2 = (1, 4, 0, 3)$,

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^{(2-1)} \quad x_2 \neq \mathbf{0}$$

所以该向量仅为一阶的广义特征向量，为此广义特征向量生成 Jordan 链为

$$v_1 = (1, 4, 0, 3) \leftarrow v_2 = \left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 2\mathbf{I} \right)^1 \quad v_1 = (-14, -35, -49, -63)$$

接下来考虑 $\lambda = 3$ 时其代数重数为 1;

1. 确定广义特征向量的最大秩

$$r \left(\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 3\mathbf{I} \right)^1 \right) = 3$$

当幂为 1 时，其秩为 3，关于 x 矩阵乘

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 3\mathbf{I} \right)^1 x = \mathbf{0}$$

的一个基础解系 B 为 $B = \{(2, 2, 4, 7)\}$

(a). 令 $x_1 = (2, 2, 4, 7)$ 可得

$$\left(\begin{pmatrix} -12 & 1 & -7 & 8 \\ -27 & 3 & -11 & 14 \\ -41 & 2 & -16 & 22 \\ -56 & 3 & -26 & 33 \end{pmatrix} - 3\mathbf{I} \right)^{(1-1)} \quad x_1 = \mathbf{I}x_1 \neq \mathbf{0}$$

所以该向量仅为特征向量，为此广义特征向量生成 Jordan 链为 $v_1 = (2, 2, 4, 7)$

最后将选取列的几个 Jordan 链为

$$(1, 2, 3, 4), (6, 15, 21, 27) \rightarrow (-1, -1, 1, 0), (2, 2, 4, 7)$$

矩阵列是来自顺序排列所选链的广义特征向量为

$$\mathbf{P} = \begin{pmatrix} 1 & 6 & -1 & 2 \\ 2 & 15 & -1 & 2 \\ 3 & 21 & 1 & 4 \\ 4 & 27 & 0 & 7 \end{pmatrix}$$

^a如果读者不知道这一部分是在做什么，请复习章节5.5

讲完了如何化为 Jordan 标准形，最后我们来看看一个推论：

推论 6.9

任何 $\mathbb{C}^{n \times n}$ 矩阵 \mathbf{A} 均存在矩阵 \mathbf{P} 和 Jordan 标准形 \mathcal{J} 使得

$$\mathcal{J} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

成立



证明 我们这里不做严格证明，首先我们选取广义特征向量的时候都会选取一个 Jordan 链，不同于对角矩阵，Jordan 标准形扩展了特征向量的意义，使得存在一个广义特征向量使得其能够化为一个个 Jordan 块，保证 \mathbf{P} 线性无关，由此任何方阵均存在一个 \mathcal{J} 与其相似。

6.5 章节练习

6.5.1 A 组

练习 49 若算子 $T \in \mathcal{L}(\mathbb{C}^3)$ 定义为 $T((x, y, z)) = (3x + z, y - x, x + y + z)$ ，求 T 的特征值以及对应的特征向量。

练习 50 已知 12 是矩阵

$$\mathbf{A} = \begin{pmatrix} 7 & 4 & -1 \\ 4 & 7 & -1 \\ -4 & a & 4 \end{pmatrix}$$

的特征值，求 a 与 $\text{Tr}\mathbf{A}$ 。

练习 51

设矩阵 $\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 1 \end{pmatrix}^{100}$ ，求 \mathbf{A}^{-1} 和 \mathbf{A}^{100} 。

练习 52

已知矩阵 $\mathbf{A} = \begin{pmatrix} -2 & 0 & 0 \\ 2 & x & 2 \\ 3 & 1 & 1 \end{pmatrix}$ 相似于对角矩阵 $\mathbf{B} = \text{diag}(-1, 2, y)$ ，求 x, y 的值。

