

# AI 中的数学

邓小铁 李翰禹

2024 年 9 月 28 日

# 目录

第零章 引言	1
第一部分 AI 的逻辑	2
第一章 合情推理	3
§1.1 命题逻辑的演绎推理	4
§1.2 合情推理的数学模型	12
§1.2.1 合情推理的基本假设，似然	14
§1.2.2 似然与概率	19
§1.2.3 先验与基率谬误	21
§1.3 合情推理的归纳强论证	23
§1.3.1 归纳强论证	23
§1.3.2 有效论证和归纳强论证的比较	28
§1.4 先验模型的存在性	33
§1.5 章末注记	35
§1.6 习题	36

<b>第二章 Markov 链与模型</b>	<b>37</b>
§2.1 Markov 链	38
§2.2 Markov 奖励过程 (MRP)	49
§2.3 Markov 决策过程 (MDP)	55
§2.4 隐 Markov 模型 (HMM)	64
§2.4.1 评估问题	67
§2.4.2 解释问题	69
§2.5 扩散模型	72
§2.5.1 采样逆向过程	77
§2.5.2 训练逆向过程	78
§2.6 章末注记	81
§2.7 习题	81
 <b>第二部分 信息与数据</b>	 <b>82</b>
<b>第三章 熵与 Kullback-Leibler 散度</b>	<b>83</b>
§3.1 熵	84
§3.1.1 概念的导出	84
§3.1.2 概念与性质	89
§3.2 Kullback-Leibler 散度	98
§3.2.1 定义	98
§3.2.2 两个关于信息的不等式	101
§3.3 编码理论	102
§3.3.1 熵与编码	102
§3.3.2 K-L 散度、交叉熵与编码	106

§3.4 在机器学习中的应用：语言生成模型 . . . . .	108
§3.5 附录：Shannon 定理的证明 . . . . .	110
§3.6 习题 . . . . .	113
§3.7 章末注记 . . . . .	115
<b>第四章 高维几何，Johnson-Lindenstrauss 引理</b>	<b>117</b>
§4.1 高维几何 . . . . .	119
§4.1.1 高维球体 . . . . .	119
§4.1.2 Stein 悖论 . . . . .	123
§4.1.3 为什么我们要正则化? 远有潜龙，勿用 . . . . .	130
§4.2 集中不等式 . . . . .	131
§4.3 J-L 引理的陈述与证明 . . . . .	138
§4.4 J-L 引理的应用 . . . . .	143
§4.5 习题 . . . . .	146
§4.6 章末注记 . . . . .	147
<b>第五章 差分隐私</b>	<b>148</b>
§5.1 数据隐私问题 . . . . .	149
§5.2 差分隐私的定义与性质 . . . . .	153
§5.3 差分隐私的应用 . . . . .	161
§5.3.1 随机反应算法 . . . . .	161
§5.3.2 全局灵敏度与 Laplace 机制 . . . . .	164
§5.3.3 DP 版本 Llyod 算法 . . . . .	167
§5.4 习题 . . . . .	170
§5.5 章末注记 . . . . .	170

### 第三部分 决策与优化 171

#### 第六章 凸分析 172

§6.1 决策与优化的基本原理 . . . . .	173
§6.1.1 统计决策理论 . . . . .	173
§6.1.2 优化问题 . . . . .	176
§6.1.3 例子: 网格搜索算法 . . . . .	183
§6.2 凸函数 . . . . .	188
§6.3 凸集 . . . . .	194
§6.3.1 基本定义和性质 . . . . .	195
§6.3.2 分离超平面定理 . . . . .	199
§6.4 习题 . . . . .	202
§6.5 章末注记 . . . . .	202

#### 第七章 对偶理论 202

§7.1 约束的几何意义 . . . . .	205
§7.2 条件极值与 Lagrange 乘子法 . . . . .	213
§7.3 Karush–Kuhn–Tucker 条件 . . . . .	217
§7.4 Lagrange 对偶 . . . . .	222
§7.4.1 原始规划与对偶规划 . . . . .	222
§7.4.2 对偶的几何意义 . . . . .	227
§7.4.3 弱对偶定理 . . . . .	229
§7.4.4 Slater 条件, 强对偶定理 . . . . .	230
§7.5 应用: 支持向量机 (SVM) . . . . .	235
§7.6 习题 . . . . .	238
§7.7 章末注记 . . . . .	238

<b>第八章 不动点理论</b>	<b>239</b>
§8.1 Banach 不动点定理 . . . . .	240
§8.2 Brouwer 不动点定理 . . . . .	251
§8.3 习题 . . . . .	258
§8.4 章末注记 . . . . .	258

## **第四部分 博弈与逻辑 259**

<b>第九章 逻辑与博弈</b>	<b>260</b>
§9.1 博弈的基本语言：以井字棋为例 . . . . .	262
§9.2 输赢博弈 . . . . .	264
§9.2.1 博弈的不同维度 . . . . .	264
§9.2.2 Zermelo 定理与 AlphaGo Zero . . . . .	266
§9.3 正则形式博弈 . . . . .	273
§9.3.1 定义 . . . . .	274
§9.3.2 理性与均衡 . . . . .	276
§9.3.3 生成对抗网络 . . . . .	279
§9.3.4 混合策略 . . . . .	282
§9.4 随机博弈 (Markov 博弈) . . . . .	288
§9.5 习题 . . . . .	299
§9.6 章末注记 . . . . .	299

## **第五部分 认知与逻辑 300**

<b>第十章 共同知识, Bayes 博弈, Aumann 知识算子</b>	<b>301</b>
§10.1 “泥泞的孩童”谜题 . . . . .	304

§10.2 不完全信息博弈 (Bayes 博弈)	309
§10.3 电子邮件博弈	319
§10.4 Aumann 知识算子	324
§10.5 习题	333
§10.6 章末注记	333
<b>第十一章 模态逻辑, 知识的逻辑</b>	<b>334</b>
§11.1 知识逻辑的形式语言	336
§11.2 Kripke 语义	341
§11.3 模态可定义性	348
§11.4 知识逻辑的基本模型与性质	351
§11.4.1 知识逻辑的 Kripke 模型与公理	351
§11.4.2 Kripke 模型与 Aumann 结构	358
§11.4.3 “泥泞的孩童”再回顾: 形式化解法	360
§11.5 对不一致达成一致	362
§11.5.1 模型	363
§11.5.2 定理及其证明	366
§11.6 习题	368
§11.7 章末注记	368
<b>第六部分 附录: 预备知识</b>	<b>369</b>
<b>附录 A 线性代数基础</b>	<b>370</b>
§A.1 线性空间	370
§A.2 线性映射	377
§A.3 矩阵	383

§A.4 双线性型与二次型 . . . . .	392
§A.5 带内积的线性空间 . . . . .	398
§A.6 行列式 . . . . .	407
§A.7 算子范数与谱理论 . . . . .	411
<b>附录 B 微分学基础</b>	<b>420</b>
§B.1 点集拓扑 . . . . .	420
§B.1.1 度量空间, 范数 . . . . .	420
§B.1.2 开集与闭集 . . . . .	425
§B.1.3 紧致性, 收敛性, 完备性 . . . . .	429
§B.1.4 连续映射 . . . . .	433
§B.1.5 与实数序有关的性质 . . . . .	438
§B.2 一元函数的微分学 . . . . .	441
§B.2.1 导数与微分的定义 . . . . .	442
§B.2.2 微分学基本定理 . . . . .	447
§B.3 多元函数的微分学 . . . . .	449
§B.3.1 微分、偏导数与导数的定义 . . . . .	449
§B.3.2 微分学基本定理 . . . . .	459
§B.3.3 隐函数定理 . . . . .	462
<b>附录 C 概率论基础</b>	<b>468</b>
§C.1 从朴素概率论到公理化概率论 . . . . .	468
§C.1.1 Kolmogorov 概率论 . . . . .	468
§C.1.2 条件概率, 独立性 . . . . .	474
§C.2 随机变量, 分布函数 . . . . .	480
§C.2.1 基本定义 . . . . .	480



§C.2.2 离散型随机变量 . . . . .	486
§C.2.3 连续型随机变量 . . . . .	487
§C.2.4 随机向量, 条件分布, 独立性 . . . . .	492
§C.2.5 随机变量 (向量) 的函数 . . . . .	499
§C.3 随机变量的数字特征, 条件数学期望 . . . . .	503
§C.3.1 数学期望, Lebesgue 积分 . . . . .	503
§C.3.2 数学期望的性质 . . . . .	510
§C.3.3 随机变量的内积空间 . . . . .	514
§C.3.4 特征函数 . . . . .	517
§C.3.5 条件数学期望 . . . . .	519
§C.4 多元正态分布 (Gauss 向量) . . . . .	526
§C.5 大数定律 . . . . .	528

# 第一部分

## AI 的逻辑

## 第二部分

### 信息与数据

## 第三部分

### 决策与优化

## 第四部分

### 博弈与逻辑

## 第五部分

### 认知与逻辑

## 第六部分

### 附录：预备知识

# 附录 A 线性代数基础

## §A.1 线性空间

从动机上说，线性空间试图将  $\mathbb{R}^n$  或者  $\mathbb{C}^n$  这样的集合连同他们上面的代数结构抽象出来. 除此之外，函数和无穷数列的集合也是非常重要的对象，比如说  $\mathbb{R}$  上的连续函数组成的集合  $C(\mathbb{R})$ ，或者具有“模长”的无穷复数列 ( $\ell^2$  空间)：

$$\ell^2 = \left\{ (x_1, x_2, \dots) \in \mathbb{C}^\infty : \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.$$

我们将这些对象的共性抽象出来，得到线性空间的概念. 线性空间都是基于某个域定义的，我们先给出域的定义.

**定义 A.1 (域)** 一个域是一个集合  $F$ ，其上定义了两种二元运算：加法  $+$  和乘法  $\cdot$ ，他们都是  $F \times F$  到  $F$  的映射，满足下面的公理：

1. (结合律) 对于任意的  $a, b, c \in F$ ，有  $(a + b) + c = a + (b + c)$  和  $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ ；
2. (交换律) 对于任意的  $a, b \in F$ ，有  $a + b = b + a$  和  $a \cdot b = b \cdot a$ ；



3. (分配律) 对于任意的  $a, b, c \in F$ , 有  $a \cdot (b + c) = a \cdot b + a \cdot c$ .
4. (单位元) 存在唯一的两个元素  $0, 1 \in F$ , 使得对于任意的  $a \in F$ , 有  $a + 0 = a$  和  $a \cdot 1 = a$ ;
5. (加法逆元) 对于任意的  $a \in F$ , 存在唯一  $b \in F$ , 使得  $a + b = 0$ , 记  $b$  作  $-a$ ;
6. (乘法逆元) 对于任意的  $a \in F$ , 如果  $a \neq 0$ , 则存在唯一  $b \in F$ , 使得  $a \cdot b = 1$ , 记  $b$  作  $a^{-1}$ .

通常将  $a \cdot b$  写作  $ab$ , 并且乘法的优先级高于加法, 即  $ab + c = (ab) + c$ .  $\square$

域的重要例子包括有理数域  $\mathbb{Q}$ , 实数域  $\mathbb{R}$  和复数域  $\mathbb{C}$ , 他们都是无限域. 我们将在后面的内容中使用这些域. 接下来, 我们定义线性空间.

**定义 A.2 (线性空间, 向量空间)** 设  $V$  是一个集合,  $F$  是一个域. 如果在  $V$  上定义了两种运算: 加法  $+$  和数乘  $\cdot$ , 使得  $V$  满足下面的公理:

1. ( $V$  的结合律) 对于任意的  $x, y, z \in V$ , 有  $(x + y) + z = x + (y + z)$ ;
2. ( $V$  的交换律) 对于任意的  $x, y \in V$ , 有  $x + y = y + x$ ;
3. (加法零元) 存在唯一的元素  $0 \in V$ , 使得对于任意的  $x \in V$ , 有  $x + 0 = x$ ;
4. (加法逆元) 对于任意的  $x \in V$ , 存在唯一  $y \in V$ , 使得  $x + y = 0$ , 记  $y$  作  $-x$ ;
5. 对于任意的  $x \in V$ , 有  $1 \cdot x = x$ ;

6. 对于任意的  $a, b \in F$  和  $x \in V$ , 有  $(ab) \cdot x = a \cdot (b \cdot x)$ ;
7. 对于任意的  $a \in F$  和  $x, y \in V$ , 有  $a \cdot (x + y) = a \cdot x + a \cdot y$ ;
8. 对于任意的  $a, b \in F$  和  $x \in V$ , 有  $(a + b) \cdot x = a \cdot x + b \cdot x$ .

则称  $V$  是一个  $F$ -线性空间, 简称线性空间, 也称向量空间.  $V$  中的元素被称为向量. 通常将数乘  $a \cdot x$  写作  $ax$ , 并且乘法的优先级高于加法, 即  $a \cdot x + y = (a \cdot x) + y$ .  $\square$

“线性”一词的含义是指的  $ax + by$  这种形式的数学对象, 线性代数就是研究这种对象的学科. 线性空间的典型例子包括:

- $\mathbb{R}^n$  和  $\mathbb{C}^n$ .
- $M_{m \times n}(F)$ , 即所有  $m \times n$  矩阵组成的集合.
- $C(\mathbb{R})$ , 即  $\mathbb{R}$  上的连续函数组成的集合.
- $C^k(\mathbb{R})$ , 即  $\mathbb{R}$  上的  $k$  次连续可微函数组成的集合.
- $\ell^2$  空间, 即所有二次可和的复数序列组成的集合.

如同所有其他的代数结构, 线性空间也有各式各样构造新的线性空间的方法. 为了看出来线性空间本质的特性, 我们有如下引理:

**引理 A.1** 设  $V$  是  $F$ -线性空间,  $W$  是  $V$  的一个子集. 则  $W$  是一个线性空间当且仅当对任意  $a, b \in F$  和  $x, y \in W$ , 有  $ax + by \in W$ .

**证明.** 按照定义即可验证.  $\square$

我们给  $ax + by$  这样的对象一个正式的定义.

**定义 A.3 (线性组合)** 设  $V$  是  $F$ -线性空间,  $x_1, \dots, x_n \in V$ ,  $a_1, \dots, a_n \in F$ , 则称  $a_1x_1 + \dots + a_nx_n$  是  $x_1, \dots, x_n$  的一个线性组合.  $\square$

接下来, 基于某些特定的线性空间, 我们构造各种新的线性空间.

**定义 A.4 (线性子空间)** 设  $V$  是  $F$ -线性空间,  $W$  是  $V$  的一个子集. 如果  $W$  是一个线性空间, 则称  $W$  是  $V$  的一个线性子空间.  $\square$

例如,  $\mathbb{Q}$  是  $\mathbb{R}$  的一个线性子空间, 但  $\mathbb{Z}$  不是  $\mathbb{R}$  的一个线性子空间. 再比如, 当  $k < l$ ,  $C^k(\mathbb{R})$  是  $C^l(\mathbb{R})$  的一个线性子空间.

**定义 A.5 (乘积空间)** 设  $V_1, \dots, V_n$  是  $F$ -线性空间, 则  $V_1 \times \dots \times V_n$  是一个  $F$ -线性空间, 其中加法和数乘分别定义为

$$\begin{aligned}(x_1, \dots, x_n) + (y_1, \dots, y_n) &= (x_1 + y_1, \dots, x_n + y_n), \\ a(x_1, \dots, x_n) &= (ax_1, \dots, ax_n).\end{aligned}\quad \square$$

例如,  $\mathbb{R}^n$  就是  $n$  个  $\mathbb{R}$  的乘积空间,  $M_{m \times n}(F)$  就是  $m \times n$  个  $F$  的乘积空间.

接下来, 我们按照表示论的观点, 引入基的概念. 线性空间是抽象的数学概念, 因此我们需要一些具体的元素去表示这整个空间.

**定义 A.6 (生成集)** 设  $V$  是  $F$ -线性空间,  $S \subseteq V$ , 如果  $V$  中的每一个元素都是  $S$  的线性组合, 则称  $S$  是  $V$  的一个生成集.

更一般地, 任意一个  $S \subseteq V$ , 我们可以定义  $S$  生成的线性子空间为所有  $S$  的线性组合的集合, 记为  $\text{Span}(S)$ .  $\square$

我们希望用尽可能少的元素来表示整个线性空间, 为此, 我们需要把“可表示”这样的概念严格化.

**定义 A.7 (线性相关)** 设  $V$  是  $F$ -线性空间,  $S \subseteq V$ , 如果存在  $x_1, \dots, x_n \in S$ ,  $a_1, \dots, a_n \in F$ , 使得  $a_1x_1 + \dots + a_nx_n = 0$ , 且至少有一个  $a_i \neq 0$ , 则称  $S$  是线性相关的, 否则称  $S$  是线性无关的.  $\square$

$S$  线性相关意味着  $S$  中的一些元素可以被另一些元素的线性组合表示出来, 因而  $S$  中有一些冗余. 线性无关意味着  $S$  中的元素都是必要的, 没有冗余. 由此, 我们可以给出基的定义.

**定义 A.8 (基)** 设  $V$  是  $F$ -线性空间,  $S \subseteq V$ , 如果  $S$  是线性无关的, 并且  $\text{Span}(S) = V$ , 则称  $S$  是  $V$  的一个基.  $\square$

线性空间的一个核心定理是基的存在性定理.

**定理 A.1 (基的存在性定理)** 设  $V$  是  $F$ -线性空间, 则  $V$  中存在一个基.

要注意, 这一定理不是平凡的. 首先, 基是线性无关的集合, 所以  $V$  本身通常就不是基. 此外, 这一定理要求有一个线性无关的集合  $S \subseteq V$ , 任意向量  $x \in V$  都可以用  $S$  中有限个元素的线性组合来表示, 这样的  $S$  并不容易找到. 该定理的证明是构造性的, 这一构造依赖于选择公理 (或者 Zorn 引理), 我们在此略去.

基的典型例子包括:

- $\mathbb{R}^n$  的标准基是  $\{e_1, \dots, e_n\}$ , 其中  $e_i$  是第  $i$  个分量为 1, 其余分量为 0 的向量.
- 特别地,  $\mathbb{R}$  的基就是  $\{1\}$ , 一般地, 域  $F$  作为线性空间的时候, 其基就是  $\{1\}$ . 但如果我们把  $\mathbb{C}$  看成  $\mathbb{R}$  的线性空间, 那么  $\mathbb{C}$  的基就是  $\{1, i\}$ .
- $M_{m \times n}(F)$  的标准基是  $\{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ , 其中  $E_{ij}$  是第  $i$  行第  $j$  列为 1, 其余元素为 0 的矩阵.

特别注意, 无穷维空间经常违背直觉. 例如, 考虑  $\ell^2$  空间和向量组  $\{e_1, e_2, \dots\}$ , 其中  $e_i$  是第  $i$  个分量为 1, 其余分量为 0 的实数列. 这个向量组看上去非常像一个基, 然而并非如此! 比如说,  $(1/n)_{n=1}^\infty \in \ell^2$ , 但是它不能写成有限个  $e_i$  的线性组合. 实际上,  $\ell^2$  空间的基一定是不可数的.

给定一个基, 我们可以用基来表示线性空间中的元素, 容易证明, 这一表示是唯一的. 因此, 我们可以把线性空间中的元素看成基的线性组合, 因而有了下面的定义.

**定义 A.9 (坐标)** 设  $V$  是  $F$ -线性空间,  $S$  是  $V$  的一个基,  $x \in V$ , 如果  $x = \sum_{v \in S} a_v v$ , 则称  $(a_v)_{v \in S}$  是  $x$  在基  $S$  下的坐标.  $\square$

例如,  $\mathbb{R}^3$  的标准基是  $\{e_1, e_2, e_3\}$ . 任意  $x \in \mathbb{R}^3$  都可以表示为  $x = a_1 e_1 + a_2 e_2 + a_3 e_3$ , 其中  $a_i$  是  $x$  的第  $i$  个分量. 因此, 我们可以把  $x$  看成一个三元组  $(a_1, a_2, a_3)$ , 这就是  $x$  在标准基下的坐标. 这样的讨论也适用于  $\mathbb{R}^n$  或  $\mathbb{C}^n$ . 另外, 坐标本身的集合也可以被看作是一个线性空间, 例如,  $\mathbb{R}^3$  的坐标集合就是  $\mathbb{R}^3$  本身.

线性空间的基可以衡量线性空间的复杂程度, 基元素越少, 线性空间越简单. 我们可以定义维数来衡量线性空间的复杂程度.

**定义 A.10 (维数)** 设  $V$  是  $F$ -线性空间, 如果  $V$  的一个基有限, 则称  $V$  是有限维的, 否则称  $V$  是无限维的. 有限维线性空间的基的元素个数称为  $V$  的维数, 记为  $\dim V$ .  $\square$

这一定义隐含的事实是, 如果  $V$  有有限基, 那么所有基都是有限的, 并且任意两个基的元素个数相同. 我们这里略去证明.

例如,  $\mathbb{R}^n$  的维数是  $n$ ,  $M_{m \times n}(F)$  的维数是  $mn$ ,  $C^k(\mathbb{R})$  和  $\ell^2$  都是无穷维的.

线性空间可以按照维数递降进行分解, 变成越来越简单的线性空间的组合. 这种组合称为直和.

**定义 A.11 (和空间与直和)** 设  $V$  是  $F$ -线性空间,  $U_1, U_2 \subseteq V$  是  $V$  的子空间, 定义他们的和空间为

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}.$$

如果  $U_1 \cap U_2 = \{0\}$ , 换句话说,  $U_1$  与  $U_2$  线性无关, 则称  $U_1 + U_2$  是直和, 记为  $U_1 \oplus U_2$ . 如果  $V = U_1 \oplus U_2$ , 则称  $U_1$  和  $U_2$  是  $V$  的直和分解.  $\square$

例如,  $\mathbb{R}^3$  可以分解为  $\mathbb{R}^3 = \mathbb{R}e_1 \oplus \mathbb{R}e_2 \oplus \mathbb{R}e_3$ , 其中  $\mathbb{R}e_i = \{\alpha e_i : \alpha \in \mathbb{R}\}$  是  $\mathbb{R}^3$  的一维子空间, 它们的直和就是  $\mathbb{R}^3$ . 注意到这个分解将三维线性空间分解成了三个一维线性空间, 这不是偶然的, 一般地, 我们有下面的定理.

**定理 A.2 (维数定理)** 设  $V$  是有限维  $F$ -线性空间,  $V = U_1 \oplus U_2$ , 则

$$\dim V = \dim U_1 + \dim U_2.$$

**证明.** 设  $S_1$  是  $U_1$  的一个基,  $S_2$  是  $U_2$  的一个基, 那么根据直和的定义,  $S_1 \cup S_2$  是  $V$  的一个基. 因为  $S_1 \cup S_2$  是线性无关的, 所以必然有  $S_1 \cap S_2 = \emptyset$ . 又由于  $V$  中的任意元素都可以写成  $S_1 \cup S_2$  中元素的线性组合, 因此

$$\dim V = |S_1 \cup S_2| = |S_1| + |S_2| = \dim U_1 + \dim U_2. \quad \square$$

通过直和分解, 我们可以把线性空间分成越来越简单的部分.

## §A.2 线性映射

接下来我们研究线性空间之间的关系. 并不是所有的关系都是重要的, 我们所关心的是保持线性空间代数结构的这种关系, 这种关系称为线性映射.

**定义 A.12 (线性映射, 线性算子, 线性函数)** 设  $V$  和  $W$  是  $F$ -线性空间, 如果映射  $f: V \rightarrow W$  满足:

1. 对任意  $x, y \in V$ ,  $f(x + y) = f(x) + f(y)$ ;
2. 对任意  $x \in V$  和  $a \in F$ ,  $f(ax) = af(x)$ ,

则称  $f$  是  $V$  到  $W$  的一个线性映射. 如果  $V = W$ , 则称  $f$  是  $V$  上的一个线性算子或线性变换. 如果  $W = F$ , 则称  $f$  是  $V$  上的一个线性函数.  $\square$

一个更简洁但也更本质的定义是, 线性映射是保持线性组合的映射.

**例 A.1** 一个平凡的例子是零映射:  $f: V \rightarrow W$ ,  $f(x) = 0$ , 这显然是线性映射, 我们通常记为  $O$ . 另一个平凡的例子是恒等映射:  $f: V \rightarrow V$ ,  $f(x) = x$ , 这也是线性映射, 我们通常记为  $\text{id}$ .  $\square$

线性映射有如下基本性质:

**命题 A.1** 设  $f: V \rightarrow W$  是域  $F$  上的线性映射, 那么

1.  $f(0) = 0$ ;
2.  $f(-x) = -f(x)$ ;
3.  $f(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i f(x_i)$ ;

4. 如果  $g: W \rightarrow Z$  是线性映射, 则  $g \circ f: V \rightarrow Z$  也是线性映射;
5. 如果  $g: V \rightarrow W$  是线性映射,  $a, b \in F$ , 则  $af + bg: x \mapsto af(x) + bg(x)$  也是线性映射; 也是线性映射;
6. 如果  $g: V \rightarrow W$  是线性映射,  $h: W \rightarrow Z$  是线性映射,  $k: Z \rightarrow V$  是线性映射, 则  $h \circ (f + g) = h \circ f + h \circ g$ ,  $(f + g) \circ k = f \circ k + g \circ k$ ;
7. 如果  $f$  是双射, 则  $f^{-1}$  也是线性映射.

**证明.** 按照定义验证即可. □

为了简化记号, 我们会将线性映射  $f$  的作用  $f(x)$  简记为  $fx$ , 线性映射的复合  $g \circ f$  简记为  $gf$ , 同一线性映射  $f$  的  $n$  次复合简记为  $f^n$ . 对于多项式函数  $G(x) = a_0 + a_1x + \cdots + a_nx^n$ , 我们可以定义一个新的线性映射  $G(f) = a_0\text{id} + a_1f + \cdots + a_nf^n$ .

线性映射可以被看成一种滤镜, 它可以将原始的空间进行变形, 变成一个新的空间. 比如说, 海上的月亮, 就是将三维空间的太阳与空间映到了海面上. 而线性算子则是一种特殊的线性映射, 它将原始空间变形成自身. 如果我们把线性空间看成一块橡皮泥, 那么线性算子可以被看成某种拉伸, 橡皮泥这个整体没有变多或者变少, 但是橡皮泥的形状发生了改变.

下面我们考虑两个线性映射的例子.

**例 A.2 (微分算子)** 考虑  $C^\infty(\mathbb{R})$ , 即任意次可微的实函数空间. 求导  $d/dx: C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$  被称为微分算子. 容易验证,  $d/dx$  是线性算子. □



**例 A.3 (投影变换)** 这个例子实际上是海上升明月的一般化. 考虑  $\mathbb{R}^n$ , 设  $m \leq n$ . 映射

$$\pi_m : (x_1, \dots, x_n) \mapsto (x_1, \dots, x_m, 0, \dots, 0)$$

称为  $\mathbb{R}^n$  的投影变换. 容易验证,  $\pi_m$  是线性算子. 此外, 实际上,  $\pi_m$  也可以被看作是  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射, 将  $(x_1, \dots, x_m, 0, \dots, 0)$  后面的 0 都丢掉, 这就是一个  $\mathbb{R}^m$  的元素.  $\square$

从投影变换的例子中, 我们可以体会到线性空间的微妙之处: 不同的线性空间可能有着完全相同的本质. 由  $(x_1, \dots, x_m, 0, \dots, 0)$  形成的空间实际上就是  $\mathbb{R}^m$ , 只是我们用了  $\mathbb{R}^n$  的元素来表示它. 这里引申出来了代数中两个重要的概念: 同态与同构.

**定义 A.13 (同态与同构)** 设  $V$  和  $W$  是  $F$ -线性空间, 如果映射  $f: V \rightarrow W$  满足:

1. 对任意  $x, y \in V$ ,  $f(x + y) = f(x) + f(y)$ ;
2. 对任意  $x \in V$  和  $a \in F$ ,  $f(ax) = af(x)$ ,

则称  $f$  是  $V$  到  $W$  的一个同态. 如果  $f$  是一个满射, 那么称  $f$  是一个满同态; 如果  $f$  还是一个单射, 那么称  $f$  是一个同构.  $\square$

线性空间之间的同态实际上就是线性空间之间的线性映射, 所以同态是平凡的概念. 同态这个词表明了两个线性空间的相似性, 一个空间丢掉一些东西之后就可以被看成另一个空间的子空间. 而满同态则是说, 丢掉一些东西之后, 这个空间就是另一个空间. 比如说, 如果  $m < n$ , 丢掉  $\mathbb{R}^n$  中元素的后面  $n - m$  个分量, 就得到了  $\mathbb{R}^m$ , 这就是

一个满同态. 同构则是说, 这两个线性空间就是一样的, 没有谁比谁更复杂, 比如说,  $\mathbb{R}^n$  和  $\mathbb{R}^m$  就是同构的, 只要  $n = m$ .

刚刚讨论的  $\mathbb{R}^m$  与  $\mathbb{R}^n$  的同构是具有一般性的, 这就是有限维线性空间的同构定理:

**定理 A.3 (有限维线性空间的同构定理)** 设  $V$  和  $W$  是有限维  $F$ -线性空间, 则  $V$  与  $W$  同构当且仅当  $\dim V = \dim W$ .

这一定理充分说明了, 有限维线性空间中维数的意义: 维数刻画了线性空间.

**证明.** 这一证明的思路是典型的: 先定义一个 (与基相关的) 基本映射, 然后进行扩张.

$\Leftarrow$ : 如果  $\dim V = \dim W = n$ , 那么两个线性空间的基的元素个数是一样的, 我们可以将它们一一对应起来. 比方说  $V$  的基是  $\{v_1, \dots, v_n\}$ ,  $W$  的基是  $\{w_1, \dots, w_n\}$ , 那么我们可以定义两个基之间的映射  $f$ , 使得  $f(v_i) = w_i$ .

我们可以将  $f$  扩张成一个线性映射. 比如说, 对于任意的  $x \in V$ , 它用基表示就是  $\sum_{i=1}^n a_i v_i$ . 我们可以定义

$$f(x) = f\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i f(v_i) = \sum_{i=1}^n a_i w_i.$$

接下来验证  $f$  是  $V$  到  $W$  同构. 首先, 按照定义就可以验证这是一个线性映射. 其次, 因为  $a_i$  是任意的, 所以这显然也是一个满射. 最后, 如果有两个不同的  $x, y$  对应相同的  $f(x) = f(y)$ , 那么  $f(x)$  和  $f(y)$  的坐标是一样的, 所以  $x$  和  $y$  的坐标也是一样的, 所以  $x = y$ , 所以  $f$  也是一个单射.

$\implies$  : 设两个线性空间由映射  $f: V \rightarrow W$  给出同构. 假设  $V$  的基是  $\{v_1, \dots, v_n\}$ , 我们证明  $W$  的基就是  $\{f(v_1), \dots, f(v_n)\}$ .

首先, 因为  $f$  是满射, 而  $v_i$  生成了整个  $V$ , 所以  $f(v_i)$  生成了整个  $W$ .

再说明  $f(v_i)$  线性无关. 假设  $\sum_{i=1}^n a_i f(v_i) = 0$ , 那么  $f(\sum_{i=1}^n a_i v_i) = 0$ , 由于  $f$  是单射, 所以  $\sum_{i=1}^n a_i v_i = 0$ , 由于  $v_i$  线性无关, 所以  $a_i = 0$ , 所以  $f(v_i)$  线性无关.

以上两点证明了  $W$  的基是  $\{f(v_1), \dots, f(v_n)\}$ , 所以  $\dim V = \dim W$ .  $\square$

此外, 同构还有一个重要性质:

**命题 A.2** 假设  $f: V \rightarrow W$  和  $g: W \rightarrow U$  是两个线性映射, 如果  $f$  和  $g$  都是同构, 那么  $g \circ f$  也是同构.

**证明.** 根据定义即可证明.  $\square$

接下来我们进一步研究线性映射所带来的结构. 我们刚刚说过, 同态就是说把一些东西丢掉, 剩下的东西可以被看成另一个空间的子空间. 丢掉的东西是和剩下的东西, 就是线性映射的核与像.

**定义 A.14 (核与像)** 设  $V$  和  $W$  是  $F$ -线性空间,  $f: V \rightarrow W$  是一个线性映射.  $f$  的核定义为  $\ker f = \{x \in V : f(x) = 0\}$ ,  $f$  的像定义为  $\operatorname{Im} f = \{f(x) : x \in V\}$ .  $\square$

“把一些东西丢掉”这一表述可以精确地由以下定理给出:

**定理 A.4** 设  $V$  和  $W$  是  $F$ -线性空间,  $f: V \rightarrow W$  是一个线性映射, 则  $\ker f$  是  $V$  的线性子空间,  $\operatorname{Im} f$  是  $W$  的线性子空间. 另外,

$$\dim V = \dim \ker f + \dim \operatorname{Im} f.$$

直观来说, 这一定理表明, 线性映射  $f$  把  $V$  抹掉了子空间  $\ker f$ , 最终得到了空间  $\operatorname{Im} f$ .

**证明.** 这一证明类似于定理 A.3 的证明, 这里只给出思路, 细节留给读者. 首先选出  $\ker f$  的基  $v_1, \dots, v_k$ , 然后添加向量  $u_1, \dots, u_l$  扩充成  $V$  的基, 然后证明  $f(u_1), \dots, f(u_l)$  是  $\operatorname{Im} f$  的基.  $\square$

一个直接但重要的推论是:

**推论 A.1** 设  $V$  和  $W$  是  $F$ -线性空间,  $f: V \rightarrow W$  是一个线性映射. 那么以下性质成立:

1.  $\dim \operatorname{Im} f \leq \dim V$ , 等号成立当且仅当  $\ker f = \{0\}$ ;
2.  $f$  是单射当且仅当  $\ker f = \{0\}$ ;
3.  $f$  是满射当且仅当  $\dim \operatorname{Im} f = \dim W$ ;
4.  $f$  是同构当且仅当  $\ker f = \{0\}$  且  $\dim \operatorname{Im} f = \dim W$ .

这一推论给了我们判断一个线性映射是否是单射、满射或者同构的方法.

最后, 我们引入线性映射的秩的概念:

**定义 A.15 (线性映射的秩)** 设  $V$  和  $W$  是  $F$ -线性空间,  $f: V \rightarrow W$  是一个线性映射.  $f$  的秩定义为  $\operatorname{rank} f = \dim \operatorname{Im} f$ .  $\square$

换言之, 线性映射的秩就是它的像的维数. 秩越高的线性映射说明它把空间“压缩”得越少, 丢掉的东西越少. 例如,  $\mathbb{R}^n$  的投影变换  $\pi_m$  的秩为  $m$ , 说明它丢掉的东西只有  $n - m$  维, 也就是后面的  $n - m$  个坐标.

推论 A.1 给出了复合线性映射秩的性质:

**推论 A.2** 设  $V, W$  和  $U$  是  $F$ -线性空间,  $f: V \rightarrow W$  和  $g: W \rightarrow U$  是两个线性映射, 则  $\text{rank}(g \circ f) \leq \text{rank } f$ , 等号成立当且仅当  $\text{Im } f \cap \ker g = \{0\}$ . 特别地, 如果  $f, g$  都是满射, 那么等号成立当且仅当  $g$  是同构. 此外,  $\text{rank}(g \circ f) \leq \text{rank } g$ , 如果  $f$  是满射, 那么等号成立.

**证明.** 根据推论 A.1, 我们有  $\text{rank}(g \circ f) = \dim \text{Im}(g \circ f) \leq \dim \text{Im } f = \text{rank } f$ . 等号成立当且仅当在空间  $\text{Im } f$  中  $g$  的核是  $\{0\}$ , 换言之,  $\text{Im } f \cap \ker g = \{0\}$ . 特别地, 如果  $f$  是满射, 那么这一条件变为  $\ker g = \{0\}$ , 如果  $g$  也是满射, 那么  $g$  是同构.

此外, 因为  $\text{Im } f \subseteq W$ , 所以  $\text{Im}(g \circ f) \subseteq \text{Im } g$ , 因此  $\text{rank}(g \circ f) = \dim \text{Im}(g \circ f) \leq \dim \text{Im } g = \text{rank } g$ . 如果  $f$  是满射, 那么  $\text{Im } f = W$ , 所以  $\text{Im}(g \circ f) = \text{Im } g$ , 因此等号成立.  $\square$

这一推论有非常直观的含义: 线性映射是同态, 因此会丢东西, 所以复合映射会丢更多的东西.

## §A.3 矩阵

我们已经用基与坐标表示了线性空间, 接下来, 我们引入矩阵的概念来表示线性映射, 我们这一节考虑的线性空间都是有限维的. 考虑一个线性映射  $f: V \rightarrow W$ , 如果  $V$  的基是  $\{v_1, \dots, v_n\}$ ,  $W$  的基是  $\{w_1, \dots, w_m\}$ , 那么  $f(v_i)$  可以用  $w_1, \dots, w_m$  的线性组合表示出来, 即

$$f(v_i) = a_{1i}w_1 + \dots + a_{mi}w_m.$$

我们把这些系数  $a_{ji}$  排成如下形状

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

此时  $A$  被称为矩阵,  $m = n$  的矩阵被称为方阵.  $A_i = (a_{i1}, \dots, a_{in})$

被称为矩阵  $A$  的第  $i$  行,  $A^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$  被称为矩阵  $A$  的第  $j$  列. 他们分

别被叫做行向量与列向量. 第  $i$  行第  $j$  列的元素, 也就是  $a_{ij}$ , 记为  $A_{ij}$ . 为了节约空间, 列向量通常被写成转置的形式, 即  $(a_{1j}, \dots, a_{mj})^T$ . 注意, 坐标向量都被视作列向量.

假设在基  $v_i$  下  $x \in V$  的坐标是  $X = (x_1, \dots, x_n)^T$ . 我们现在来计算  $f(x)$  在基  $w_i$  下的坐标  $Y = (y_1, \dots, y_m)^T$ . 因为  $x = x_1v_1 + \cdots + x_nv_n$ , 所以

$$\begin{aligned} f(x) &= f(x_1v_1 + \cdots + x_nv_n) \\ &= x_1f(v_1) + \cdots + x_nf(v_n) \\ &= x_1(a_{11}w_1 + \cdots + a_{m1}w_m) + \cdots + x_n(a_{1n}w_1 + \cdots + a_{mn}w_m) \\ &= (a_{11}x_1 + \cdots + a_{1n}x_n)w_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)w_m. \end{aligned}$$

因此,  $f(x)$  在基  $w_i$  下的坐标是  $(a_{11}x_1 + \cdots + a_{1n}x_n, \dots, a_{m1}x_1 + \cdots + a_{mn}x_n)$

$\cdots + a_{mn}x_n)^T$ ,  $y_i = a_{i1}x_1 + \cdots + a_{in}x_n$ . 我们将这一计算结果写作

$$Y = AX.$$

这就是矩阵与向量的乘法.

通过矩阵, 线性映射作用在向量上的结果可以被具体算出来. 从这个意义上说, 矩阵表示了线性映射. 这一观点可以用下图来表示:

$$\begin{array}{ccc} x & \xrightarrow{f} & f(x) \\ \downarrow & & \downarrow \\ X & \xrightarrow{A} & Y \end{array}$$

反之, 给定一个域  $F$  上的  $m \times n$  矩阵  $A$ , 我们可以定义一个线性映射  $f_A: F^n \rightarrow F^m$ : 对  $X \in F^n$ ,  $f_A(X) = AX$ .  $f_A$  被称为  $A$  诱导的线性映射. 所以, 矩阵本身也可以看成是一个线性映射, 不仅仅只是线性映射的表示.

我们看一个平凡的例子. 考虑零映射  $O: x \mapsto 0$ . 不管在什么基下,  $O$  的矩阵都是全零矩阵, 我们称为零矩阵, 依然使用符号  $O$  表示. 反过来, 如果一个线性映射的矩阵是零矩阵, 那么这个线性映射也是零映射.

我们再看一个的例子, 这个例子说明利用矩阵如何给出不同基之下的坐标变换公式. 设  $V$  是一个  $n$  维线性空间,  $f: V \rightarrow V$  是一个线性算子.  $V$  的一组基是  $\{v_1, \dots, v_n\}$ , 另一组基是  $\{v'_1, \dots, v'_n\}$ . 定义一个  $V$  的自同构满足  $f(v_i) = v'_i$ . 假设  $f$  在基  $\{v_i\}$  下的矩阵是  $A$ , 这被称为基  $\{v_i\}$  到基  $\{v'_i\}$  的过渡矩阵.

考虑一个点  $x \in V$ , 它在基  $\{v_i\}$  下的坐标是  $X = (x_1, \dots, x_n)^T$ , 在

基  $\{v'_i\}$  下的坐标是  $X' = (x'_1, \dots, x'_n)^\top$ . 我们来计算  $x$  在基  $\{v'_i\}$  下的坐标. 因为  $f(x) = f(x_1v_1 + \dots + x_nv_n) = x_1f(v_1) + \dots + x_nf(v_n)$ , 而

$$f(v_i) = v'_i = \sum_{j=1}^n a_{ji}v_j,$$

所以

$$\begin{aligned} f(x) &= \sum_{i=1}^n x_i \sum_{j=1}^n a_{ji}v_j = \sum_{j=1}^n \left( \sum_{i=1}^n a_{ji}x_i \right) v_j \implies \\ x &= \sum_{j=1}^n \left( \sum_{i=1}^n a_{ji}x_i \right) f^{-1}(v_j) = \sum_{j=1}^n \left( \sum_{i=1}^n a_{ji}x_i \right) v'_j. \end{aligned}$$

因此,  $X = AX'$ .

线性映射相关的概念就可以被迁移到矩阵中来.

首先我们考虑映射的线性组合. 设  $V$  是  $F$ -线性空间,  $f: V \rightarrow W$  和  $g: V \rightarrow W$  是两个线性映射,  $\lambda, \mu \in F$ , 那么  $\lambda f + \mu g$  也是一个线性映射. 如果  $V$  的基是  $\{v_1, \dots, v_n\}$ ,  $W$  的基是  $\{w_1, \dots, w_m\}$ . 假设在这些基下,  $f$  和  $g$  的矩阵分别是  $A$  和  $B$ , 那么  $\lambda f + \mu g$  的矩阵可以很自然地记作  $\lambda A + \mu B$ . 容易验证,  $\lambda A + \mu B$  的第  $i$  行第  $j$  列的元素是  $\lambda a_{ij} + \mu b_{ij}$ . 用这样的办法, 我们就定义了矩阵的数乘和加法.

然后再考虑映射的复合. 设  $V$  是  $F$ -线性空间,  $f: V \rightarrow W$  和  $g: W \rightarrow U$  是两个线性映射. 如果  $V$  的基是  $\{v_1, \dots, v_n\}$ ,  $W$  的基是  $\{w_1, \dots, w_m\}$ ,  $U$  的基是  $\{u_1, \dots, u_l\}$ . 假设在这些基下,  $g$  和  $f$  的矩阵分别是  $A$  和  $B$ , 那么复合  $gf$  的矩阵可以很自然地记作  $AB$ . 我们来计



算  $AB$  的第  $i$  行第  $j$  列的元素. 因为  $gf(v_i) = g(f(v_i))$ , 所以

$$\begin{aligned}
 gf(v_i) &= g(f(v_i)) \\
 &= g(a_{1i}w_1 + \cdots + a_{mi}w_m) \\
 &= a_{1i}g(w_1) + \cdots + a_{mi}g(w_m) \\
 &= a_{1i}(b_{11}u_1 + \cdots + b_{l1}u_l) + \cdots + a_{mi}(b_{1m}u_1 + \cdots + b_{lm}u_l) \\
 &= (a_{1i}b_{11} + \cdots + a_{mi}b_{1m})u_1 + \cdots + (a_{1i}b_{l1} + \cdots + a_{mi}b_{lm})u_l.
 \end{aligned}$$

因此,  $AB$  的第  $i$  行第  $j$  列的元素是  $a_{1i}b_{j1} + \cdots + a_{mi}b_{jm}$ . 这就是矩阵乘法的定义. 当有多个相同矩阵相乘时, 我们可以写成幂的形式. 比如,  $A^2 = AA$ ,  $A^3 = AAA$  等等.

接下来, 我们考虑同构对应的矩阵. 最简单的同构是恒等映射  $\text{id}: V \rightarrow V$ , 它的矩阵是单位矩阵  $I_n$ . 容易看出, 无论在什么基下,  $I_n$  都等于

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

有了单位矩阵, 类似算子的多项式, 给定多项式  $G(x) = a_0 + a_1x + \cdots + a_nx^n$ , 我们可以定义矩阵  $G(A) = a_0I_n + a_1A + \cdots + a_nA^n$ .

在更一般的情况下, 考虑  $V$  和  $W$  是  $n$  维的  $F$ -线性空间, 基分别是  $\{v_1, \dots, v_n\}$  和  $\{w_1, \dots, w_n\}$ . 如果线性映射  $f: V \rightarrow W$  将  $v_i$  映到  $\lambda_i w_i$ ,

那么  $f$  的矩阵就是

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

我们将这样的矩阵称为对角矩阵.

对于一般的同构映射  $f: V \rightarrow W$ , 它有一个同构逆映射  $f^{-1}: W \rightarrow V$ . 假设  $V$  的基是  $\{v_1, \dots, v_n\}$ ,  $W$  的基是  $\{w_1, \dots, w_n\}$ , 那么  $f$  和  $f^{-1}$  的矩阵分别是  $A$  和  $B$ . 我们来计算  $AB$  和  $BA$ . 因为  $f^{-1}f = \text{id}$ , 所以  $AB = I_n$ . 同理,  $BA = I_n$ .  $B$  可以被看成  $A$  的逆元, 我们记作  $B = A^{-1}$ . 这就是矩阵的逆的定义.

接下来, 我们引入矩阵转置的概念.

**定义 A.16 (矩阵转置)** 设  $A = (a_{ij})$  是一个  $m \times n$  矩阵, 我们定义  $A$  的转置为一个  $n \times m$  矩阵  $A^T = (a_{ji})$ .  $\square$

实矩阵的转置在线性映射中对应的是对偶空间的对偶映射, 这里我们不展开讨论了.

对于满足  $A^T = A$  的矩阵, 我们称之为**对称矩阵**. 对于满足  $A^T = -A$  的矩阵, 我们称之为**反对称矩阵**.

现在, 矩阵作为一个代数结构所需要的要素都已经给出了. 我们来看看矩阵的一些基本性质. 这些性质大多是从线性映射继承过来的 (命题 A.1), 我们在此略去证明.

**命题 A.3** 设  $A, B, C$  都是域  $F$  上的  $n$  阶方阵,  $\lambda, \mu \in F$ , 那么

$$1. (\lambda A + \mu B)C = \lambda AC + \mu BC;$$

2.  $A(\lambda B + \mu C) = \lambda AB + \mu AC;$
3.  $(AB)C = A(BC);$
4.  $A(B + C) = AB + AC;$
5.  $(A + B)C = AC + BC;$
6.  $AI_n = I_n A = A;$
7.  $AO = OA = O;$
8.  $A(-B) = (-A)B = A(-B) = -(AB);$
9.  $(AB)^{-1} = B^{-1}A^{-1};$
10.  $(A^{-1})^{-1} = A;$
11.  $(\lambda A)^{-1} = \lambda^{-1}A^{-1};$
12.  $(A^T)^{-1} = (A^{-1})^T;$
13.  $(A^T)^T = A;$
14.  $(A + B)^T = A^T + B^T;$
15.  $(AB)^T = B^T A^T;$
16.  $(\lambda A)^T = \lambda A^T;$
17.  $(A^{-1})^T = (A^T)^{-1};$
18.  $(A^T)^{-1} = (A^{-1})^T.$

回顾前面关于矩阵的表示论观点, 从线性映射  $f$  得到矩阵  $A$  需要基于给定的基. 一个自然的问题是, 如果我们换了基, 那么矩阵  $A$  会怎么变化? 下面我们来具体计算.

设  $f: V \rightarrow V$  是一个线性算子,  $V$  的两个基分别是  $\{v_1, \dots, v_n\}$  和  $\{v'_1, \dots, v'_n\}$ . 假设在这两个基下,  $f$  的矩阵分别是  $A = (a_{ij})$  和  $A' = (a'_{ij})$ . 我们来计算  $A'$  和  $A$  的关系.

假设  $g$  是一个自同构, 使得  $g(v_i) = v'_i$ , 在  $\{v_1, \dots, v_n\}$  下对应的矩阵是  $B$ . 我们下面证明  $A' = B^{-1}AB$ , 注意到

$$fg(v_i) = f(g(v_i)) = f(v'_i) = \sum_j a'_{ji} v'_j = \sum_j a'_{ji} g(v_j),$$

因为  $g$  可逆, 所以

$$g^{-1}fg(v_i) = \sum_j a'_{ji} v_j.$$

左边是  $B^{-1}AB$ , 而右边对应的就是  $A'$ . 所以我们证明了

**定理 A.5** 设  $V$  是  $F$ -线性空间,  $f: V \rightarrow V$  是一个线性算子,  $\{v_1, \dots, v_n\}$  和  $\{w_1, \dots, w_n\}$  是  $V$  的两个基. 假设  $A$  和  $A'$  分别是  $f$  在这两个基下的矩阵,  $B$  是从  $\{v_1, \dots, v_n\}$  到  $\{w_1, \dots, w_n\}$  的过渡矩阵, 那么  $A' = B^{-1}AB$ .

矩阵  $A$  和  $A'$  通过可逆矩阵  $B$  联系了起来:  $A' = B^{-1}AB$ , 行如这样的矩阵关系被称为相似, 记作  $A \sim A'$ . 容易验证, 相似是一个等价关系, 也就是说:

- $A \sim A$ ;
- 如果  $A \sim A'$ , 那么  $A' \sim A$ ;

- 如果  $A \sim A'$ ,  $A' \sim A''$ , 那么  $A \sim A''$ .

根据上面的讨论, 矩阵的相似关系对应了基的变换: 如果我们把  $V$  的基换成  $V$  的另一个基, 那么线性算子  $f$  的矩阵也会变成相似的另一个矩阵.

最后, 我们讨论矩阵的秩. 同样, 这一定义来自线性映射的秩:

**定义 A.17 (矩阵的秩, 行空间, 列空间)** 设  $A$  是一个  $m \times n$  矩阵,  $f_A: F^n \rightarrow F^m$  是它诱导的线性映射.

- 定义  $A$  的秩为  $f_A$  的秩, 记为  $\text{rank } A$ .
- $A$  的  $m$  个行向量生成了一个线性空间, 称为矩阵  $A$  的行空间; 类似地, 所有的列向量生成了一个线性空间, 称为矩阵  $A$  的列空间.
- 行空间的维数称为矩阵  $A$  的行秩, 列空间的维数称为矩阵  $A$  的列秩.
- 如果行秩等于  $m$ , 即行空间的一组基就是所有行向量, 那么我们称  $A$  是行满秩的; 如果列秩等于  $n$ , 即列空间的一组基就是所有列向量, 那么我们称  $A$  是列满秩的.
- 对于方阵来说, 如果它同时是行满秩的和列满秩的, 那么我们称它是满秩的. □

矩阵的秩最核心的定理是:

**定理 A.6** 设  $A$  是一个  $m \times n$  矩阵,  $A$  的行秩、列秩与秩都相等.

这一定理的证明通常涉及到矩阵的初等行变换, 也可以用对偶空间理论的证明, 我们这里就不给出了.

从线性映射复合的秩关系（推论 A.2），我们直接得到了矩阵乘法秩的性质：

**命题 A.4** 设  $A$  是一个  $m \times n$  矩阵， $B$  是一个  $n \times p$  矩阵，那么

$$\text{rank}(AB) \leq \min\{\text{rank } A, \text{rank } B\}.$$

如果  $A$  和  $B$  都是方阵，那么当  $A$  可逆时， $\text{rank}(AB) = \text{rank } B$ ；当  $B$  可逆时， $\text{rank}(AB) = \text{rank } A$ 。

## §A.4 双线性型与二次型

本节考虑线性函数的一种推广，即双线性函数。它是两个变量的函数，而且每个变量都是线性的。这种函数在几何上有很多应用，比如内积。我们先来给出它的定义。

**定义 A.18 (双线性型)** 设  $V$  是一个  $F$ -线性空间，如果  $V$  上有一个映射  $f: V \times V \rightarrow F$ ，满足

1. 对于任意的  $v \in V$ ， $f(v, \cdot): V \rightarrow F$  是一个线性映射；
2. 对于任意的  $w \in V$ ， $f(\cdot, w): V \rightarrow F$  是一个线性映射。

那么称  $f$  是  $V$  上的一个双线性型。 □

类似线性映射，我们的首要任务是表示一个双线性型。实际上，双线性型也可以用矩阵来表示。选定一组  $V$  的基  $\{v_1, \dots, v_n\}$ ，任意给定

两个向量  $x = \sum_{i=1}^n x_i v_i$  和  $y = \sum_{i=1}^n y_i v_i$ , 我们有

$$\begin{aligned} f(x, y) &= f\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i y_j f(v_i, v_j). \end{aligned}$$

如果我们知道了  $f(v_i, v_j)$ , 那么  $f(x, y)$  就完全可以用  $x, y$  的坐标表示出来. 这就是说, 我们可以用矩阵来表示双线性型. 定义  $A_{ij} = f(v_i, v_j)$ , 我们将矩阵  $A = (A_{ij})$  称为  $f$  在基  $\{v_1, \dots, v_n\}$  下的矩阵. 假设  $x, y$  的坐标是  $X, Y$ , 那么我们有

$$f(x, y) = X^T A Y.$$

反过来, 如果给定了一个  $n$  阶方阵  $A$ , 那么我们可以定义一个双线性型  $f_A$ , 使得  $f_A(v_i, v_j) = a_{ij}$ , 这样的双线性型称为由矩阵  $A$  诱导的双线性型. 于是, 双线性型就和矩阵一一对应了.

类似线性映射, 双线性映射关心的一个重要问题是基变换. 设  $A$  是一个双线性型  $f$  在基  $\{v_1, \dots, v_n\}$  下的矩阵, 而  $A'$  是基  $\{v'_1, \dots, v'_n\}$  下的矩阵, 假设  $\{v_i\}$  到  $\{v'_i\}$  的过渡矩阵是  $P$ . 现在任取  $x, y \in V$ , 他们在基  $\{v_i\}$  下的坐标分别是  $X, Y$ , 在基  $\{v'_i\}$  下的坐标分别是  $X', Y'$ , 那么我们有

$$f(x, y) = X^T A Y = X'^T A' Y'.$$

根据坐标的基变换公式,  $X = P X'$ ,  $Y = P Y'$ , 所以

$$X^T A Y = (A X')^T P (A Y') = X'^T (A^T P A) Y'.$$

由于  $x, y$  是任意的,  $X, Y$  也是任意的, 联立上面两式,  $A' = A^T P A$ , 这就是基变换公式. 对应到矩阵中, 这被称为合同变换.

**定义 A.19 (合同矩阵)** 设  $A, B$  都是  $n$  阶方阵, 如果存在一个可逆方阵  $P$ , 使得  $B = P^T A P$ , 那么称  $A$  和  $B$  是合同的.  $\square$

容易验证, 合同关系是一个等价关系, 这与相似关系是类似的. 根据命题 A.4, 可逆矩阵相乘不改变矩阵的秩, 所以合同矩阵的秩是相同的. 于是, 双线性型的任意矩阵表示都有相同的秩, 我们因此可以定义双线性型的秩:

**定义 A.20 (双线性型的秩)** 设  $f$  是  $V$  上的一个双线性型, 如果  $f$  在某个基下的矩阵的秩是  $r$ , 那么称  $f$  的秩是  $r$ .  $\square$

接下来, 我们考虑一种特殊的双线性型: 对称双线性型, 即  $f(x, y) = f(y, x)$  对任意  $x, y \in V$  成立. 注意到, 对称双线性型对应的矩阵是对称矩阵. 我们现在令  $x = y$ , 定义  $q(x) = f(x, x)$ , 那么  $q$  是一个实值函数, 这样的函数便是二次型.

**定义 A.21 (二次型)** 设  $V$  是数域  $F$  上的线性空间,  $f$  是  $V$  上的一个对称双线性型, 那么定义  $q: V \rightarrow F$  为  $q(x) = f(x, x)$ , 称  $q$  是  $f$  诱导的二次型.  $\square$

自然, 秩的概念也可以被迁移到二次型上:

**定义 A.22 (二次型的秩)** 设  $q$  是  $V$  上的一个二次型, 如果定义  $q$  的秩为诱导它的双线性型  $f$  的秩.  $\square$

实际上, 二次型本身也可以算出双线性型: 设  $q$  是二次型, 那么我们可以定义

$$f(x, y) = \frac{1}{2}(q(x+y) - q(x) - q(y)).$$



容易验证,  $f$  是一个对称双线性型. 于是, 二次型和对称双线性型是一一对应的.

自然, 二次型也可以用坐标表示. 选定一组基, 假设二次型  $q$  对应的对称双线性型是  $f$ , 那么  $q$  在这组基下的矩阵就是  $f$  在这组基下的矩阵. 这样, 二次型就和对称矩阵一一对应了. 如果再给定  $x \in V$  的坐标  $X$ , 那么  $q(x) = X^T A X$ . 如果把它展开写, 就是

$$q(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

这是一个关于  $x_i$  的二次函数. 通常来说, 我们希望化简这一表示, 如果能写成  $\sum_{i=1}^n a_{ii} x_i^2$  的形式, 那么计算都会变得非常容易. 此时, 二次型  $q$  的矩阵是对角矩阵.

**定义 A.23 (规范型)** 设  $V$  是数域  $F$  上的线性空间,  $q$  是  $V$  上的一个二次型, 如果存在一组基, 使得  $q$  在这组基下的矩阵是对角矩阵, 那么称这个对角矩阵是  $q$  的规范型, 这组基是规范基.  $\square$

二次型的一个核心定理是, 规范型总是存在:

**定理 A.7** 设  $V$  是  $F$ -线性空间,  $q$  是  $V$  上的一个二次型, 那么  $q$  存在规范型.

**证明.** 首先, 选定一组基  $\{v_1, \dots, v_n\}$ , 保证  $q(v_1) \neq 0$  (这样的  $v_1$  一定存在, 否则  $q$  就是零映射, 自然有规范型) 我们想办法把它变成另外一组基, 使得二次型  $q$  的矩阵是对角矩阵. 假设  $q$  对应的双线性型是  $f$ . 对维数  $n$  用归纳法.

如果  $n = 1$ , 这是显然的.

现在考虑一般的  $n$ , 我们想办法将矩阵的第一行和第一列非对角的元素都变成 0, 那么剩下的矩阵实际上就是在一个  $n-1$  维空间上的双线性型, 于是就可以用归纳假设了. 注意到这些元素其实就是  $f(v_1, v_j)$ , 所以只需要把他们变成 0 就可以了.

注意到  $f(v_1, \cdot)$  是非零线性函数, 秩是 1, 所以  $\dim \ker f(v_1, \cdot) = n-1$ , 于是我们选出核的基  $v'_2, \dots, v'_n$ . 另外  $v_1 \notin \ker f(v_1, \cdot)$ , 所以  $v_1$  与  $\{v'_2, \dots, v'_n\}$  线性无关, 于是  $\{v_1, v'_2, \dots, v'_n\}$  是一组基. 根据核的定义, 此时  $f(v_1, v'_i) = 0$ , 因此这样就把第一行非对角的元素都变成了 0.  $\square$

相应地, 在矩阵上, 这一定理的表述为:

**推论 A.3** 任何对称矩阵  $A$  都合同于一个对角矩阵.

在实数域上, 这一定理还可以被加强:

**定理 A.8 (惯性定理)** 设  $V$  是  $\mathbb{R}$  上的  $n$  维线性空间,  $q$  是  $V$  上的一个二次型, 那么  $q$  存在形如

$$q(x) = \sum_{i=1}^r \lambda_i x_i^2$$

的规范型 ( $(x_i)$  是  $x$  的坐标), 其中  $\lambda_i \in \{1, -1\}$ ,  $r$  是  $q$  的秩, 且  $\lambda_i$  中 1 的个数和  $-1$  的个数只依赖于  $q$ , 不依赖于规范基的选取.

这一定理我们就不再给出证明了.

惯性定理给出了几类特殊的二次型:

**定义 A.24 (正定, 半正定, 负定, 半负定)** 设  $V$  是  $\mathbb{R}$  上的  $n$  维线性空间,  $q$  是  $V$  上的一个二次型.

- 如果  $q$  的规范型是

$$q(x) = \sum_{i=1}^n x_i^2,$$

那么称  $q$  是正定的.

- 如果  $q$  的规范型是

$$q(x) = \sum_{i=1}^r x_i^2 \quad (r \leq n),$$

那么称  $q$  是半正定的.

- 如果  $q$  的规范型是

$$q(x) = -\sum_{i=1}^r x_i^2,$$

那么称  $q$  是负定的.

- 如果  $q$  的规范型是

$$q(x) = -\sum_{i=1}^r x_i^2 \quad (r \leq n),$$

那么称  $q$  是半负定的.

对于实对称矩阵  $A$ , 如果它对应的二次型是正定/半正定/负定/半负定的, 那么称  $A$  是正定/半正定/负定/半负定的.  $\square$

以上概念都可以直接用二次型的取值去等价定义:

**命题 A.5** 设  $V$  是  $\mathbb{R}$  上的  $n$  维线性空间,  $q$  是  $V$  上的一个二次型.

- $q$  是正定的当且仅当对任意的非零向量  $x \in V$ , 都有  $q(x) > 0$ .

- $q$  是半正定的当且仅当对任意的非零向量  $x \in V$ , 都有  $q(x) \geq 0$ .
- $q$  是负定的当且仅当对任意的非零向量  $x \in V$ , 都有  $q(x) < 0$ .
- $q$  是半负定的当且仅当对任意的非零向量  $x \in V$ , 都有  $q(x) \leq 0$ .

证明. 选定一组规范基, 按照定义验证即可.  $\square$

自然, 这一命题的矩阵版本也是成立的. 一个直接的推论是:

**推论 A.4** 设  $A$  是一个矩阵, 那么  $A^T A$  和  $AA^T$  都是半正定的. 此外,  $B$  是一个正定矩阵当且仅当存在可逆矩阵  $P$ , 使得  $B = P^T P$ .

## §A.5 带内积的线性空间

内积的考虑是从几何中来的. 一个典型的例子是平面欧氏几何. 我们知道, 笛卡尔的平面解析几何等价于平面欧氏几何. 建立坐标系的过程实际上就是选定了一个基, 而坐标就是基的坐标. 在这个基下, 平面上的点可以用坐标表示. 然而, 并不是所有的坐标轴都是好计算的, 我们考虑的是互相垂直的坐标轴, 此时, 平面上点的坐标就完全可以用投影来表示了. 计算投影的过程实际上就是内积的过程. 将平面解析几何的内积定义一般化, 我们就得到了线性空间的内积.

**定义 A.25 (内积)** 设  $V$  是一个实线性空间, 如果  $V$  上有一个对称双线性型  $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ , 它诱导的二次型是正定的, 那么称  $\langle \cdot, \cdot \rangle$  是  $V$  上的一个内积, 称  $V$  是一个内积空间.  $\square$

注意, 讨论内积的时候, 我们只考虑实线性空间, 这是因为实数可以比大小, 并且不会像有理数那样对根号不封闭, 所以可以定义模长. 自然,  $\mathbb{R}^n$  是内积空间, 因为我们可以定义  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ .

利用内积, 我们可以定义模长.

**定义 A.26 (模)** 设  $V$  是一个内积空间,  $v \in V$ , 定义  $v$  的模或内积诱导的范数为  $\|v\| = \sqrt{\langle v, v \rangle}$ .  $\square$

容易证明, 向量的模等于零当且仅当它是零向量, 此外, 对任意的  $v \in V$  和  $\lambda \in \mathbb{R}$ , 有  $\|\lambda v\| = |\lambda| \|v\|$ . 模长为 1 的向量称为单位向量.

反过来, 内积诱导的范数也可以表示内积:

$$\langle v, w \rangle = \frac{1}{4} (\|v + w\|^2 - \|v - w\|^2). \quad (\text{A.1})$$

利用内积, 我们可以推广平面几何中的各种概念. 首先是垂直的概念.

**定义 A.27 (正交)** 设  $V$  是一个内积空间,  $v, w \in V$ , 如果  $\langle v, w \rangle = 0$ , 那么称  $v$  与  $w$  正交, 记作  $v \perp w$ .  $\square$

对于一般情况, 两个向量会有夹角的概念, 我们可以利用内积来定义.

**定义 A.28 (夹角)** 设  $V$  是一个内积空间,  $v, w \in V$ , 如果  $\theta \in [0, \pi]$  满足

$$\cos \theta = \frac{\langle v, w \rangle}{\|v\| \|w\|},$$

那么称  $\theta$  是  $v$  与  $w$  的夹角.  $\square$

夹角对任意非零向量都可以定义, 这是因为内积有 Cauchy 不等式:

**定理 A.9 (Cauchy 不等式)** 设  $V$  是一个内积空间,  $v, w \in V$ , 那么有

$$|\langle v, w \rangle| \leq \|v\| \|w\|.$$

**证明.** 取  $\lambda \in \mathbb{R}$ , 那么

$$0 \leq \langle v + \lambda w, v + \lambda w \rangle = \|v\|^2 + 2\lambda \langle v, w \rangle + \lambda^2 \|w\|^2.$$

将最右边看作是  $\lambda$  的函数, 这是一个二次函数, 因为它恒大于等于 0, 所以判别式  $\Delta \leq 0$ , 即

$$4 \langle v, w \rangle^2 - 4 \|v\|^2 \|w\|^2 \leq 0 \iff |\langle v, w \rangle| \leq \|v\| \|w\|. \quad \square$$

利用 Cauchy 不等式, 我们可以证明模长满足三角不等式:

**定理 A.10 (三角不等式)** 设  $V$  是一个内积空间,  $v, w \in V$ , 那么有

$$\|v + w\| \leq \|v\| + \|w\|.$$

**证明.** 由 Cauchy 不等式, 我们有

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle \\ &= \|v\|^2 + 2 \langle v, w \rangle + \|w\|^2 \\ &\leq \|v\|^2 + 2 \|v\| \|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2. \end{aligned} \quad \square$$

利用以上性质, 容易验证, 模实际上给了  $V$  一个范数. 关于范数的详细讨论, 见附录 B.1.1. 对于一般的范数, 我们无法像 (A.1) 一样去定义内积, 所以内积有它独特的性质.

内积还给出了投影的概念:

**定义 A.29 (投影)** 设  $V$  是一个内积空间,  $v, w \in V$ , 如果  $\lambda \in \mathbb{R}$  满足  $\langle v - \lambda w, w \rangle = 0$ , 那么称  $\lambda w$  是  $v$  在  $w$  上的投影, 其中  $\lambda = \langle v, w \rangle / \|w\|^2$ .  $\square$

接下来, 我们继续表示论的观点, 讨论内积空间中的基与坐标. 首先是正交与线性无关的关系.

**命题 A.6** 设  $V$  是一个内积空间, 两两正交的非零向量  $v_1, \dots, v_n \in V$  是线性无关的.

**证明.** 设  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  满足  $\sum_{i=1}^n \lambda_i v_i = 0$ , 那么

$$0 = \left\langle \sum_{i=1}^n \lambda_i v_i, v_j \right\rangle = \sum_{i=1}^n \lambda_i \langle v_i, v_j \rangle = \lambda_j \|v_j\|^2.$$

因为  $v_j \neq 0$ , 所以  $\lambda_j = 0$ , 这就证明了线性无关性.  $\square$

两两正交的基被称为**正交基**, 如果正交基的每个向量都是单位向量, 那么称为**标准正交基**. 在内积空间中, 基存在性定理 (定理 A.1) 可以被加强为标准正交基存在性定理:

**定理 A.11 (标准正交基存在性定理)** 设  $V$  是一个有限维内积空间, 那么  $V$  中存在一个标准正交基.

我们只提示这一定理的证明思路. 首先选择一个基, 然后利用 Gram-Schmidt 正交化方法将它正交化, 再将它单位化.

我们来看看这一定理的用处. 首先, 它给出了计算坐标的简易方式.

**命题 A.7** 设  $V$  是一个内积空间, 它的标准正交基是  $e_1, \dots, e_n$ ,  $v \in V$ , 那么

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

另外, 成立勾股定理:

$$\|v\|^2 = \sum_{i=1}^n \langle v, e_i \rangle^2.$$

在标准正交基下, 投影的系数就是坐标. 此时, 内积也可以被写成矩阵的形式. 假设  $x, y \in V$ , 他们的坐标分别是  $X$  和  $Y$ , 那么  $\langle x, y \rangle = X^T Y$ .

使用标准正交基的另一个好处是, 线性函数的表示变得简单了. 给定  $V$  的一个标准正交基  $\{e_i\}_{i=1}^n$  和  $V$  上的线性函数  $f$ , 对任意一个向量  $v = \sum_{i=1}^n v_i e_i$ ,

$$f(v) = \sum_{i=1}^n v_i f(e_i).$$

考虑向量  $e_f = \sum_{i=1}^n f(e_i) e_i$ , 容易验证  $f(v) = \langle e_f, v \rangle$ .

另一方面, 这样的  $e_f$  必定是唯一的. 如果有两个  $e_f, e'_f$  使得  $\langle e_f, v \rangle = \langle e'_f, v \rangle = f(v)$  对任意的  $v$  都成立, 那么取  $v = e_f - e'_f$  得

$$\langle e_f - e'_f, e_f - e'_f \rangle = 0 \iff e_f - e'_f = 0.$$

综上, 我们可以用一个向量的内积来表示线性函数:

**定理 A.12 (Riesz 表示定理)** 设  $V$  是一个内积空间,  $f$  是  $V$  上的一个线性函数, 那么存在唯一的向量  $u \in V$ , 使得  $f(v) = \langle u, v \rangle$  对任意的  $v \in V$  成立.

反之, 给定一个向量  $u$ , 很容易验证,  $f_u(\cdot) = \langle u, \cdot \rangle$  就是一个线性函数. 如此我们就给出了内积空间中的线性函数和向量的一一对应.

此外, 利用标准正交基, 内积空间中一组向量的线性无关性也可以用内积来判断, 这就是 *Gram* 矩阵.



**定义 A.30 (Gram 矩阵)** 给定内积空间  $V$  的一组向量  $v_1, \dots, v_k$ , 定义他们的 **Gram 矩阵** 为  $G = (\langle v_i, v_j \rangle)_{k \times k}$ .  $\square$

利用标准正交基, 很容易计算 Gram 矩阵. 我们其实已经见过这样的例子. 给定任意一个实矩阵  $A$ ,  $A^T A$  就是列向量的 Gram 矩阵,  $AA^T$  就是行向量的 Gram 矩阵.

Gram 矩阵的基本性质是:

**命题 A.8** 设  $V$  是一个内积空间,  $v_1, \dots, v_k \in V$ , 他们的 Gram 矩阵为  $G$ , 那么

1.  $G$  是对称矩阵;
2.  $G$  是半正定的;
3.  $v_1, \dots, v_k$  线性无关当且仅当  $G$  正定.

**证明.** 1. 显然.

2. 考虑  $\mathbb{R}^k$  上的二次型  $f(x) = x^T G x$ . 对任意的  $x \in \mathbb{R}^k$ , 有

$$x^T G x = \sum_{i,j=1}^k x_i x_j \langle v_i, v_j \rangle = \left\langle \sum_{i=1}^k x_i v_i, \sum_{j=1}^k x_j v_j \right\rangle = \left\| \sum_{i=1}^k x_i v_i \right\|^2 \geq 0.$$

因此  $G$  是半正定的.

3. 我们已经证明  $f(x) \geq 0$ . 由命题 A.5,  $G$  是正定的当且仅当等价式  $f(x) = 0 \iff x = 0$  成立. 而  $f(x) = \left\| \sum_{i=1}^k x_i v_i \right\|^2$ , 因此  $f(x) = 0 \iff \sum_{i=1}^k x_i v_i = 0$ , 所以  $x = 0 \iff v_1, \dots, v_k$  线性无关.  $\square$

利用第三点, 我们可以很容易地判断一组向量的线性无关性.

半正定和对称性还暗示着,  $G$  可以形成某种半正定的二次型. 我们在证明中已经给出这样的二次型  $f$ . 在附录 C.3.3, 我们会遇到这样的例子, 即一族随机变量的协方差矩阵.

向量组之间的正交性也可以用 Gram 矩阵来刻画:

**命题 A.9** 设  $V$  是一个内积空间,  $v_1, \dots, v_k \in V$ , 那么下列命题等价:

1.  $v_1, \dots, v_k$  两两正交;
2.  $G$  是对角矩阵.

特别地,  $v_1, \dots, v_k$  是标准正交的当且仅当  $G = I_k$ .

这一命题的证明是显然的.

内积空间中, 直和分解也可以被加强. 为此, 我们先引入正交补的概念.

**定义 A.31 (正交补)** 设  $V$  是一个内积空间,  $W \subseteq V$ , 那么  $W$  的正交补是

$$W^\perp = \{v \in V : \forall w \in W, \langle v, w \rangle = 0\}. \quad \square$$

**定理 A.13** 设  $V$  是一个内积空间,  $W \subseteq V$  是一个有限维子空间, 那么

$$V = W \oplus W^\perp, \quad (W^\perp)^\perp = W.$$

这一定理的证明思路类似定理 A.4 的证明, 区别是这里需要扩充标准正交基. 这里不再给出具体证明.

最后, 我们考虑标准正交基之间的过渡矩阵. 设  $V$  是一个内积空间,  $e_1, \dots, e_n$  和  $e'_1, \dots, e'_n$  都是  $V$  的标准正交基. 设  $e'_j = a_{1j}e_1 + \dots + a_{nj}e_n$ , 如此就得到了过渡矩阵  $A$ . 我们来看看  $A$  的性质.

**命题 A.10** 设  $V$  是一个内积空间,  $e_1, \dots, e_n$  和  $e'_1, \dots, e'_n$  都是  $V$  的标准正交基. 设  $e_i$  到  $e'_i$  的过渡矩阵是  $A$ , 那么  $AA^T = A^T A = I_n$ .

**证明.** 设  $A = (a_{ij})$ , 那么

$$\begin{aligned} \langle e'_i, e'_j \rangle &= \left\langle \sum_{k=1}^n a_{ki} e_k, \sum_{l=1}^n a_{lj} e_l \right\rangle \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{ki} a_{lj} \langle e_k, e_l \rangle \\ &= \sum_{k=1}^n a_{ki} a_{kj}. \end{aligned}$$

当  $i = j$ , 上式就是  $\sum_{k=1}^n a_{ki}^2 = 1$ , 当  $i \neq j$ , 上式就是  $\sum_{k=1}^n a_{ki} a_{kj} = 0$ . 这就证明了  $AA^T = I_n$ . 同理可证  $A^T A = I_n$ .  $\square$

我们将满足  $A^T A = AA^T = I_n$  的矩阵称为正交矩阵, 它的逆矩阵就是它的转置矩阵.

正交矩阵有很多等价的刻画:

**定理 A.14** 设  $A$  是一个  $n$  阶方阵. 下列陈述等价:

1.  $A$  是一个正交矩阵.
2. 对任意  $v \in \mathbb{R}^n$ , 都有  $\|Av\| = \|v\|$ .
3.  $A$  的行向量是两两正交的单位向量.
4.  $A$  的列向量是两两正交的单位向量.
5.  $A^T = A^{-1}$ .

**证明.** 按照定义写出即可.  $\square$

利用基变换, 我们马上可以得到以下正交矩阵的性质:

**命题 A.11** 设  $A, B$  是  $n$  阶正交矩阵, 那么

1.  $I_n$  是正交矩阵;
2.  $AB$  是正交矩阵;
3.  $A^{-1}$  是正交矩阵;

这些性质使得正交矩阵构成了一个群, 称为正交群, 记作  $O(n)$ . 这超出了本书的范围, 我们就不继续深入了.

接下来, 我们讨论内积空间的同构.

**定义 A.32 (等距映射与等距同构)** 设  $V$  和  $W$  都是内积空间,  $T: V \rightarrow W$  是一个线性映射, 如果对任意的  $v_1, v_2 \in V$ , 都有  $\langle Tv_1, Tv_2 \rangle = \langle v_1, v_2 \rangle$ , 那么称  $T$  是一个等距映射. 如果  $T$  是一个双射, 那么称  $T$  是一个等距同构. □

内积空间的同构是线性空间同构的加强版, 因为它还要求保持内积.

同样, 内积空间的等距同构类只取决于维数:

**定理 A.15** 设  $V$  和  $W$  都是有限维内积空间, 那么  $V$  与  $W$  等距同构当且仅当  $\dim V = \dim W$ .

**证明.** 证明完全类似定理 A.3 的证明, 此时将  $V$  的标准正交基一一对应到  $W$  的标准正交基上. □

等距同构对应的矩阵恰好就是正交矩阵.

**定理 A.16** 设  $V$  和  $W$  都是有限维内积空间,  $f: V \rightarrow W$  是一个等距同构, 那么在  $V$  和  $W$  的标准正交基之下  $f$  是一个正交矩阵. 反之,  $n$  阶正交矩阵  $A$  诱导的线性映射  $f_A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是一个等距同构.

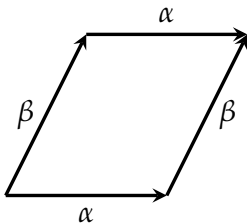
**证明.** 后半部分(“反之”之后)由定理 A.14 的第二点直接得出. 我们来证明前半部分.

设  $v_1, \dots, v_n$  是  $V$  的标准正交基,  $w_1, \dots, w_n$  是  $W$  的标准正交基,  $A$  是  $f$  在这两组基下的矩阵. 考虑任意一个点  $x \in V$ , 它的坐标是  $X$ . 那么  $f(x)$  的坐标是  $AX$ . 由于  $f$  是等距同构, 所以  $\|f(x)\| = \|x\|$ , 根据命题 A.7, 这等价于  $\|AX\| = \|X\|$ , 由定理 A.14 的第二点,  $A$  是一个正交矩阵.  $\square$

## §A.6 行列式

行列式可以进一步理解为矩阵的表示: 将很多个数的矩阵压缩到一个数. 我们将从几何观点讨论, 先从平面开始.

考虑平面  $\mathbb{R}^2$  上的两个向量  $\alpha = (a_1, a_2)^T$  和  $\beta = (b_1, b_2)^T$ , 我们可以用这两个向量作为平行四边形的两条边, 构造一个平行四边形:



现在我们定义这个平行四边形的有向面积. 数值上, 有向面积就是我们通常理解的平行四边形面积. 有向面积的符号按照如下的规则给

出. 从  $\alpha$  旋转到  $\beta$  所在的方向, 转动一个平角以内的角度. 如果这个角度是逆时针的, 面积就是正的, 否则就是负的.

容易算出, 这个有向面积是  $a_1b_2 - a_2b_1$ , 可以使用如下形象的记号表示:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \begin{vmatrix} \alpha & \beta \end{vmatrix}.$$

我们可以把这个面积的计算推广到  $n$  维空间中去. 设  $A$  是一个  $n$  阶方阵, 它的列向量是  $A^1, \dots, A^n$ , 我们考虑这些列向量张成的  $n$  维平行体:

$$\Pi(A) = \{x_1A^1 + \dots + x_nA^n : 0 \leq x_i \leq 1\}.$$

平行体的体积可以归纳定义. 一维的情况下, 向量就是实数, 这个实数的绝对值就作为体积. 假设已经定义了  $n-1$  维平行体的体积, 那么  $n$  维平行体的体积就是它的底的体积乘以高. 我们需要定义什么是底什么是高. 底是  $n-1$  维平行体  $\Pi(A^1, \dots, A^{n-1})$ . 为定义高, 先将  $A^n$  投影到  $\text{Span}(A^1, \dots, A^{n-1})$ , 投影  $A_*^n$  就是垂足, 高就是  $A^n - A_*^n$  的长度. 这样, 我们就得到了  $n$  维平行体的体积的定义.

行列式与有向体积有关. 我们这里不会专门定义有向体积的概念, 只给出一个二维情况下的直观. 通常  $\mathbb{R}^2$  的标准正交基  $e_1, e_2$ , 从  $e_1$  到  $e_2$  是逆时针的. 我们之前定义有向面积的时候也遵循了这样的原则: 与  $e_1, e_2$  形成相同定向 (即顺逆时针) 的面积是正的, 与  $e_1, e_2$  形成相反定向的面积是负的. 定向这一概念本质地反映了一组向量的顺序.

对于一般情况, 向量  $A^1, \dots, A^n$  和  $\mathbb{R}^n$  的标准正交基  $e_1, \dots, e_n$  之间同定向时  $\Pi(A)$  的体积是正的, 否则就是负的. 这样, 我们就定义了有向体积. 我们指出, 行列式的定义恰好就给出了有向体积的计算公式. 下面我们给出行列式的定义.

**定义 A.33 (行列式)** 设  $A$  是一个  $n$  阶方阵,  $A$  的行列式  $\det A$  归纳定义为

1.  $n = 1$  时,  $\det A = A_{11}$ ;
2.  $\det A = A_{11} \det R_{11} - A_{12} \det R_{12} + \cdots + (-1)^{1+n} A_{1n} \det R_{1n}$ ,  
其中  $R_{ij}$  是  $A$  去掉第  $i$  行第  $j$  列以后得到的矩阵.  $\square$

下面仅罗列一些行列式的性质, 但不给出证明.

**命题 A.12** 设  $A, B$  是两个  $n$  阶方阵, 则

1.  $\det I_n = 1$ ;
2.  $\det(AB) = (\det A)(\det B)$ ;
3.  $\det(\lambda A) = \lambda^n \det A$ ;
4.  $\det(A^{-1}) = (\det A)^{-1}$ ;
5.  $\det A = 0$  当且仅当  $A$  不可逆;
6.  $\det(A^T) = \det A$ ;

矩阵的行列式自然地定义了线性算子的行列式. 这一定义是良定义的, 因为根据定理 A.5, 如果  $A$  和  $A'$  是同一个线性算子在不同基下的矩阵, 由过渡矩阵  $B$  给出, 那么  $A' = B^{-1}AB$ , 因此根据命题 A.12,

$$\det A' = \det(B^{-1}AB) = (\det B^{-1})(\det A)(\det B) = (\det B^{-1})(\det B)(\det A) = \det A.$$

因而行列式也是刻画线性算子的一个不变量.

**定义 A.34 (线性算子的行列式)** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个线性算子,  $A$  是  $f$  在  $V$  的一个基下的矩阵, 则  $f$  的行列式  $\det f$  定义为  $\det A$ .  $\square$

线性算子的行列式表明了行列式的几何意义. 一个线性算子将一个平行体每条边 (也就是基向量) 映射到另一个平行体的每条边, 如此就将一个平行体映射到了另一个. 原来平行体的体积在这个线性算子作用后会发生改变, 这个变化的比率就是行列式.

具体来说, 我们选择  $V$  的标准正交基  $e_1, \dots, e_n$ , 他们的坐标是  $E_1, \dots, E_n$ . 考虑线性算子  $f$ , 它对应的矩阵是  $A$ . 现在考虑单位立方体 (自然也是平行体)  $\Pi(e_1, \dots, e_n)$ , 那么在映射作用下平行体的边就变成了  $\Pi(f(e_1), \dots, f(e_n))$ . 因为是  $\{e_i\}$  是标准正交基, 所以这些平行体实际上也可以写作  $\Pi(E_1, \dots, E_n)$  和  $\Pi(AE_1, \dots, AE_n)$ , 注意到, 后者实际上就是  $\Pi(A^1, \dots, A^n)$ , 考虑他们的有向体积, 这正好是行列式的定义.

作为一个注记, 从线性算子的角度来看, 行列式的性质 (命题 A.12) 就是非常自然的结果了. 我们将矩阵对应的线性映射写出. 那么, 除了最后一条, 命题 A.12 的性质都可以逐一解释:

1. 恒等算子不改变有向体积, 所以行列式为 1;
2. 两个算子  $f_A$  和  $f_B$  的复合对有向体积的改变是累乘, 即先按比例  $\det f_A$  变, 再按比例  $\det f_B$  变, 因此矩阵乘积的行列式等于行列式的乘积;
3.  $\lambda A$  对应的算子就是  $A$  算子作用后再按照  $\lambda$  的比率等比例伸缩, 对于一个  $n$  维图形来说, 这样的变化对有向体积的改变是  $\lambda^n$ ;



4. 考虑可逆线性算子  $f$ , 那么, 先作用  $f$  再作用  $f^{-1}$ , 体积不变, 所以他们对体积变化的比率乘起来 1, 即逆的行列式等于行列式的逆;
5. 最后, 如果线性算子不可逆, 那么像是更低维的, 比如在三维空间中的二维平面, 那么显然有向体积就是 0 了, 所以不可逆的行列式为 0.

行列式还有其他的一些性质, 这里不再讨论.

## §A.7 算子范数与谱理论

本节讨论如何给算子定义范数, 以及如何利用范数来研究算子的性质, 特别是特征值相关的性质. 本节需要一些点集拓扑的知识, 请参阅附录 B.1.

考虑一个  $n$  维内积空间  $V$  以及其上的一个线性算子  $f$ . 给定一个  $V$  的标准正交基  $e_1, \dots, e_n$ , 我们可以用矩阵  $A$  来表示  $f$ . 注意到对任意一个  $x \in V$ , 假设它的坐标是  $X = (x_1, \dots, x_n)^T$ , 那么

$$\begin{aligned}
 \|f(x)\|^2 &= \|AX\|^2 \\
 &= \left( \sum_{j=1}^n \left( \sum_{i=1}^n a_{ij} \right) x_j \right)^2 \\
 &\leq \max_k \sum_{i=1}^n |a_{ik}|^2 \cdot \sum_{j=1}^n x_j^2 \\
 &= C \|X\| = C \|x\|.
 \end{aligned}$$

这里  $C = \max_k \sum_{i=1}^n |a_{ik}|^2$ , 这说明, 对给定的算子  $f$ , 我们可以找到一

个常数  $C$  使得  $\|f(x)\| \leq C \|x\|$  对任意的  $x \in V$  都成立, 因此, 我们可以定义算子  $f$  的范数为最小的这样的常数  $C$ .

**定义 A.35 (算子范数)** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个线性算子, 那么  $f$  的范数定义为

$$\|f\| = \inf \{C \geq 0 : \|f(x)\| \leq C \|x\|, \forall x \in V\}.$$

或者等价地,

$$\|f\| = \sup_{x \in V} \frac{\|f(x)\|}{\|x\|} = \sup_{\|x\|=1} \|f(x)\|. \quad \square$$

要保证定义出来的确实是范数, 我们需要验证它满足非负性、齐次性和三角不等式. 我们只证明三角不等式, 其他两个类似. 考虑两个算子  $f, g$ ,

$$\begin{aligned} \|f + g\| &= \sup_{\|x\|=1} \|(f + g)(x)\| \\ &= \sup_{\|x\|=1} \|f(x) + g(x)\| \\ &\leq \sup_{\|x\|=1} (\|f(x)\| + \|g(x)\|) \\ &\leq \sup_{\|x\|=1} \|f(x)\| + \sup_{\|x\|=1} \|g(x)\| \\ &= \|f\| + \|g\|. \end{aligned}$$

这里我们用到了  $V$  中向量的三角不等式.

需要注意的是, 我们这里定义的算子范数是非常受限的一个定义: 我们只考虑了有限维内积空间, 由内积诱导的范数定义的算子范数. 一

般地, 任意两个线性赋范空间之间的线性映射都可以定义范数:

**定义 A.36 (线性映射的范数)** 设  $V, W$  是两个线性赋范空间,  $f$  是  $V$  到  $W$  的一个线性映射, 如果

$$\|f\| = \sup_{\|x\|=1} \|f(x)\|$$

不是无穷大, 那么称  $f$  是有界的,  $\|f\|$  是  $f$  的算子范数. □

如果线性映射的定义域是有限维空间, 那么范数一定存在. 无限维内积空间中算子范数不一定存在, 所以我们接下来的讨论都默认有限维线性空间.

注意到, 算子范数自然地诱导了矩阵的范数:

**定义 A.37 (矩阵范数)** 设  $A$  是一个  $m \times n$  的矩阵, 那么  $A$  的范数定义为

$$\|A\| = \sup_{\|x\|=1} \|Ax\|. \quad \square$$

给定标准正交基底, 利用矩阵  $A$  表示线性映射  $f$ . 假设此时  $x$  的坐标是  $X$ , 那么  $\|f(x)\| = \|AX\|$ , 由例 B.6、命题 B.10 和推论 B.2,  $\|AX\|$  是  $X$  的连续函数, 因此在紧集  $\{X: \|X\| = 1\}$  上取到最大值. 因此定义中的  $\sup$  实际上是一个  $\max$ . 这一结论对任意范数定义的算子范数都成立 (而不仅仅只是内积诱导的范数). 我们后面会利用内积的特性显式给出取到最大值的向量.

算子范数一个显然的性质是:

**命题 A.13** 对有限维内积空间  $V$  中的线性算子  $f$ ,  $\|f(x)\| \leq \|f\| \|x\|$  对任意  $x \in V$  成立.

利用这一条, 我们马上得到

**命题 A.14** 对有限维内积空间  $V$  中的线性算子  $f, g$ ,  $\|f \circ g\| \leq \|f\| \|g\|$ .

**命题 A.15** 对有限维内积空间  $V$  中的线性算子  $f$ ,  $\|f^k\| \leq \|f\|^k$ , 这里  $f^k$  表示  $f$  的  $k$  次复合.

以上性质都可以迁移到一般的线性映射以及矩阵上, 这里不再赘述.

上面的性质并不依赖于“内积”的性质, 主要是依赖“范数”的性质. 接下来, 我们将深入利用内积的性质来研究算子的性质. 这里面的关键概念是谱, 或者特征值.

**定义 A.38 (特征值, 特征向量, 谱)** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个线性算子,  $\lambda \in \mathbb{R}$ . 如果存在一个非零向量  $x \in V$  使得  $f(x) = \lambda x$ , 那么称  $\lambda$  是  $f$  的一个特征值,  $x$  是  $\lambda$  对应的特征向量.  $f$  的所有特征值的集合称为  $f$  的谱, 记作  $\sigma(f)$ .  $\square$

我们这里限制特征值为实数. 实际上, 一般的情况下, 特征值应该为复数, 而内积应该定义为复共轭定义的内积. 但是, 我们不关心会出现复数的情况, 因此这里只考虑实数的情况.

我们下面的任务是给出特征值的刻画. 首先定义特征子空间:

**定义 A.39 (特征子空间)** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个线性算子,  $\lambda \in \mathbb{R}$ . 定义

$$V^\lambda = \{x \in V : f(x) = \lambda x\}.$$

称  $V^\lambda$  是  $f$  的特征子空间.  $\square$

显然,  $V^\lambda$  是  $V$  的一个线性子空间. 我们下面的任务是刻画特征子空间.

特征向量存在的意味着  $V^\lambda$  非零, 也就是  $\ker(f - \lambda \cdot \text{id}) \neq \{0\}$ . 因此, 根据推论 A.1,  $f - \lambda \cdot \text{id}$  不是满射, 因此也不是双射, 从而  $\det(f - \lambda \cdot \text{id}) = 0$ . 当选择一组基之后,  $\det(f - \lambda \cdot \text{id})$  就可以写成  $\det(A - \lambda I_n)$  的形式, 其中  $A$  是  $f$  在这组基下的矩阵. 因此, 我们得到了一个关于  $\lambda$  的方程:

$$\det(A - \lambda I_n) = 0. \quad (\text{A.2})$$

将  $\lambda$  展开, 我们得到一个关于  $\lambda$  的  $n$  次多项式, 称为  $f$  的特征多项式. 这个多项式的根就是  $f$  的特征值.

我们需要验证特征多项式对于算子来说是良定义的, 也就是不管怎么选取基, 得到的特征多项式都是一样的.

**命题 A.16** 设  $A$  是一个  $n$  阶方阵,  $P$  是一个可逆方阵, 那么  $A$  与  $P^{-1}AP$  有相同的特征多项式.

**证明.**  $\det(P^{-1}AP - \lambda I_n) = \det(P^{-1}(A - \lambda I_n)P) = \det(P^{-1}) \det(A - \lambda I_n) \det(P) = \det(A - \lambda I_n)$ .  $\square$

有了特征值, 其对应的特征向量完全由  $V^\lambda$  刻画. 在特定的基之下, 我们可以用线性方程组的方式求出这个子空间一组基的坐标.

至此, 我们有了求解特征值和特征向量的方法.

**注.** 根据上面的讨论,  $n$  维内积空间的谱至多有  $n$  个元素, 因而是离散的. 这一点在无穷维内积空间中也不成立, 因而无穷维算子的谱要复杂得多. 无穷维谱理论在泛函分析、量子力学等领域有着广泛的应用. 然而遗憾的是, 这些都超出了本书的范围.

利用定义, 很容易写出矩阵特征值的性质:

**命题 A.17** 设  $A$  是一个  $n$  阶方阵,  $\lambda$  是  $A$  的一个特征值,  $v$  是  $\lambda$  对应的特征向量, 那么

- 如果  $A$  可逆, 那么  $\lambda^{-1}$  是  $A^{-1}$  的一个特征值,  $v$  是  $\lambda^{-1}$  对应的特征向量;
- 任意多项式  $p$ ,  $p(\lambda)$  是  $p(A)$  的一个特征值,  $v$  是  $p(A)$  的一个特征向量.

接下来, 我们考虑一类特殊的线性算子, 被称为自伴算子.

**定义 A.40 (自伴算子)** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个线性算子. 如果对任意  $x, y \in V$ , 都有  $\langle f(x), y \rangle = \langle x, f(y) \rangle$ , 那么称  $f$  是自伴算子.  $\square$

我们指出, 自伴算子的矩阵在标准正交基下是实对称矩阵, 这可以从内积的矩阵表示看出. 假设  $A$  是  $f$  在标准正交基下的矩阵, 对任意  $x, y \in V$ , 他们的坐标是  $X, Y$ , 那么  $\langle f(x), y \rangle = \langle x, f(y) \rangle$  等价于  $(AX)^T Y = X^T (AY)$ , 也就是  $X^T A^T Y = X^T AY$ . 将  $X, Y$  取遍所有可能的向量, 我们得到  $A^T = A$ .

自伴算子的谱可以有一个非常好的刻画:

**定理 A.17** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个自伴算子,  $\lambda \in \mathbb{R}$ . 那么  $f$  的特征多项式所有根都是实根, 因此  $\sigma(f)$  是一个非空实数集. 此外,

$$V = \bigoplus_{\lambda \in \sigma(f)} V^\lambda.$$

因此, 存在一组标准正交基, 使得  $f$  在这组基下的矩阵是对角矩阵, 这组基就是  $f$  的特征向量. 此外, 对角线上的元素恰好是  $f$  的特征值, 对于  $\lambda \in \sigma(f)$ , 它在对角线上出现的次数就是  $\dim V^\lambda$ .

这一证明非常类似定理 A.11 和定理 A.7 的证明, 我们这里就不再赘述.

这一定理在矩阵上表述为:

**推论 A.5** 对任意实对称矩阵  $A$ , 都存在一个正交矩阵  $P$  使得  $P^{-1}AP = P^TAP$  是对角矩阵, 这个对角矩阵的对角线上的元素就是  $A$  的特征值.

因此实对称矩阵可以通过一个正交矩阵相似并合同到对角矩阵.

这一结果在几何上有明确的意义. 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个自伴算子,  $v_1, \dots, v_n$  是  $f$  对应的标准正交的特征向量基, 那么  $f$  的作用就是将这些坐标轴拉伸或者压缩, 或者反转. 例如, 在  $\mathbb{R}^2$  上, 假设  $f(v_1) = -1.5v_1$ ,  $f(v_2) = 0.5v_2$ , 这一算子的效果可以用图 A.1 表示出来.

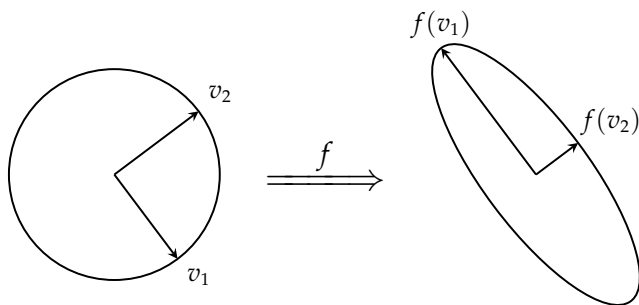


图 A.1: 自伴算子的作用

接下来, 我们考虑自伴算子的范数与谱的关系. 设  $f$  是  $V$  上的一个自伴算子, 那么根据定理 A.17, 存在一组标准正交基  $e_1, \dots, e_n$ , 使得

$f(e_i) = \lambda_i e_i$ ,  $\lambda_i$  是特征值 (可重复). 此时  $f$  对应的矩阵记为  $A$ , 它是一个对角矩阵. 考虑任何一个向量  $x \in V$ , 它的坐标是  $X$ , 那么

$$\|f(x)\|^2 = \|AX\|^2 = \sum_{i=1}^n \lambda_i^2 x_i^2.$$

假设  $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ , 那么

$$|\lambda_n|^2 \sum_{i=1}^n x_i^2 \leq \sum_{i=1}^n \lambda_i^2 x_i^2 \leq |\lambda_1|^2 \sum_{i=1}^n x_i^2.$$

因此

$$|\lambda_n| \|x\| \leq \|f(x)\| \leq |\lambda_1| \|x\|.$$

从左到右, 等号分别在  $x = e_n$  和  $x = e_1$  时取到. 因此, 我们得到了

**定理 A.18** 设  $V$  是一个  $n$  维内积空间,  $f$  是  $V$  上的一个自伴算子, 那么

$$\|f\| = \max_{\lambda \in \sigma(f)} |\lambda|.$$

如果最大值在特征值  $\lambda_0$  取到, 其对应单位特征向量是  $e_0$ , 那么  $\|f(e_0)\| = \|f\| = |\lambda_0|$ .

最大的特征值模称为算子的谱半径. 从几何意义来说, 谱半径就是算子对应的线性变换对应的线性变换对向量的最大拉伸率. 一个更直观的理解方式是, 将这些基向量画一个球包起来, 算子会将这个球映射到一个椭球, 这个椭球的最长轴就是谱半径. 例如, 对于图 A.1 中的算子, 谱半径就是 1.5.



# 参考文献

- [Bre57] Leo Breiman. The Individual Ergodic Theorem of Information Theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 1957.
- [CT12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.
- [Huf52] David A. Huffman. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*, 40(9):1098–1101, September 1952.
- [Inf] Information | Etymology, origin and meaning of information by etymonline. <https://www.etymonline.com/word/information>. (accessed 2023-07-10).
- [Jay02] Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2002.

- [KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [LLG<sup>+</sup>19] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, October 2019.
- [McM53] Brockway McMillan. The Basic Theorems of Information Theory. *The Annals of Mathematical Statistics*, 24(2):196–219, June 1953.
- [RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, pages 318–362. MIT Press, Cambridge, MA, USA, January 1986.
- [Rob49] Robert M. Fano. *The Transmission of Information*. March 1949.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1996.

- [Uff22] Jos Uffink. Boltzmann's Work in Statistical Physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2022 edition, 2022.
- [李 10] 李贤平. 概率论基础. 高等教育出版社, 2010.