

“错位”  
的

# 线性 代数

(美) S.Treil 著

董耀择 译



原书 Linear Algebra Done Wrong

中文精译版

ISBN 114-514191981-0



1 145141 919810

---

# 「错位」的线性代数

## ——Linear Algebra Done Wrong 翻译版

---

作者：[美]Sergei Treil（谢尔盖·特雷尔）

布朗大学数学系

译者：董耀择

南京大学匡亚明学院 2025 级本科生

发布日期：2025 年 12 月 24 日

当前版本：9.8.6

NaN

出版社



To my parents.

谨以此书献给我的父母。

——来自译者

Copyright © Sergei Treil, 2004, 2009, 2011, 2014, 2017, 2020, 2021, 2024



本书基于完整署名-禁止商用-禁止演绎 3.0 许可协议，参见

<https://creativecommons.org/licenses/by-nc-nd/3.0/>

**额外详情：**您可以免费使用本书用于非商业目的，特别是用于学习和（或）教学。您可以使用个人打印机或专业打印服务打印本书的纸质副本或其部分内容。教师在开设课程时（或其所属机构）可以为学生提供本书的印刷副本，并收取印刷成本费，但学生应有使用免费电子版本的选项。

**译者补充：**该译本仅用于个人学习交流，且完全开源、免费发布。本书版权归原作者所有，请勿用作其他非法、商业用途。

# 推荐序

一部刻意“错位”的线性代数教材，恰好对了现代学习的“位”。

如果你已经翻到这里，多半对这本书的名字产生过疑惑——为什么要叫《“错位”的线性代数》？线性代数这样一门讲究严谨与规范的课程，居然以“Wrong”来命名？

真正的“错”不在数学，而在路线。

多数线性代数教材遵循着一条自洽却有些“传统”的道路：先从线性方程组和初等行变换开始，再谈矩阵、行列式、特征值与特征向量，最后才把抽象的向量空间和线性变换请上场。而在中国，绝大多数教材更是沿袭了苏式教材的讲法：先进行行列式，然后是向量，矩阵，线性方程组……

这些写法对第一次接触线代的学生当然友好，却令人摸不着头脑，在一开始引入一些莫名其妙的概念不知道是为了干什么；这也容易让人形成一种印象：线性代数就是一套“行列式、矩阵计算技术”和“考试题型模板”。

这本书，刻意与那条熟悉的道路“错位”。

作者 Sergei Treil 把向量空间、基、线性变换放在了非常靠前的位置；他坚持先把“线性代数在想什么”讲清楚，然后才教“线性代数怎么算”。他宁愿花时间讨论：为什么要用基、为什么矩阵乘法只能那样定义、为什么行列式本质上是“有符号体积”，也不愿把篇幅都交给“教你十种快速算行列式的方法”。在很多传统教材的章节安排中，这样的路线确实有些“错位”——但也正因为这份“错位感”，你会更早、更直接地接触线性代数真正的灵魂。

## 这本书到底在讲什么？

如果用一句话来概括：它试图用一门线性代数课，完成学生从“会算题”到“会理解抽象数学”的跨越。

全书的结构，与其说是“从易到难”，不如说是“从直观到本质”：

- 在前几章中，你会看到我们熟悉的对象——向量、矩阵、线性方程组——但它们被放在统一的“线性变换”视角下讨论。基的选择、坐标变换、矩阵乘法的定义，都被解释为某种“不得不如此”的自然选择，而不是一个人类任意约定的规则。

- 行列式那一章，作者几乎“拒绝”从公式和展开式讲起，而是从体积、定向和多线性出发，一步步推导出经典行列式的性质。你会看到：那些在习题课里被当作“记忆负担”的公式，其实都隐含着几何和代数的深层结构。
- 从第四章开始，谱理论、特征值与特征向量登场，实数空间自然而然扩展到复数空间。书中不只关心“怎么求特征值”，更关心“特征分解究竟在数学与应用中扮演怎样的角色”。
- 接着是内积空间、正交投影、正规算子、自伴算子、极分解与奇异值分解(SVD)等主题——这些在多数初级教材里只是“略提一二”的内容，在本书中都获得了扎实的展开。它们直接连接了数据分析、数值线性代数和现代应用数学的核心方法。
- 在更靠后的章节中，作者引入了对偶空间和张量，并在有限维框架内勾勒出张量思想的轮廓。对于尚未接触高等代数和张量分析的读者，这一章既是挑战，也是通往更高层次数学的一块“垫脚石”。
- 若尔当标准型只在第九章亮相，且被作者明确地标记为“可选内容”。这不是数学上的轻视，而是一种取舍：对大多数走向分析、概率、几何和应用数学的学生而言，理解正交对角化、SVD、二次型与正定性，比写出一个复杂矩阵的若尔当分解更重要。

如果你已经学过一轮“常规”线性代数，再来读这本书，你会体验到一种熟悉结构被重新拆解、再组装的快感——许多曾经记不住的公式与结论，会因为背后的“为什么”而突然变得自然起来。

### 它和普通线性代数教材，有什么不一样？

- 它把“基”放在舞台中央。在本书中，线性无关和生成并不是孤立的定义，而是“基”这个核心概念的两个侧面：存在性与唯一性。作者一开始就强调：我们真正关心的是“能不能找一组适合做坐标的向量”，因为一旦有了基，我们就可以用  $\mathbb{R}^n$  或  $\mathbb{C}^n$  来取代一切抽象空间。
- 它用线性变换统摄方程组和矩阵运算。行约简、高斯消元、主元计数等熟悉技巧，被统一解释为对线性算子的研究工具，而不只是“考试技巧”。证明“任意两个基有相同大小”、“秩的各种性质”等关键命题时，本书坦率地承认：这些结论的深处，其实都在使用高斯消元。
- 它从几何直观中抽出代数结构。行列式来自体积，正交变换来自旋转与反射，正定矩阵与二次型对应几何上的椭球与距离。许多定理在书中先以几何语言被“讲通”，然后才化为公式与证明。这使得抽象概念不再悬空，而是牢牢地钉在图像和直觉上。
- 它刻意连接后续课程和应用方向。作者本人的背景偏分析与应用，而非纯代



数，因此书中对谱理论、自伴算子、极分解、SVD、正定性判据等内容着墨颇多。这些正是现代数值分析、信号处理、机器学习、优化理论等领域的基础工具。

- 它并不追求“形式上的完全无坐标化”。很多“高级”的线性代数教材，刻意避免出现具体矩阵，把一切都写成抽象映射与同构。但这本书则坦率而务实：在多数地方直接把算子和矩阵视为同物，只在讨论基变换等问题时才刻意区分。这让初学者能把精力更多地放在概念本身，而不是符号系统的切换上。最重要的是，如果读者将来遇到其他领域内的线性算子，那么本书所讲的内容都对其适用。这种学科视角的“高度”是可贵的。

换句话说，这本书并不想再造一门“更难线性代数”，而是想通过一种略微“错位”的讲法，让你在第一次认真学习线性代数时，就习惯用“线性变换 - 空间结构 - 证明思维”的方式来理解问题，而不是停留在机械操作层面。

### 为什么值得读的是这一中文版？

你的手上，并不是一本“机翻修订本”，而是一位华五相关专业的在校学生于课后有限的时间中，逐行推敲、完全重排、完美复刻版的成果。

译者在“译者的话”中已经讲得很坦诚：

- 这是为解决真实学习痛点而发起的项目——原书在课程中极其重要，但全英文对很多同学构成门槛；
- 中文版不仅追求语义的准确传达，更追求排版维度上的“高仿原书”：所有公式、符号、例题结构尽可能与原版一一对应；
- 全书以  $\text{\LaTeX}$  精心排版，并完全开源于 [GitHub](#)，任何人都可以勘误、重排、增补习题解答——它是一部可以持续进化的教材。

从读者角度看，这意味着：

- 你可以在中文语境中精确地理解原作者的思路，不必在晦涩英语与新概念之间来回切换；
- 你可以顺畅地对照原文与译文，自学或预习时随时切换语言；
- 你甚至可以参与到这个项目中来，在 [GitHub](#) 上提交问题和修改建议，亲手推动一本高质量线性代数教材在中文世界的完善。

这部《“错位”的线性代数》，因此不只是一本文本意义上的书，更像是一门开放的课程、一场正在进行中的协作实践。

### 谁适合读这本书？

- 对数学有兴趣、愿意接受一点抽象推理训练的一年级或二年级本科生——尤其是理科、工科、计算机和经济金融方向的学生；

- 已经学过“标准线性代数”课程，但总觉得概念散乱、只会做题不会理解的同学——这本书可以帮你“重建内功”；
- 准备进一步学习实分析、抽象代数、泛函分析、微分几何、概率论等高年级课程的学生——它会让你更早习惯现代数学的语言与视角；
- 任何对数学教材的写法、结构与美感感兴趣的读者——这本书本身就是一本“如何讲清一门抽象学科”的范例。

### **如果你决定读下去，可以期待什么？**

你会发现：

- 线性代数远不止“解线性方程组”和“算行列式”，而是一种在整个现代数学和应用科学中反复出现的结构语言；
- 许多曾经以为“只能硬背”的东西，实际上背后都有几何、代数或分析上的必然性；
- 严谨的推理并不意味着枯燥，相反，它可以像拆解一台精巧机器那样令人愉快；
- 你不只是“学会了线性代数”，而是被推了一把，跨进了现代数学的门槛。

如果你只是想多拿几分考试分数，这本书也许不是最高“性价比”的选择；但如果你希望真正理解自己在学什么，并愿意在大学阶段为自己的数学根基多投入一点时间，那么这本《“错位”的线性代数》，值得你从头到尾走一遍。

推荐者 GPT-5.1

# 译者的话

## 我为什么翻译这本书？

在南京大学匡亚明学院，时间是每位学生最宝贵的资源。我们的学习强度很高，课业繁重，几乎没有空闲。在这种情况下，再额外开启一个耗时巨大的翻译项目，似乎是一个不理智的选择。

但我还是决定这么做。

因为《Linear Algebra Done Wrong》是我们线代课程至关重要的参考书，而语言的障碍确实困扰着不少同学。我相信，一份高质量的中文译本能够实实在在地帮助到大家。

南京大学的校训精神中，“诚”字所蕴含的力量给了我巨大的鼓舞。“大哉一诚天下动”，它告诉我，一个真诚的、以服务之心出发的行动，即便微小，也能产生积极的影响。与其独自感叹学习之艰辛，不如动手为大家做一点实事。

因此，我启动了这个项目。我希望这份译稿不仅仅是我个人学习的沉淀，更能成为一份小小的礼物，送给每一位在知识海洋中奋力前行的同学。希望它能为你节省一些时间，扫清一些迷茫，让你更专注于领略线性代数的核心思想。

本书译自 2025 年 8 月 25 日作者发布在[个人主页](#)上的版本，中文版全文采用 L<sup>A</sup>T<sub>E</sub>X 精心排版，力求高度还原原书中的所有公式与数学符号，以保证其准确性和专业性。本书的排版模板来源于[zhbook](#)项目，在此对原作者的贡献表示由衷的感谢。

本书的 GitHub 开源项目地址为：<http://github.com/DongYaoZe/Translate-LADW> 诚挚欢迎各位同学加入翻译、贡献代码、参与纠错，共同完善此项目。若上述链接无法访问，也可以通过点击[此处](#)访问。每次版本更新亦将会在该页面同步发布。

若您对本翻译项目有任何疑问、建议或想法，欢迎通过我的电子邮箱 [251840159@smail.nju.edu.cn](mailto:251840159@smail.nju.edu.cn) 与我联系。

董耀择

2025 年 10 月





# 前言

本书的标题听起来有些神秘。为什么有人会读这本书，如果它以一种错位的方式呈现了这个主题呢？这本书里到底哪里是“错”的呢？在回答这些问题之前，请允许我先描述一下本书的目标读者。

本书源于“荣誉线性代数”（Honors Linear Algebra）课程的讲义。它旨在成为一门面向已在数学上训练有素的学生们的入门线性代数课程。它适用于那些虽然还不太熟悉抽象推理，但愿意学习比“菜谱式”（cookbook style）的微积分课程更严谨数学的学生。除了作为线性代数的第一门课程，它还旨在成为介绍严谨证明、形式化定义——简而言之，介绍现代理论（抽象）数学风格的第一门课程。本书的目标读者解释了它为何如此特别地融合了初级概念和具体示例（通常在入门线性代数教材中呈现），以及更抽象的定义和构造（通常在高级书籍中典型出现）。本书的另一个特点是它不是由代数专家所写，也不是为代数专家所写。因此，我试图强调那些对分析、几何、概率等重要主题，而没有包含一些传统的主题。例如，我只考虑实数或复数域上的向量空间。完全不考虑其他域上的线性空间，因为我认为花费时间介绍和解释抽象域的知识，不如花在一些在其他学科中更必需的经典主题上。

后来，当学生在抽象代数课程中学习一般域时，他们会明白本书研究的许多构造也适用于一般域。

本书仅考虑有限维空间，并且基总是指有限基。原因是，要对无限维空间说出一些有意义的话，就必须引入收敛、范数、完备性等概念，即泛函分析的基础。而这绝对是一个单独的课程（教材）的主题。

因此，我在此不考虑无限维 Hamel 基：它们在大多数分析和几何应用中是不需要的，而且我认为它们属于抽象代数课程。

## 给教师的说明

本书的某些细节使其与标准的进阶线性代数教材有所不同。

首先是关于基、线性无关和生成集的定义。在书中，我首先将基定义为这样的系统：任何向量都能唯一地表示为其线性组合。然后，线性无关和生成系统的性质自然地成为基的性质的组成部分，一个代表表示的唯一性，另一个代表表示的存在性。这样处理的原因是，我认为基的概念比线性无关的概念更重要：在大多数应用中，我们并不真正关心线性无关，我们需要的是一个可以作为基的系统。

例如，在求解齐次方程组时，我们不仅仅是在寻找线性无关的解，而是在寻找解空间中的一组基。而且，向学生解释基的重要性很容易：它允许我们引入坐标，并使用  $\mathbb{R}^n$ （或  $\mathbb{C}^n$ ）来代替处理抽象向量空间。此外，我们需要坐标来进行计算机计算，而计算机非常擅长处理矩阵。而且，我真的不知道线性无关概念有什么简单的动机。

另一个细节是，我先介绍线性变换，然后再教授如何求解线性方程组。

一个缺点是我们直到第二章才证明只有方阵是可逆的，以及其他一些重要事实。然而，已经定义的线性变换允许更系统的行约简的呈现。此外，我花了很多时间（两节）来阐述矩阵乘法的动机。我希望我能很好地解释为什么这种看起来很奇怪的乘法规则实际上非常自然，以至于我们别无选择。

许多关于基、线性变换等重要事实，例如在任何向量空间中，两个基具有相同数量的向量，都是通过行约简中的主元计数来证明的。虽然这些事实中的大多数都具有“无坐标”的证明，形式上不涉及高斯消元，但仔细分析这些证明会发现，高斯消元和主元计数并没有消失，它们只是在大多数证明中被隐藏起来了。因此，与其呈现非常优雅（但对于初学者来说很难理解）的“无坐标”证明，这些证明通常在进阶线性代数书籍中呈现，我们则使用了“行约简”证明，这在“微积分”教材中更为常见。这样做的优点是，可以轻松地看到所有证明背后的共同思想，并且对于不那么数学化的读者来说，这些证明更容易理解和记住。

我还将在第二章的第 8 节中介绍一个简单且易于记忆的公式，用于计算基变换。

第三章处理行列式。我花了很多时间来阐述行列式的动机，然后才给出正式定义。行列式被介绍为计算体积的一种方式。书中表明，如果我们允许了有符号体积，使得行列式在每列上都是线性的（此时学生应该很清楚线性关系非常有帮助，而允许负体积是为此付出的很小的代价），并且假设一些非常自然的性质，那么我们就别无选择，只能得到行列式的经典定义。我想强调的是，我一开始并没有假定行列式的反对称性，而是将其从体积的一些非常自然的性质中推导出来。

请注意，虽然在形式上，第一至第三章主要处理实数空间，但其中所有内容都适用于复数空间，甚至适用于任意域上的空间。

第四章是谱理论的介绍，这是复数空间  $\mathbb{C}^n$  自然出现的地方。虽然  $\mathbb{C}^n$  在本书开头被形式化定义，我们也给出了复数向量空间的定义，但在第四章之前，主要对象是实数空间  $\mathbb{R}^n$ 。现在，复数特征值的出现表明，对于谱理论，最自然的不是实数空间  $\mathbb{R}^n$ ，而是复数空间  $\mathbb{C}^n$ ，即使我们最初处理的是实数矩阵（实数空间中的算子）。这里的重点是特征值分解，并且特征值空间的基的概念也被引入。

第五章是内积空间，它出现在谱理论之后，因为我希望同时处理复数和实数两种情况，而谱理论为复数空间的出场提供了强有力的动机。除了引入的动机之外，第四章和第五章不互相依赖，教师可以先讲第五章。

尽管我在第九章中介绍了若尔当标准型，但我通常没有时间在一学期的课程中讲授它。我更倾向于花更多的时间在第六章和第七章讨论的主题上，例如正规算子和自伴算子的对角化、极分解和奇异值分解、正交矩阵的结构和定向，以及二次型理论。

我认为这些主题比若尔当标准型对于应用更重要，尽管后者确实很优美。但是，我新增了第九章，以便教师可以跳过第六章和第七章中的某些主题，转而讲授若尔当分解定理。

我还新增了（2009 年新增）第八章，讨论对偶空间和张量。我认为其中的材料，特别是关于张量的部分，对于一年级的线性代数课程来说还是太难了，但有些主题（例如，对偶空间中的坐标变换）可以很容易地包含在教学大纲中。它还可以作为进阶课程中张量理论的介绍。请注意，本章中介绍的结果适用于任意域。

我试图以较为非正式的方式呈现本书的材料，偏爱直观的几何推理而不是形式化的代数运算，因此对于一些纯粹主义者来说，本书可能不够严谨。在本书通篇，我通常（当不引起混淆时）将线性变换与其矩阵等同起来，让我们能使用更简单的符号。而且我认为，对于没有经验的学生来说，过度强调变换与矩阵之间的区别可能会造成混淆。只有当区别很重要时，例如在分析一个变换的矩阵如何在基变换下变化时，我才会使用特殊的符号来区分变换和它的矩阵。

Sergei Treil





# 目 录

推荐序	i
译者的话	v
前 言	vii
目 录	xi
主要符号表	xxi
第一章 基本概念	1
1. 向量空间	1
1.1. 一些例子	2
1.2. 矩阵表示	3
2. 线性组合, 基	5
2.1. 生成系统与线性无系统	6
3. 线性变换 矩阵-向量乘法	10
3.1. 一些例子	10
3.2. 线性变换 $\mathbb{F}^n \rightarrow \mathbb{F}^m$ 矩阵-向量乘法	12
3.3. 线性变换与生成集	13
3.4. 结论	14
4. 线性变换作为向量空间	15
5. 线性变换的复合与矩阵乘法	17
5.1. 矩阵乘法的定义	17
5.2. 动机: 线性变换的复合	17
5.3. 矩阵乘法的性质	19
5.4. 转置矩阵与乘法	19
5.5. 迹与矩阵乘法	20
6. 可逆变换与矩阵 同构	21
6.1. 恒等变换与单位矩阵	21

6.2. 可逆变换	21
6.2.1. 逆变换的性质	23
6.3. 同构 同构空间	24
6.4. 可逆性与方程	25
7. 子空间	27
8. 应用于计算机图形学	28
8.1. 二维操作	28
8.2. 三维图形	29
<b>第二章 线性方程组</b>	<b>33</b>
1. 线性方程组的不同表示	33
2. 线性方程组的求解 阶梯形与简化阶梯形	34
2.1. 行运算	34
2.1.1. 行运算与初等矩阵的乘法	34
2.2. 行约简	36
2.2.1. 行约简的一个例子	36
2.3. 阶梯形	38
3. 主元的分析	40
3.1. 关于线性无关和基的推论 维数	41
3.2. 可逆矩阵的推论	43
4. 通过行约简求 $A^{-1}$	44
5. 维数 有限维空间	46
5.1. 将线性无关系统补全为基	47
5.2. 有限维空间的子空间	47
6. 线性系统的通解	48
7. 矩阵的基本子空间 秩	51
7.1. 计算基本子空间和秩	52
7.2. 基本子空间基的计算方法的解释	54
7.2.1. 零空间 $\text{Ker } A$	54
7.2.2. 列空间 $\text{Ran } A$	54
7.2.3. 行空间 $\text{Ran } A^T$	55

7.3. 秩定理 基本子空间的维数	56
7.4. 将线性无关系统补全为基	58
8. 任意基下线性变换的表示 坐标变换公式	60
8.1. 坐标向量	60
8.2. 线性变换的矩阵	60
8.3. 坐标变换矩阵	61
8.3.1. 从标准基进行坐标变换的例子	61
8.3.2. 通过标准基进行变换的例子	62
8.4. 变换的矩阵与坐标变换	63
8.5. 单组基的情况: 相似矩阵	63
<b>第三章 行列式</b>	<b>65</b>
1. 引言	65
2. 行列式应具备的性质	66
2.1. 关于列向量的线性性质	66
2.2. “列替换”下保持不变性	67
2.3. 反对称性	67
2.4. 归一化	68
3. 行列式的构造	68
3.1. 基本性质	68
3.2. 从行列式基本性质中推导出的性质	69
3.3. 对角和三角矩阵的行列式	70
3.4. 计算行列式	71
3.5. 矩阵转置和乘积的行列式 初等矩阵的行列式	71
3.6. 行列式的性质总结	73
4. 行列式的正式定义 存在性与唯一性	75
5. 代数余子式展开	77
5.1. 逆矩阵的代数余子式公式	80
5.2. 逆矩阵的代数余子式公式的应用	81
6. 子式与秩	82
7. 第三章复习题	83

<b>第四章 谱理论简介 (特征值与特征向量)</b>	<b>85</b>
1. 基本定义	85
1.1. 特征值、特征向量、谱	85
1.2. 寻找特征值: 特征多项式	86
1.3. 寻找抽象算子的特征多项式和特征值	86
1.4. 复空间与实空间	87
1.5. 特征值的重数	88
1.6. 迹与行列式	88
1.7. 三角矩阵的特征值	88
2. 对角化	90
2.1. 预备知识	90
2.2. 一些动机: 算子函数	92
2.3. $n$ 个不同特征值的情况	93
2.4. 子空间的基 (又名子空间的直和)	93
2.5. 可对角化判据	95
2.6. 实数分解	96
2.7. 一些例子	96
2.7.1. 实数特征值	96
2.7.2. 复数特征值	97
2.7.3. 一个不可对角化的矩阵	97
<b>第五章 内积空间</b>	<b>101</b>
1. $\mathbb{R}^n$ 与 $\mathbb{C}^n$ 中的内积 内积空间	101
1.1. $\mathbb{R}^n$ 中的内积和范数	101
1.2. $\mathbb{C}^n$ 中的内积和范数	102
1.3. 内积空间	103
1.3.1. 例子	103
1.4. 内积的性质	104
1.5. 范数 赋范空间	106
2. 正交性 正交与标准正交基	108
2.1. 正交和标准正交基	110



3. 正交投影和格拉姆-施密特正交化	112
3.1. 格拉姆-施密特正交化方法	114
3.2. 一个例子	115
3.3. 正交补 正交分解 $V = E \oplus E^\perp$	116
4. 最小二乘解 正交投影的公式	117
4.1. 最小二乘解	117
4.1.1. 几何方法	118
4.1.2. 正规方程	118
4.2. 正交投影公式	119
4.3. 一个例子: 直线拟合	119
4.3.1. 一个例子	120
4.4. 其他例子: 曲线和平面拟合	121
4.4.1. 曲线拟合例子	121
4.4.2. 平面拟合	122
5. 线性变换的伴随, 基本子空间的再次回顾	123
5.1. 伴随矩阵与伴随算子	123
5.1.1. 伴随的唯一性	124
5.1.2. 抽象环境下的伴随变换	124
5.1.3. 有用的公式	124
5.2. 基本子空间之间的关系	125
5.3. 线性变换的“本质”部分	126
6. 等距同构和酉算子 酉矩阵和正交矩阵	127
6.1. 基本定义	127
6.2. 例子	129
6.3. 酉算子的性质	129
6.4. 酉等价算子	130
7. $\mathbb{R}^n$ 中的刚性运动	132
8. 复化与反复化	134
8.1. 反复化	134
8.1.1. 向量空间的复化	134
8.1.2. 内积的复化	134

8.2. 复化	135
8.3. 向实数空间引入复数结构	136
8.3.1. 引入复数结构的初等方法	136
8.3.2. 从初等到抽象构造复数结构	137
8.3.3. 复结构的抽象构造	137
8.3.4. 通过初等方法进行的抽象构造	139
<b>第六章 内积空间中算子的结构</b>	<b>141</b>
1. 算子的上三角 (舒尔) 表示	141
2. 自伴随和正规算子的谱定理	143
3. 极分解与奇异值分解	148
3.1. 正定算子 平方根	148
3.2. 算子的模 奇异值	149
3.3. 奇异值 施密特分解	150
3.4. 施密特分解的矩阵表示 奇异值分解	152
3.4.1. 从奇异值分解到极分解	153
4. 奇异值分解的应用	155
4.1. 单位球的像	155
4.2. 线性变换的算子范数	157
4.3. 矩阵的条件数	158
4.4. 矩阵的有效秩	159
4.5. 摩尔-彭罗斯 (伪) 逆	160
5. 正交矩阵的结构	162
6. 定向	167
6.1. 动机	167
6.2. 形式定义	168
6.3. 基的连续变换与定向	169
<b>第七章 双线性型与二次型</b>	<b>171</b>
1. 主要定义	171
1.1. $\mathbb{R}^n$ 上的双线性型	171
1.2. $\mathbb{R}^n$ 上的二次型	171

1.3. $\mathbb{C}^n$ 上的二次型	172
2. 二次型的对角化	173
2.1. 正交 (相似) 对角化	174
2.2. 非正交对角化	175
2.2.1. 通过配方法对角化	175
2.2.2. 使用行/列运算进行对角化	176
3. 惯性定理	179
4. 正定二次型 极小极大特征值刻画与塞尔维斯特正定性判据	181
4.1. 塞尔维斯特正定性判据	181
4.2. 特征值的极小极大刻画	183
4.3. 一些注记	185
5. 正定二次型与内积	186
<b>第八章 对偶空间与张量</b>	<b>189</b>
1. 对偶空间	189
1.1. 线性泛函与对偶空间 对偶空间中的坐标变换	189
1.1.1. 坐标变换公式	190
1.1.2. 唯一性定理	191
1.2. 二次对偶空间	191
1.3. 对偶基, 又称双正交基	192
1.3.1. 抽象非正交傅里叶分解	193
1.4. 对偶系统的例子	193
1.4.1. 泰勒公式	194
1.4.2. 拉格朗日插值	194
2. 内积空间的对偶	196
2.1. 里斯表示定理	196
2.2. 内积空间自身是其对偶吗?	197
2.3. 双正交系统与标准正交基	198
3. 对偶 (伴随) 变换与转置, 基本子空间再回顾 (又一次)	198
3.1. 对偶 (伴随) 变换	199
3.1.1. $A: \mathbb{F}^n \rightarrow \mathbb{F}^m$ 情况下的对偶变换	199
3.1.2. 抽象设置下的对偶变换	200

3.1.3. 定义对偶变换的无坐标方法	200
3.2. 零化子与基本子空间之间的关系	201
4. 空间与其对偶之间的区别	203
4.1. $X$ 与 $X'$ 之间的同构	203
4.2. 例子: 速度 (微分算子), 微分形式作为向量, 线性泛函	204
4.2.1. 速度作为向量	204
4.2.2. 微分形式作为线性泛函 (余向量)	205
4.2.3. 微分算子作为向量	205
4.3. 实内积空间的情况	206
4.3.1. 爱因斯坦求和约定, 度量张量	206
4.3.2. 协变和逆变坐标 上/下指标的升降	207
4.4. 结论	209
5. 多线性函数 张量	209
5.1. 多线性函数	209
5.1.1. 多线性函数构成向量空间	210
5.1.2. $L(V_1, V_2, \dots, V_p; V)$ 的维数	210
5.2. 张量积	211
5.2.1. 将多线性函数提升到张量积上的线性变换	212
5.2.2. 张量积的对偶	213
5.3. 协变和逆变张量	213
5.3.1. 线性变换作为张量	214
5.3.2. 多线性变换作为张量	215
6. 张量的坐标变换公式	216
6.1. 张量的坐标表示	217
6.2. 爱因斯坦求和约定中的坐标变换公式	218
6.3. 张量的坐标变换公式	218
<b>第九章 高级谱理论</b>	<b>221</b>
1. 凯莱-哈密顿定理	221
2. 谱映射定理	224
2.1. 算子的多项式	224
2.2. 谱映射定理	224

3. 广义特征子空间 代数重数的几何意义 .....	226
3.1. 不变子空间 .....	226
3.2. 广义特征子空间 .....	227
3.3. 代数重数的几何意义 .....	231
3.4. 一个重要的应用 .....	231
4. 幂零算子的结构 .....	232
4.1. 广义特征向量循环 .....	233
4.2. 幂零算子的若尔当标准型 .....	234
4.3. 点图 若尔当标准型的唯一性 .....	236
4.4. 计算若尔当标准基 .....	237
5. 若尔当分解定理 .....	238
5.1. 关于计算若尔当标准基的注记 .....	238
<b>附录 课后习题解答 .....</b>	<b>241</b>
第一章习题解答 .....	241
第二章习题解答 .....	252
第三章习题解答 .....	264
第四章习题解答 .....	272
第五章习题解答 .....	279
第六章习题解答 .....	294
第七章习题解答 .....	303
第八章习题解答 .....	307
第九章习题解答 .....	312
<b>索 引 .....</b>	<b>315</b>
<b>译 后 记 .....</b>	<b>317</b>

*This page intentionally left blank.*

# 主要符号表

**说明：**本符号表为译者基于原书内容梳理而成，旨在为读者提供一个快速查阅的索引。原书中虽然大部分符号都能通过上下文语境自然理解，但通过列表盘点，可以更直观地呈现全书的数学语言体系。

表中“最早出现在”仅为大致位置参考（精确到章节，如需具体页码请移步“索引”），旨在帮助读者定位相关定义的引入背景。初学者完全不必逐条阅读或刻意背诵这些符号，在实际阅读中遇到又不明白时再回查即可，也可以直接跳过本节。

符号	描述	最早出现在（章. 节）
$x \in A$	元素 $x$ 属于集合 $A$	
$A \subset B$	集合 $A$ 包含于集合 $B$ （子集）	
$A \subsetneq B$	集合 $A$ 真包含于集合 $B$ ，真子集（ $A \neq B$ ）	
$\forall$	对于所有（全称量词）	
$\exists$	存在（存在量词）	
$:=$	定义符号。 $A := B$ 表示“将 $B$ 作为 $A$ 的定义”，自此以后 $A$ 与 $B$ 视为同一表达式。	
$\square$	证明结束（Quod Erat Demonstrandum, Q.E.D.）	
$\lim$	极限符号，如 $\lim_{n \rightarrow \infty} x_n$ 、 $\lim_{\varepsilon \rightarrow 0^+} A_\varepsilon$ 等	
$\prod$	连乘记号： $\prod_{k=1}^n a_k = a_1 a_2 \cdots a_n$	
$\mathbb{R}$	实数域	1.1
$\mathbb{C}$	复数域	1.1
$\mathbb{F}$	任意域，一般的标量域，通常为 $\mathbb{R}$ 或 $\mathbb{C}$	1.1
$\mathbb{R}^n, \mathbb{C}^n, \mathbb{F}^n$	$n$ 维实/复/一般向量空间	1.1



$\mathbf{v}, \mathbf{x}, \mathbf{y}$	粗体小写字母, 表示列向量, 一般写作 $\mathbf{v} = (v_1, \dots, v_n)^T$ . 本书中所有向量默认都是列向量	1.1
$\mathbf{0}$	零向量或零矩阵	1.1
$V, W, X, Y$	一般的 (有限维) 向量空间	1.1
$M_{m \times n}$	$m \times n$ 矩阵的集合	1.1
$A = (a_{j,k})_{m \times n}^n$	以 $a_{j,k}$ 为元素的矩阵	1.1
$A^T$	矩阵 $A$ 的转置	1.1
$\mathbf{e}_k$	标准基向量 (第 $k$ 个分量为 1, 其余为 0)	1.2
$T: V \rightarrow W$	从空间 $V$ 到 $W$ 的线性变换	1.3
$\mathbf{x} \mapsto \mathbf{y}$	元素间的映射关系 ( $\mathbf{x}$ 被映射为 $\mathbf{y}$ )	
$T_1 T_2$ 或 $T_1 \circ T_2$	线性变换 (算子) 的复合/乘积	1.5
$R_\gamma$	绕原点逆时针旋转 $\gamma$ 角的变换矩阵	1.5
$\text{trace}(A), \text{tr}(A)$	矩阵 $A$ 的迹 (对角元之和)	1.5
$I$ 或 $I_V$	单位矩阵, 或空间 $V$ 上的恒等算子	1.6
$A^{-1}$	矩阵或算子 $A$ 的逆	1.6
$V \cong W$	向量空间 $V$ 和 $W$ 同构	1.6
$\text{Ker } A, \text{Null } A$	算子 $A$ 的核 (Kernel) 或零空间	1.7
$\text{Ran } A$	算子 $A$ 的像空间	1.7
$\mathcal{L}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$	向量系统 $\mathbf{v}_1, \dots, \mathbf{v}_n$ 的线性张成, 有时也用 $\text{span}$ 替代 $\mathcal{L}$	1.7
$A\mathbf{x} = \mathbf{b}$	线性方程组的矩阵形式	2.1
$E$	初等矩阵	2.2
$A_e$	矩阵的阶梯形 (Echelon form)	2.2
$A_{re}$	简化阶梯形 (Reduced Echelon form)	2.2
$\dim V$	向量空间 $V$ 的维数	2.5
$\text{rank } A$	矩阵 $A$ 的秩	2.7
$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$	向量空间的一组基 $\mathcal{B}$	2.8
$[T]_{\mathcal{B}}$	线性变换 $T$ 在基 $\mathcal{B}$ 下的矩阵表示	2.8
$A \sim B$	矩阵 $A$ 与 $B$ 相似 ( $B = Q A Q^{-1}$ )	2.8

$\det A$	矩阵 $A$ 的行列式	3.1
$D(\mathbf{v}_1, \dots, \mathbf{v}_n)$	作为列向量函数的行列式	3.1
$\text{diag}\{a_1, \dots, a_n\}$	对角元素为 $a_1, \dots, a_n$ 的对角矩阵	3.3
$*$	矩阵中不关心的, 未具体指明的或任意的元素 (通配符)	3.3
$\text{Perm}(n)$	$n$ 个元素的所有排列的集合	3.4
$K(\sigma)$	排列 $\sigma$ 的逆序对数量	3.4
$\text{sign}(\sigma)$	排列 $\sigma$ 的符号 (1 或 $-1$ )	3.4
$A_{j,k}$	矩阵的余子式 (Cofactor), $A_{j,k} = (-1)^{j+k}C_{j,k}$	3.5
$C_{j,k}$	矩阵的代数余子式 (Minor)	3.5
$C^T$	伴随矩阵 (Classical Adjoint) / 代数余子式矩阵	3.5
$\lambda$	特征值	4.1
$\sigma(A)$	算子 $A$ 的谱 (所有特征值的集合)	4.1
$\det(A - \lambda I)$	$A$ 的特征多项式	4.1
$\bar{z}$	复数 $z$ 的共轭	5.1
$\text{Re}(z), \text{Im}(z)$	复数 $z$ 的实部与虚部	5.1
$(\mathbf{x}, \mathbf{y})$	向量 $\mathbf{x}$ 与 $\mathbf{y}$ 的内积	5.1
$A^*$	矩阵 $A$ 的共轭转置, 即 $A^* = \overline{A}^T$ , 埃尔米特伴随	5.1
$\ \mathbf{x}\ , \ \mathbf{x}\ _2$	向量的范数 (通常指欧几里得范数)	5.1
$\ \mathbf{x}\ _p$	向量的 $p$ -范数 ( $(\sum  x_k ^p)^{1/p}$ )	5.1
$\ \mathbf{x}\ _\infty$	向量的 $\infty$ -范数 (最大值范数)	5.1
$\mathbf{x} \perp \mathbf{y}$	向量 $\mathbf{x}$ 与 $\mathbf{y}$ 正交 (垂直)	5.2
$E^\perp$	子空间 $E$ 的正交补	5.3
$V = E \oplus E^\perp$	正交分解, 其中 $\oplus$ 为直和	5.3
$P_E \mathbf{v}$	向量 $\mathbf{v}$ 在子空间 $E$ 上的正交投影。 $P_E : X \rightarrow E$ 是满足 $\mathbf{v} - P_E \mathbf{v} \in E^\perp$ 的线性算子	5.3
$U$	酉 (Unitary) 矩阵, 保持内积的线性算子	5.6

$\sqrt{A}, A^{1/2}$	正定算子 $A$ 的平方根	6.3
$ A $	算子 $A$ 的模, 定义为 $\sqrt{A^*A}$	6.3
$\sigma_k$	奇异值	6.3
$\Sigma$	对角上为非负奇异值的对角矩阵	6.3
$\delta_{k,j}$	克罗内克 (Kronecker) 符号	6.3
$\ A\ $	算子 (矩阵) $A$ 的算子范数	6.4
$\ A\  \cdot \ A^{-1}\ $	矩阵 $A$ 的条件数 (也作 $\kappa(A)$ )	6.4
$\Delta \mathbf{b}$	向量 $\mathbf{b}$ 的微小扰动	6.4
$A^+$	摩尔-彭罗斯伪逆	6.4
$\mathcal{V}(t)$	一族随参数 $t$ 连续变化的基	6.6
$L(\mathbf{x}, \mathbf{y})$	双线性型	7.1
$Q[\mathbf{x}]$	由双线性型或算子生成的二次型	7.1
$\text{codim } E$	子空间 $E$ 的余维度, $\text{codim } E = \dim(E^\perp)$	7.4
$\deg p$	多项式 $p$ 的次数	8.1
$V'$	向量空间 $V$ 的对偶空间	8.1
$\langle \mathbf{x}, \mathbf{f} \rangle$	对偶配对 (线性泛函 $\mathbf{f}$ 作用于向量 $\mathbf{x}$ )	8.1
$A'$	线性变换 $A$ 的对偶 (转置) 变换	8.2
$\mathbf{v} \otimes \mathbf{w}$	线性泛函 $\mathbf{v}$ 和 $\mathbf{w}$ 的张量积	8.5
$V \otimes W$	向量空间 $V$ 和 $W$ 的张量积	8.5
$p(A)$	矩阵多项式	9.1
$E_\lambda$	特征值 $\lambda$ 对应的广义特征子空间	9.3
$N$	幂零算子 ( $\exists k \in \mathbb{N}_+$ s.t. $N^k = \mathbf{0}$ )	9.4
$\mathcal{C}$	广义特征向量循环 (Cycle)	9.4
$J, J_k(\lambda)$	若尔当块 (Jordan block)	9.4

**缩写****全称**

SVD	Singular Value Decomposition, 奇异值分解
iff	if and only if, 当且仅当

# 第一章 基本概念

## 1. 向量空间

向量空间  $V$  由一些被称为向量（本书中用小写粗体字母表示，如  $\mathbf{v}$ ）的对象组成，并且包含两个运算：向量加法和数（标）量乘法<sup>1</sup>，使得以下 8 个性质（所谓的向量空间公理 (axiom)）成立：

前 4 个性质涉及加法：<sup>2</sup>

1. 交换律： $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$  对所有  $\mathbf{v}, \mathbf{w} \in V$ ；
2. 结合律： $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  对所有  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ ；
3. 零向量：存在一个特殊的向量，记作  $\mathbf{0}$ ，使得  $\mathbf{v} + \mathbf{0} = \mathbf{v}$  对所有  $\mathbf{v} \in V$ ；
4. 加法逆元：对于每个向量  $\mathbf{v} \in V$ ，都存在一个向量  $\mathbf{w} \in V$  使得  $\mathbf{v} + \mathbf{w} = \mathbf{0}$ 。

这样的加法逆元通常记作  $-\mathbf{v}$ ；

接下来的两个性质涉及乘法：

5. 乘法单位元： $1\mathbf{v} = \mathbf{v}$  对所有  $\mathbf{v} \in V$ ；
6. 乘法结合律： $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$  对所有  $\mathbf{v} \in V$  和所有标量  $\alpha, \beta$ ；

最后是两个分配律，它们连接了乘法和加法：

7.  $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$  对所有  $\mathbf{u}, \mathbf{v} \in V$  和所有标量  $\alpha$ ；
8.  $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$  对所有  $\mathbf{v} \in V$  和所有标量  $\alpha, \beta$ 。

**注记** 上述性质似乎很难记忆，但读者没有必要死记硬背。它们只是我们熟知的从高中学到的关于对数字进行代数运算的规则。这里唯一的陌生之处在于，你必须理解你可以在什么对象上进行什么运算。你可以将向量相加，也可以用数字（标量）乘以向量。当然，你可以对数字进行所有你以前学过的运算。但是，你不能将两个向量相乘，也不能将一个数字加到一个向量上。

**注记** 可以很容易地证明零向量  $\mathbf{0}$  是唯一的，并且给定  $\mathbf{v} \in V$ ，其加法逆元  $-\mathbf{v}$  也是唯一的。

---

<sup>1</sup>为了在向量和其他对象之间做出视觉区分，本书使用**加粗的小写字母**来表示向量，而使用**普通小写字母**来表示数字（标量）。在一些（更高级的）书中，拉丁字母留给向量使用，而希腊字母被用作标量；在更高级的文本中，任何字母都可以用于任何目的，读者必须根据上下文理解每个符号的含义。我认为，尤其对于初学者来说，在不同对象之间作一定的视觉区分是有帮助的，所以加粗的小写字母将始终表示一个向量。而在黑板上，通常会使用箭头（如  $\vec{v}$ ）来标识一个向量。

<sup>2</sup>这时会引出一个问题：“我们该如何记住上述性质呢？”而答案是，根本不需要记住，请看下文！

通过使用向量空间的性质 5, 6 和 8, 也不难证明  $\mathbf{0} = 0\mathbf{v}$  对任何  $\mathbf{v} \in V$ , 并且  $-\mathbf{v} = (-1)\mathbf{v}$ . 注意, 要做到这一点, 仍然需要使用向量空间的其它性质来证明, 特别是性质 3 和 4。

如果标量是通常的实数, 我们称空间  $V$  为**实向量空间** (real vector space)。如果标量是复数, 即如果我们能用复数乘以向量, 我们称空间  $V$  为**复向量空间** (complex vector space)。

注意, 任何复向量空间也都是实向量空间 (因为如果我们能用复数乘以向量, 那么我们也能用实数乘以向量), 但反之则不然。

我们可能会考虑标量是任意域  $\mathbb{F}$  中的元素的情况。在这种情况下, 我们说  $V$  是域  $\mathbb{F}$  上的向量空间。虽然本书中的许多构造 (特别是第一至第三章中的所有内容) 适用于一般域, 但本书仅考虑实数和复数向量空间。

如果我们不指定标量集, 或者用字母  $\mathbb{F}$  来表示它, 那么结果对实数和复数空间都成立。如果我们想区分实数和复数情况, 我们会明确说明我们正在考虑哪种情况。

请注意, 在定义域  $\mathbb{F}$  上的向量空间定义中, 我们**要求**标量集是一个域, 因此我们始终可以进行除法 (无余数), 尽管不能进行整数除法。在这种情况下, 我们可以考虑有理数域上的向量空间, 但不能考虑整数环上的。

### 1.1. 一些例子

**例子** 空间  $\mathbb{R}^n$  由所有大小为  $n$  的列向量组成:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

其中各项均为实数。加法和乘法是逐项定义的, 即

$$\alpha \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{pmatrix}, \quad \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}$$

**例子** 空间  $\mathbb{C}^n$  也由大小为  $n$  的列向量组成, 只是各项现在是复数。加法和乘法与  $\mathbb{R}^n$  中的定义完全相同, 唯一的区别是我们现在可以乘以**复数**, 即  $\mathbb{C}^n$  是一个**复向量空间**。

本书中的许多结果对于  $\mathbb{R}^n$  和  $\mathbb{C}^n$  都成立。这种情况下我们将使用符号  $\mathbb{F}^n$ 。

**例子** 空间  $M_{m \times n}$  (也记作  $M_{m,n}$ ) 是  $m \times n$  矩阵的集合: 加法和标量乘法是逐项定义的。如果我们只允许实数项存在 (因此只允许实数乘法), 那么我们会得到一个实向量空间; 如果我们允许复数项和复数乘法存在, 那么我们会得到一个复向量空间。

形式上, 我们必须区分实数情况和复数情况, 即写成  $M_{m,n}^{\mathbb{R}}$  或  $M_{m,n}^{\mathbb{C}}$ . 然而, 在大多数情况下, 实数和复数情况之间没有区别, 我们也无需指明正在考虑哪种情况。当有区别时, 我们会明确说明正在考虑哪种情况。

**注记** 正如我们上面提到的, 向量空间的公理仅仅是我们所熟知的 (实数或复数) 数字代数运算的规则, 所以如果我们把标量 (数字) 当作向量, 所有公理都会得到满足! 因此, 实数集  $\mathbb{R}$  也是一个实向量空间, 复数集  $\mathbb{C}$  也是一个复向量空间。

更重要的是, 由于在上面的例子中, 所有向量运算 (加法和标量乘法) 都是逐项执行的, 因此对于这些例子, 向量空间的公理自动满足, 因为它们对于标量也是成立的 (你能看出为什么吗?). 所以, 我们不必检验公理本身, 就自然地获得了这些例子确实是向量空间的事实!

同样的情况也适用于下面的例子, 即多项式, 其中多项式的系数起着项的作用。

**例子** 考虑最多  $n$  次的多项式的空间  $\mathbb{P}_n$ , 包含所有形式为

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$$

的多项式, 其中  $t$  是自变量。注意, 一些或甚至所有系数  $a_k$  可以是 0。

在  $a_k$  为实数系数的情况下, 我们得到一个实向量空间, 复数系数则构成一个复向量空间。同样, 我们只在区别至关重要时才明确说明我们正在处理实数或复数情况; 否则, 讨论的一切都适用于这两种情况。

**问题** 在以上每个例子中, 零向量各是什么?

## 1.2. 矩阵表示

一个  $m \times n$  矩阵是含有  $m$  行和  $n$  列的矩形数组。<sup>3</sup>数组中的元素称为矩阵的项 (entry)。

通常我们可以方便地用带下标的字母来表示矩阵的项: 第一个下标表示项所

---

<sup>3</sup>译者注: 需要注意的是, 矩阵中元素的定位规则与中学里平面直角坐标系点的查找方式恰好相反。中学是“纵横坐标”, 坐标轴分别在水平和竖直方向上延伸; 而矩阵中, 行的增加在竖直方向, 列的增加在水平方向。

在的行号，第二个下标表示列号。例如，

$$(1.1) \quad A = (a_{j,k})_{m \times n, j=1, k=1}^n = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

是一种一般地书写  $m \times n$  矩阵的方式。

对于矩阵  $A$ ，位于第  $j$  行和第  $k$  列的项常常被表示为  $A_{j,k}$  或  $(A)_{j,k}$ ，有时也会像上面 (1.1) 那个例子一样，用相同的小写字母来表示矩阵的项。

给定矩阵  $A$ ，它的**转置 (transpose)** (或转置矩阵)  $A^T$  是通过将  $A$  的行变为列来定义的。例如

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

所以， $A^T$  的列是  $A$  的行，反之亦然， $A^T$  的行是  $A$  的列。

正式定义如下： $(A^T)_{j,k} = (A)_{k,j}$ 。这种表示的意思是  $A^T$  中第  $j$  行第  $k$  列的项等于  $A$  中第  $k$  行第  $j$  列的项。

对矩阵的转置有一个线性变换上很好的解释，即它给出了所谓的**伴随 (adjoint)** 变换。我们将在后面详细讨论这一点，现在转置只是一个有用的形式运算。

转置的一个早期用途是我们可以将列向量  $\mathbf{x} \in \mathbb{F}^n$  (回想一下  $\mathbb{F}$  是  $\mathbb{R}$  或  $\mathbb{C}$ ) 写成  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 。如果我们将列向量垂直放置，它将占用纸面上更多的空间。

## 练习

**译者注：**本书练习题解答放在了附录中，请读者先思考，先动笔算，而不是一上来就看答案。本书的一大特色是，正文内容较为抽象且简略，而对于一些概念的更深入理解与辨析、应用、具象化、直观化，实践算法如何进行，以及少量提高内容，甚至一些定理的证明，作者都将其放在了练习中。故读者即使不将本书作为主要教材，仅是看书的话，也建议看着答案把题目过一遍。只有这样，你才知道自己是否真正理解了正文内容，适不适合立即往后看。以适量练习题作为补充，你的收获也将会更多。

1.1. 令  $\mathbf{x} = (1, 2, 3)^T$ ,  $\mathbf{y} = (y_1, y_2, y_3)^T$ ,  $\mathbf{z} = (4, 2, 1)^T$ . 计算  $2\mathbf{x}$ ,  $3\mathbf{y}$ ,  $\mathbf{x} + 2\mathbf{y} - 3\mathbf{z}$ .

1.2. 下列集合 (在自然定义的加法和标量乘法下) 哪些是向量空间? 请给出你的理由。

- a)  $[0, 1]$  区间上所有连续函数的集合;
- b)  $[0, 1]$  区间上所有非负函数的集合;
- c) 恰好  $n$  次多项式的集合;
- d) 所有  $n \times n$  对称矩阵的集合, 即满足  $A^T = A$  的矩阵  $A = \{a_{j,k}\}_{j,k=1}^n$ .

1.3. 判断正误:

- a) 每个向量空间都包含一个零向量;
- b) 一个向量空间可以有多个零向量;
- c) 一个  $m \times n$  矩阵有  $m$  行和  $n$  列;



- d) 如果  $f$  和  $g$  是  $n$  次多项式, 那么  $f + g$  也是正好为  $n$  次的多项式;  
 e) 如果  $f$  和  $g$  是最高为  $n$  次的多项式, 那么  $f + g$  也是最高为  $n$  次的多项式。

- 1.4. 证明向量空间  $V$  的零向量  $\mathbf{0}$  是唯一的。  
 1.5. 空间  $M_{2 \times 3}$  的零向量是什么样子的矩阵? 请写出来。  
 1.6. 证明向量空间公理 4 中定义加法逆元是唯一的。  
 1.7. 证明  $0\mathbf{v} = \mathbf{0}$  对任何向量  $\mathbf{v} \in V$ 。  
 1.8. 证明对任何向量  $\mathbf{v}$ , 其加法逆元  $-\mathbf{v}$  由  $(-1)\mathbf{v}$  给出。

## 2. 线性组合, 基

设  $V$  为向量空间, 又设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in V$  为一组向量。向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  的线性组合 (linear combination) 是形式为

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p = \sum_{k=1}^p \alpha_k \mathbf{v}_k$$

的和。

**定义** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$  称为  $V$  的基 (base) (或基底), 如果任何向量  $\mathbf{v} \in V$  都可以唯一地表示为其下的线性组合

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \sum_{k=1}^n \alpha_k \mathbf{v}_k.$$

系数  $\alpha_1, \alpha_2, \dots, \alpha_n$  称为向量  $\mathbf{v}$  的坐标 (coordinates) (在基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  下, 或相对于基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ )。

另一种说  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是基的方式是说, 对于任何可能的上文等号右侧  $\mathbf{v}_k$  的选择, 方程  $x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_n \mathbf{v}_n = \mathbf{v}$  (未知数为  $x_k$ ) 有唯一解。

在讨论基<sup>4</sup>的任何性质之前, 我会给出几个例子, 以说明这些对象确实存在, 并且研究它们是有意义的。

**例子 2.2.** 在第一个例子中, 空间  $V$  是  $\mathbb{F}^n$ , 其中  $\mathbb{F}$  是实数  $\mathbb{R}$  或复数  $\mathbb{C}$ . 考虑向量

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

(向量  $\mathbf{e}_k$  除第  $k$  个分量为 1 外, 其余分量均为 0)。向量组  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{F}^n$  的一

<sup>4</sup>基”basis”的复数是”bases”, 与”base”的复数相同。

组基。事实上, 任意向量  $\mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$  都可以表示为线性组合

$$\mathbf{v} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n = \sum_{k=1}^n x_k \mathbf{e}_k,$$

并且这种表示是唯一的。向量组  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n \in \mathbb{F}^n$  被称为  $\mathbb{F}^n$  中的**标准基** (standard basis)。

**例子 2.3.** 在这个例子中,  $\mathbb{P}_n$  是至多  $n$  次多项式构成的空间。考虑向量 (多项式)  $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n \in \mathbb{P}_n$  定义为

$$\mathbf{e}_0 := 1, \quad \mathbf{e}_1 := t, \quad \mathbf{e}_2 := t^2, \quad \mathbf{e}_3 := t^3, \quad \dots, \quad \mathbf{e}_n := t^n.$$

显然, 任意多项式  $p$ ,  $p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$  都存在唯一的表示

$$p = a_0 \mathbf{e}_0 + a_1 \mathbf{e}_1 + \cdots + a_n \mathbf{e}_n.$$

因此, 向量组  $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n \in \mathbb{P}_n$  是  $\mathbb{P}_n$  中的一组基。我们将它称为  $\mathbb{P}_n$  中的**标准基**。

**注记** 如果一个向量空间  $V$  拥有基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , 那么任何向量  $\mathbf{v}$  都可以由其在分解  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{v}_k$  中的系数唯一确定<sup>5</sup>。因此, 如果我们把系数  $\alpha_k$  堆叠成一个列向量, 我们就可以像处理列向量一样处理它们, 即像处理  $\mathbb{F}^n$  的元素一样 (同样, 这里的  $\mathbb{F}$  是  $\mathbb{R}$  或  $\mathbb{C}$ , 但所有内容也都适用于抽象域  $\mathbb{F}$ )。

具体来说, 如果  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{v}_k$  且  $\mathbf{w} = \sum_{k=1}^n \beta_k \mathbf{v}_k$ , 那么

$$\mathbf{v} + \mathbf{w} = \sum_{k=1}^n \alpha_k \mathbf{v}_k + \sum_{k=1}^n \beta_k \mathbf{v}_k = \sum_{k=1}^n (\alpha_k + \beta_k) \mathbf{v}_k.$$

也就是说, 要得到和的坐标列, 只需将各个向量的坐标列相加。类似地, 要得到  $\alpha \mathbf{v}$  的坐标, 只需将  $\mathbf{v}$  的坐标列乘以  $\alpha$ 。

## 2.1. 生成系统与线性无关系统

基的定义是任何向量都可以表示为其线性组合。这句话实际上包含两个陈述, 即表示存在和表示唯一。让我们分别分析这两个陈述。

如果我们只考虑存在性, 我们就得到以下概念:

**定义** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in V$  称为  $V$  中的**生成系统** (generating system) (也称为**张成系统** (spanning system) 或**完备系统** (complete system)), 如果任何向量

<sup>5</sup>这是一个非常重要的注记, 将在本书中贯穿使用。它允许我们将任何关于标准列空间  $\mathbb{F}^n$  的陈述转化为关于具有基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的向量空间  $V$  的陈述。

$\mathbf{v} \in V$  都可以表示为线性组合

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p = \sum_{k=1}^p \alpha_k \mathbf{v}_k.$$

与基的定义不同之处在于, 我们不假定上面的表示是唯一的。

这里, “生成”、“张成”和“完备”是同义词。我个人更喜欢“完备”这个词, 只因我的算子理论研究背景。生成和张成往往在更常见的线性代数教材中得到使用。

显然, 任何基都是生成(完备)系统。此外, 如果我们有一组基, 例如  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , 并且我们往其中添加几个向量, 例如  $\mathbf{v}_{n+1}, \dots, \mathbf{v}_p$ , 那么新的系统也将是生成(完备)系统。实际上, 我们可以将任一向量表示为向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的线性组合, 并将新向量(通过将相应的系数  $\alpha_k = 0$ )忽略掉。

现在来关注唯一性。这里我们不想担心存在性, 所以让我们考虑零向量  $\mathbf{0}$ , 它总是可以表示为线性组合。

**定义** 线性组合  $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p$  称为**平凡的**(trivial), 如果  $\alpha_k = 0 \forall k$ 。

平凡线性组合总是(对于所有选择的向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ )等于  $\mathbf{0}$ , 这也许就是“平凡”这个名字的由来。

**定义** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in V$  称为**线性无关**(linearly independent)的, 如果只有平凡线性组合( $\sum_{k=1}^p \alpha_k \mathbf{v}_k$  其中  $\alpha_k = 0 \forall k$ )等于  $\mathbf{0}$ 。

换句话说, 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  是线性无关的, 当且仅当方程  $x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_p \mathbf{v}_p = \mathbf{0}$  (未知数为  $x_k$ ) 只有一个平凡解  $x_1 = x_2 = \cdots = x_p = 0$ 。

如果系统不是线性无关的, 则称为是**线性相关**(linearly dependent)的。通过将线性无关的定义取否, 我们得到以下定义:

**定义** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  称为**线性相关的**, 如果  $\mathbf{0}$  可以表示为**非平凡**的线性组合, 即  $\mathbf{0} = \sum_{k=1}^p \alpha_k \mathbf{v}_k$ . 非平凡意味着至少有一个系数  $\alpha_k$  非零。这可以(并且通常)写成  $\sum_{k=1}^p |\alpha_k| \neq 0$ 。

因此, 重申定义, 我们可以说, 一个系统是线性相关的, 当且仅当存在不全为零的标量  $\alpha_1, \alpha_2, \dots, \alpha_p$ , 使得

$$\sum_{k=1}^p \alpha_k \mathbf{v}_k = \mathbf{0}.$$

另一个定义(关于方程)是, 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  是线性相关的, 当且仅当方程

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_p \mathbf{v}_p = \mathbf{0}$$

(未知数为  $x_k$ ) 有一个非平凡解。非平凡, 再次意味着至少有一个  $x_k$  不为零, 这可以写成  $\sum_{k=1}^p |x_k| \neq 0$ 。

以下命题提供了线性相关系统的一个替代描述。

**命题 2.6.** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in V$  是线性相关的, 当且仅当其中一个向量  $\mathbf{v}_k$  可以表示为其他向量的线性组合,

$$(2.1) \quad \mathbf{v}_k = \sum_{j=1, j \neq k}^p \beta_j \mathbf{v}_j.$$

**证明:** 假设系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  是线性相关的。那么存在不全为零的标量  $\alpha_k$  ( $\sum_{k=1}^p |\alpha_k| \neq 0$ ), 使得

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_p \mathbf{v}_p = \mathbf{0}.$$

设  $k$  是  $\alpha_k \neq 0$  的下标。那么, 将除  $\alpha_k \mathbf{v}_k$  之外的所有项移到右侧, 我们得到

$$\alpha_k \mathbf{v}_k = - \sum_{j=1, j \neq k}^p \alpha_j \mathbf{v}_j.$$

将两边除以  $\alpha_k$ , 我们得到 (2.1) 式, 其中  $\beta_j = -\alpha_j/\alpha_k$ 。

另一方面, 如果 (2.1) 式成立, 则  $\mathbf{0}$  可以表示为非平凡线性组合

$$\mathbf{v}_k - \sum_{j=1, j \neq k}^p \beta_j \mathbf{v}_j = \mathbf{0}.$$

□

显然, 任何基都是线性无关的系统。实际上, 如果一个系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是基, 则  $\mathbf{0}$  允许唯一表示

$$\mathbf{0} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n = \sum_{k=1}^n \alpha_k \mathbf{v}_k.$$

因为平凡线性组合总是给出  $\mathbf{0}$ , 所以平凡组合必须是给出  $\mathbf{0}$  的唯一组合。

因此, 正如我们已经讨论过的, 如果一个系统是基, 那么它就是完备 (生成) 的并且线性无关的系统。以下命题表明其必要性也成立。

**命题 2.7.** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$  是基, 当且仅当它线性无关且完备 (生成)。

**证明:** 我们已经知道基总是线性无关且完备的, 所以命题的一个方向已经证明。

让我们证明另一个方向。假设系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是线性完备的。取任意向量  $\mathbf{v} \in V$ . 由于系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是线性完备的 (生成的),  $\mathbf{v}$  可以表示为

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n = \sum_{k=1}^n \alpha_k \mathbf{v}_k.$$

我们只需要证明这个表示是唯一的。

假设  $\mathbf{v}$  还有另一个表示

$$\mathbf{v} = \sum_{k=1}^n \tilde{\alpha}_k \mathbf{v}_k.$$

那么

$$\sum_{k=1}^n (\alpha_k - \tilde{\alpha}_k) \mathbf{v}_k = \sum_{k=1}^n \alpha_k \mathbf{v}_k - \sum_{k=1}^n \tilde{\alpha}_k \mathbf{v}_k = \mathbf{v} - \mathbf{v} = \mathbf{0}.$$

由于系统是线性无关的,  $\alpha_k - \tilde{\alpha}_k = 0 \forall k$ , 因此, 表示  $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n$  是唯一的。□

**注记** 在许多教材中, 基被定义为完备且线性无关的系统 (根据命题 2.7, 这个定义等价于我们的定义)。虽然这个定义比本书提出的定义更常见, 但我更喜欢后者。它强调了基的主要性质, 即任何向量都可以唯一地表示为基下的线性组合。

**命题 2.8.** 任何 (有限) 生成系统都包含一组基。

**证明:** 假设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p \in V$  是生成 (完备) 集。如果它是线性无关的, 那么它就是基, 我们就完成了证明。

假设它不是线性无关的, 即它是线性相关的。那么将存在一个向量  $\mathbf{v}_k$  可以表示为向量  $\mathbf{v}_j$  ( $j \neq k$ ) 的线性组合。

由于  $\mathbf{v}_k$  可以表示为向量  $\mathbf{v}_j$  ( $j \neq k$ ) 的线性组合, 因此任何向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  的线性组合都可以表示为相同的向量 (即  $\mathbf{v}_j, 1 \leq j \leq p, j \neq k$ ) 的线性组合 (即删掉  $\mathbf{v}_k$  之后的向量)。因此, 如果我们删除向量  $\mathbf{v}_k$ , 新的系统仍然是完备的。

如果新的系统是线性无关的, 我们就完成了证明。如果不是, 则重复这个过程。

重复有限次该过程后, 我们将得到一个线性无关且完备的系统, 否则我们将删除所有向量, 最后得到一个空集。

因此, 任何有限的完备 (生成) 集都包含一个完备的线性无关子集, 即一组基。□

## 练习

2.1. 在  $3 \times 2$  矩阵空间  $M_{3 \times 2}$  中找到一组基。

- 2.2. 判断正误：
- 包含零向量的任何集合都是线性相关的；
  - 基必须包含  $\mathbf{0}$ ；
  - 线性相关集的子集是线性相关的；
  - 线性无关集的子集是线性无关的；
  - 如果  $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}$ ，那么所有标量  $\alpha_k$  都为零。
- 2.3. 回忆一下，如果  $A^T = A$ ，则矩阵称为**对称**矩阵。写下一个  $2 \times 2$  对称矩阵空间的基（有许多可能的答案）。基中有多少个元素？
- 2.4. 写出以下空间的基：
- $3 \times 3$  对称矩阵；
  - $n \times n$  对称矩阵；
  - $n \times n$  反对称矩阵 ( $A^T = -A$ )。
- 2.5. 设向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是线性无关的，但不是生成的。证明可以找到向量  $\mathbf{v}_{r+1}$  使得系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}$  是线性无关的。  
提示：选择任何不能表示为  $\sum_{k=1}^r \alpha_k \mathbf{v}_k$  的向量作为  $\mathbf{v}_{r+1}$ ，并证明系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}$  是线性无关的。
- 2.6. 是否可能存在，向量  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  是线性相关的，而向量  $\mathbf{w}_1 = \mathbf{v}_1 + \mathbf{v}_2$ ,  $\mathbf{w}_2 = \mathbf{v}_2 + \mathbf{v}_3$  和  $\mathbf{w}_3 = \mathbf{v}_3 + \mathbf{v}_1$  是线性无关的？

### 3. 线性变换 矩阵-向量乘法

从集合  $X$  到集合  $Y$  的**变换**<sup>6</sup> (transformation) $T$  是一个规则，它为每个自变量（输入） $x \in X$  分配一个值（输出） $y = T(x) \in Y$ 。

集合  $X$  称为  $T$  的**定义域** (domain)，集合  $Y$  称为  $T$  的**目标空间** (target space) 或**上域** (codomain)。

我们用  $T: X \rightarrow Y$  来表示  $T$  是一个定义域为  $X$ 、目标空间为  $Y$  的变换。

**定义** 设  $V, W$  为向量空间（在同一域  $\mathbb{F}$  上）。变换  $T: V \rightarrow W$  称为**线性** (linear) 的（或具有线性性质），如果

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in V$ ;
- $T(\alpha \mathbf{v}) = \alpha T(\mathbf{v})$  对所有  $\mathbf{v} \in V$  和所有标量  $\alpha \in \mathbb{F}$ 。

性质 1 和 2 一起等价于这个性质： $T(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$  对所有  $\mathbf{u}, \mathbf{v} \in V$  和所有标量  $\alpha, \beta$ 。

#### 3.1. 一些例子

你以前一定接触过线性变换，甚至可能没有意识到，如下例所示。

---

<sup>6</sup>单词“变换” (transformation)、“映射” (map, mapping)、“算子” (operator)、“函数” (function) 这些词都表示同一个概念。

**例子 求导:** 设  $V = \mathbb{P}_n$  (至多  $n$  次多项式的集合),  $W = \mathbb{P}_{n-1}$ , 令  $T: \mathbb{P}_n \rightarrow \mathbb{P}_{n-1}$  为求导的算子,

$$T(p) := p' \quad \forall p \in \mathbb{P}_n.$$

因为  $(f+g)' = f' + g'$  且  $(\alpha f)' = \alpha f'$ , 所以这是一个线性变换。

**例子 旋转:** 在这个例子中,  $V = W = \mathbb{R}^2$  (通常的坐标平面), 一个变换  $T_\gamma: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  将  $\mathbb{R}^2$  中的一个向量逆时针旋转  $\gamma$  弧度。由于  $T_\gamma$  整体将平面旋转, 它也整体旋转了用于定义两个向量之和的平行四边形 (向量加法的平行四边形法则)。因此, 线性变换的性质 1 成立。也很容易看出性质 2 也成立。

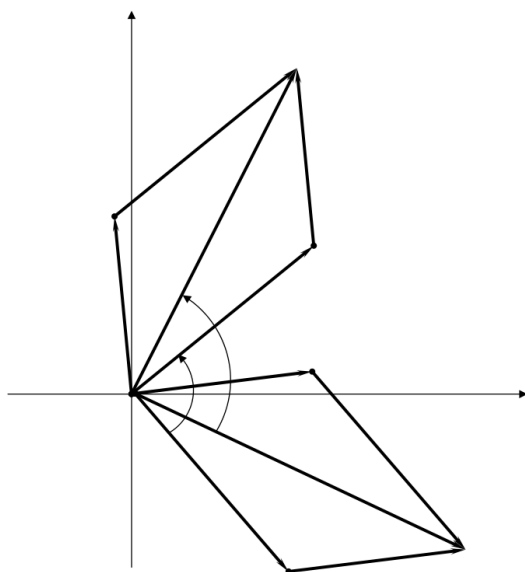


图 0.1 旋转

**例子 反射:** 在这个例子中, 同样  $V = W = \mathbb{R}^2$ , 变换  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  是关于第一个坐标轴的反射, 参见图 0.1。也可以几何地证明这个变换是线性的, 但我们将使用另一种方法来证明。

即, 很容易写出  $T$  的表达式:

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$$

并且从该表达式可以看出, 这个变换是线性的。

**例子** 让我们来研究线性变换  $T: \mathbb{R} \rightarrow \mathbb{R}$ . 任何这样的变换都由公式  $T(x) = ax$  给出, 其中  $a = T(1)$ . 事实上,

$$T(x) = T(x \times 1) = xT(1) = xa = ax.$$

因此,  $\mathbb{R}$  的任何线性变换仅仅是将一个数乘上一个常数。



### 3.2. 线性变换 $\mathbb{F}^n \rightarrow \mathbb{F}^m$ 矩阵-向量乘法

事实证明, 从  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性变换  $T$  也能表示为一种乘法, 但这不是乘标量, 而是乘矩阵。我们来看看怎么做。

设  $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$  是一个线性变换。计算  $T(\mathbf{x})$  对所有向量  $\mathbf{x} \in \mathbb{F}^n$ , 需要哪些信息? 我的主张是, 知道  $T$  如何作用于  $\mathbb{F}^n$  的标准基  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  就足够了。也就是说, 知道  $n$  个  $\mathbb{F}^m$  中的向量 (即  $n$  个大小为  $m$  的向量),

$$\mathbf{a}_1 = T(\mathbf{e}_1), \quad \mathbf{a}_2 := T(\mathbf{e}_2), \quad \dots, \quad \mathbf{a}_n := T(\mathbf{e}_n)$$

就足够了。

实际上, 设

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

那么  $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n = \sum_{k=1}^n x_k\mathbf{e}_k$  并且

$$T(\mathbf{x}) = T\left(\sum_{k=1}^n x_k\mathbf{e}_k\right) = \sum_{k=1}^n T(x_k\mathbf{e}_k) = \sum_{k=1}^n x_k T(\mathbf{e}_k) = \sum_{k=1}^n x_k\mathbf{a}_k.$$

因此, 如果我们把 (列) 向量  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  组合成一个矩阵  $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$  ( $\mathbf{a}_k$  是  $A$  的第  $k$  列,  $k = 1, 2, \dots, n$ ), 这个矩阵就包含了关于  $T$  的所有信息。让我们看看如何定义矩阵与 (列) 向量的乘积以将变换  $T$  表示为乘积,  $T(\mathbf{x}) = A\mathbf{x}$ . 设

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}.$$

回忆一下,  $A$  的第  $k$  列是向量  $\mathbf{a}_k$ , 即

$$\mathbf{a}_k = \begin{pmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{m,k} \end{pmatrix}.$$

那么, 如果我们希望  $A\mathbf{x} = T(\mathbf{x})$ , 我们得到

$$A\mathbf{x} = \sum_{k=1}^n x_k\mathbf{a}_k = x_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + x_2 \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix}$$

所以，矩阵-向量乘法应该通过以下**按列坐标规则** (column by coordinating rule) 执行：

分别将矩阵的每一列乘以向量的相应坐标。

**例子**

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 3 \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+4+9 \\ 3+4+3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}.$$

“按列坐标规则”对于表示乘积的变换非常适用。它在后面不同的理论构造中也将会非常重要。

然而，在手动计算时，逐项计算结果更方便。这可以表示为以下**按行列规则** (row by column rule)：

要得到结果的第  $k$  项，需要将矩阵的第  $k$  行乘以向量，即，如果  $A\mathbf{x} = \mathbf{y}$ ，那么  $y_k = \sum_{j=1}^n a_{k,j}x_j$ ,  $k = 1, 2, \dots, m$ ;

这里  $x_j$  和  $y_k$  分别是向量  $\mathbf{x}$  和  $\mathbf{y}$  的坐标，而  $a_{j,k}$  是矩阵  $A$  的项。

**例子**

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 \\ 4 \cdot 1 + 5 \cdot 2 + 6 \cdot 3 \end{pmatrix} = \begin{pmatrix} 1+4+9 \\ 4+10+18 \end{pmatrix} = \begin{pmatrix} 14 \\ 32 \end{pmatrix}.$$

### 3.3. 线性变换与生成集

正如我们在上面讨论的，作用于  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性变换  $T$  完全由其在  $\mathbb{F}^n$  标准基上的值定义。

我们虽然在上文用了标准基，但并没有利用其特殊性，其实也可以考虑任何基，甚至任何生成（张成）集。也就是说，

线性变换  $T: V \rightarrow W$  完全由其在生成集上的值（特别地，由其在基上的值）定义。

因此，如果  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的生成集（特别地，如果它是基），并且  $T$  和  $T_1$  是（具有相同定义域和目标空间的）两个线性变换 ( $T, T_1: V \rightarrow W$ ) 并且

$$T\mathbf{v}_k = T_1\mathbf{v}_k, k = 1, 2, \dots, n,$$

则  $T = T_1$ .

这个命题的证明是显然的，留作读者的练习。

### 3.4. 结论

- 要获得  $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$  的线性变换的矩阵, 只需将向量  $\mathbf{a}_k = T\mathbf{e}_k$  (其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{F}^n$  的标准基) 组合成一个矩阵: 矩阵的第  $k$  列是  $\mathbf{a}_k$ ,  $k = 1, 2, \dots, n$ .
- 如果已知线性变换  $T$  的矩阵  $A$ , 则  $T(\mathbf{x})$  可以通过矩阵-向量乘法找到,  $T(\mathbf{x}) = A\mathbf{x}$ . 要执行矩阵-向量乘法, 可以使用“按列坐标规则”或“按行列规则”。

后者似乎更适合手动计算。前者非常适合并行计算, 并且将在后面不同的理论构造中使用。

对于线性变换  $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ , 其矩阵通常记作  $[T]$ . 然而, 人们常常不区分线性变换和它的矩阵, 并使用相同的符号表示两者。当这不引起混淆时, 我们也将使用相同的符号表示变换和它的矩阵。

由于线性变换本质上是乘法, 因此  $T\mathbf{v}$  这个表示通常会被采用, 而不是  $T(\mathbf{v})$ <sup>7</sup>。我们将使用这种表示法。注意, 通常的代数运算顺序(优先级)是适用的, 即  $T\mathbf{v} + \mathbf{u}$  表示  $T(\mathbf{v}) + \mathbf{u}$ , 而不是  $T(\mathbf{v} + \mathbf{u})$ 。

**注记** 在矩阵-向量乘法  $A\mathbf{x}$  中, 矩阵  $A$  的列数必须与向量  $\mathbf{x}$  的大小一致<sup>8</sup>, 即  $\mathbb{F}^n$  中的向量只能被  $m \times n$  矩阵相乘。

这是有意义的, 因为  $m \times n$  矩阵定义了一个从  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性变换, 所以向量  $\mathbf{x}$  必须属于  $\mathbb{F}^n$ 。

最简单的记住这个事实的方法是, 如果在进行乘法时, 只要你先用完了一种种类的项, 这时还有另一些项未使用, 那么乘法就是不被定义的。

**注记** 不需要将自己局限于标准基与  $\mathbb{F}^n$  的情况: 当存在定义域和目标空间中的基时, 本节中描述的所有内容都适用于任意向量空间。当然, 如果改变了基, 线性变换的矩阵也会不同。这将在后面第二章第 8 节中讨论。

#### 练习

3.1. 进行乘法运算:

$$\begin{aligned} \text{a)} & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}; \\ \text{b)} & \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix}; \end{aligned}$$

<sup>7</sup> $T\mathbf{v}$  的表示比  $T(\mathbf{v})$  更常用。

<sup>8</sup>在使用“按行列规则”进行矩阵-向量乘法时, 请确保行中的项数与列中的项数相同。行和列的项应该同时结束: 如果不满足, 则乘法不被定义。

$$\begin{aligned} \text{c)} & \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}; \\ \text{d)} & \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}^9. \end{aligned}$$

3.2. 找到  $\mathbb{R}^2$  中关于直线  $x_1 = 3x_2$  的反射的线性变换的矩阵。

3.3. 求下列线性变换对应的矩阵：

a)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ , 定义为  $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x + 2y \\ 2x - 5y \\ 7y \end{pmatrix}$ ;

b)  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^3$ , 定义为  $T(x_1, x_2, x_3, x_4)^T = (x_1 + x_2 + x_3 + x_4, x_2 - x_4, x_1 + 3x_2 + 6x_4)^T$ ;

c)  $T: \mathbb{P}_n \rightarrow \mathbb{P}_n$ ,  $Tf(t) = f'(t)$  (在标准基  $1, t, t^2, \dots, t^n$  下找到矩阵) ;

d)  $T: \mathbb{P}_n \rightarrow \mathbb{P}_n$ ,  $Tf(t) = 2f(t) + 3f'(t) - 4f''(t)$  (同样在标准基  $1, t, t^2, \dots, t^n$  下找到矩阵) .

3.4. 以下线性变换在  $\mathbb{R}^3$  中作用。分别求它们对应的  $3 \times 3$  矩阵：

a) 将每个向量投影到  $x - y$  平面；

b) 将每个向量反射到  $x - y$  平面；

c) 将  $x - y$  平面绕  $z$  轴旋转  $30^\circ$ ，同时保持  $z$  轴不变。

3.5. 设  $A$  是一个线性变换。如果  $\mathbf{z}$  是线段  $[\mathbf{x}, \mathbf{y}]$  的中点，证明  $A\mathbf{z}$  是线段  $[A\mathbf{x}, A\mathbf{y}]$  的中点。

提示： $\mathbf{z}$  是线段  $[\mathbf{x}, \mathbf{y}]$  的中点意味着什么？

3.6. 复数集  $\mathbb{C}$  可以通过将  $z = x + iy \in \mathbb{C}$  视为列向量  $(x, y)^T \in \mathbb{R}^2$  来规范对应 (be canonically identified)。

a) 将  $\mathbb{C}$  视为复向量空间，证明通过  $\alpha = a + ib \in \mathbb{C}$  的乘法是在  $\mathbb{C}$  中的线性变换。它的矩阵是什么？

b) 将  $\mathbb{C}$  视为实向量空间  $\mathbb{R}^2$ ，证明通过  $\alpha = a + ib \in \mathbb{C}$  的乘法在那里定义了一个线性变换。它的矩阵是什么？

c) 定义  $T(x + iy) = 2x - y + i(x - 3y)$ 。证明这个变换不是  $\mathbb{C}$  复向量空间中的线性变换，但如果我们把  $\mathbb{C}$  视为实向量空间  $\mathbb{R}^2$ ，那么它在那里是一个线性变换 (即  $T$  是一个**实线性**但不是**复线性**变换)。找到这个实线性变换的矩阵。

3.7. 证明  $\mathbb{C}$  中的任何线性变换 (视为复向量空间) 都是通过乘以  $\alpha \in \mathbb{C}$  来实现的。

## 4. 线性变换作为向量空间

我们可以对线性变换进行哪些运算？我们总是可以在一个线性变换上乘上一个标量，也即，如果我们有一个线性变换  $T: V \rightarrow W$  和一个标量  $\alpha$ ，我们可以定义一个新的变换  $\alpha T$  为

$$(\alpha T)\mathbf{v} = \alpha(T\mathbf{v}) \quad \forall \mathbf{v} \in V.$$

<sup>9</sup>译者注：原书如此。

可以很容易地检验,  $\alpha T$  也是一个线性变换:

$$\begin{aligned} (\alpha T)(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2) &= \alpha(T(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2)) \quad (\text{根据 } \alpha T \text{ 的定义}) \\ &= \alpha(\alpha_1 T \mathbf{v}_1 + \alpha_2 T \mathbf{v}_2) \quad (\text{根据 } T \text{ 的线性性质}) \\ &= \alpha_1 \alpha T \mathbf{v}_1 + \alpha_2 \alpha T \mathbf{v}_2 \\ &= \alpha_1 (\alpha T) \mathbf{v}_1 + \alpha_2 (\alpha T) \mathbf{v}_2. \end{aligned}$$

如果  $T_1$  和  $T_2$  是具有相同定义域和目标空间的线性变换 ( $T_1 : V \rightarrow W$  且  $T_2 : V \rightarrow W$ , 或简写为  $T_1, T_2 : V \rightarrow W$ ), 那么我们可以将这些变换相加, 即定义一个新的变换  $T = (T_1 + T_2) : V \rightarrow W$  为

$$(T_1 + T_2)\mathbf{v} = T_1\mathbf{v} + T_2\mathbf{v} \quad \forall \mathbf{v} \in V.$$

可以很容易地检验出变换  $T_1 + T_2$  也是线性的, 只需重复上面关于  $\alpha T$  线性的推理即可。

因此, 如果我们固定向量空间  $V$  和  $W$  并考虑从  $V$  到  $W$  的所有线性变换的集合 (我们将其表示为  $\mathcal{L}(V, W)$ ), 我们可以定义  $\mathcal{L}(V, W)$  上的 2 个运算: 标量乘法和加法。可以很容易地证明这些运算满足向量空间公理 (该公理在第 1 节中定义)。

这里, 读者不应该感到惊讶, 因为向量空间的公理本质上意味着有关向量的运算应该遵循自然的代数运算规则。而线性变换上的运算定义就是为了满足这些规则!

作为说明, 让我们为向量空间公理的第一个分配律 (公理 7) 写下正式证明。我们想证明  $\alpha(T_1 + T_2) = \alpha T_1 + \alpha T_2$ . 对于  $V$  中的任何  $\mathbf{v}$ ,

$$\begin{aligned} \alpha(T_1 + T_2)\mathbf{v} &= \alpha((T_1 + T_2)\mathbf{v}) \quad (\text{根据乘法的定义}) \\ &= \alpha(T_1\mathbf{v} + T_2\mathbf{v}) \quad (\text{根据和的定义}) \\ &= \alpha T_1\mathbf{v} + \alpha T_2\mathbf{v} \quad (\text{根据 } W \text{ 中的公理 7}) \\ &= (\alpha T_1 + \alpha T_2)\mathbf{v} \quad (\text{根据和的定义}) \end{aligned}$$

因此, 确实  $\alpha(T_1 + T_2) = \alpha T_1 + \alpha T_2$ .

**注记** 线性运算 (加法和标量乘法) 在  $T : \mathbb{F}^n \rightarrow \mathbb{F}^m$  的线性变换上对应于它们矩阵上的相应运算。由于我们知道  $m \times n$  矩阵的集合是一个向量空间, 这就立即意味着  $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$  也是一个向量空间。

我们率先提出了抽象证明, 首先是因为它适用于一般的空间, 例如, 适用于没有基的向量空间, 在那里我们不能使用坐标; 其次, 这里展示的抽象推理在其他许多地方都会使用, 所以读者将受益于理解它。

并且随着读者对数学的理解慢慢深入, 他/她将看到这种抽象推理确实是非常简单的, 几乎可以自动完成。

## 5. 线性变换的复合与矩阵乘法

### 5.1. 矩阵乘法的定义

知道了矩阵-向量乘法，人们很容易猜出两个矩阵乘积  $AB$  的自然定义：让我们用  $A$  乘以  $B$  的每个列（矩阵-向量乘法），并将得到的列向量连接成一个矩阵。形式上，

如果  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r$  是  $B$  的列，那么  $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_r$  就是矩阵  $AB$  的列。

回忆矩阵-向量乘法的“按行列规则”，我们得到矩阵的**按行列规则**：

$AB$  的项  $(AB)_{j,k}$ （第  $j$  行第  $k$  列的项）定义为  
 $(AB)_{j,k} = (A \text{ 的第 } j \text{ 行}) \cdot (B \text{ 的第 } k \text{ 列})$ 。

形式上可以写成

$$(AB)_{j,k} = \sum_l a_{j,l} b_{l,k},$$

如果  $a_{j,k}$  和  $b_{j,k}$  分别是矩阵  $A$  和  $B$  的项。

我特意没有提及矩阵  $A$  和  $B$  的大小，但如果我们回忆矩阵-向量乘法的按行列规则，我们可以看到，为了使乘法有定义， $A$  的行的大小应等于  $B$  的列的大小。

换句话说，乘积  $AB$  有定义当且仅当  $A$  是  $m \times n$  的矩阵， $B$  是  $n \times r$  的矩阵。（这与从按行列规则获得的条件相同。）

### 5.2. 动机：线性变换的复合

在这里可以问问自己：为什么我们要使用如此复杂的乘法规则？为什么我们不直接逐项对应地将矩阵相乘<sup>10</sup>？

答案是，如上定义的乘法自然地源于线性变换的复合。

假设我们有两个线性变换， $T_1: \mathbb{F}^n \rightarrow \mathbb{F}^m$  和  $T_2: \mathbb{F}^r \rightarrow \mathbb{F}^n$ 。定义变换的**复合** (composition)  $T = T_1 \circ T_2$  为

$$T(\mathbf{x}) = T_1(T_2(\mathbf{x})) \quad \forall \mathbf{x} \in \mathbb{F}^r.$$

请注意， $T_2(\mathbf{x}) \in \mathbb{F}^n$ 。由于  $T_1: \mathbb{F}^n \rightarrow \mathbb{F}^m$ ，表达式  $T_1(T_2(\mathbf{x}))$  是有定义的，并且结果属于  $\mathbb{F}^m$ 。所以， $T: \mathbb{F}^r \rightarrow \mathbb{F}^m$ 。<sup>11</sup>

可以很容易地证明  $T$  是一个线性变换（留作练习），所以它由一个  $m \times r$  矩阵定义。已知  $T_1$  和  $T_2$  的矩阵，如何找到  $T$  的矩阵？

令  $A$  为  $T_1$  的矩阵，令  $B$  为  $T_2$  的矩阵。正如我们在上一节中所讨论的， $T$  的列是向量  $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_r)$ ，其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r$  是  $\mathbb{F}^r$  中的标准基。对于

<sup>10</sup>译者注：此即矩阵逐项积的定义。

<sup>11</sup>我们通常将线性变换与其矩阵等同，但在接下来的几个段落中，我们将区分它们。

$k = 1, 2, \dots, r$ , 我们有

$$T(\mathbf{e}_k) = T_1(T_2(\mathbf{e}_k)) = T_1(B\mathbf{e}_k) = T_1(\mathbf{b}_k) = A\mathbf{b}_k$$

(变换  $T_2$  和  $T_1$  分别就是乘  $B$  和乘  $A$ )。

所以,  $T$  的矩阵的列是  $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_r$ , 这正是矩阵  $AB$  的定义方式!

让我们回到与之等同的观点。由于矩阵乘法与复合一致, 我们可以 (并且将) 写作  $T_1T_2$  而不是  $T_1 \circ T_2$ , 以及  $T_1T_2\mathbf{x}$  而不是  $T_1(T_2(\mathbf{x}))$ .<sup>12</sup>

注意在组合  $T_1T_2$  中, 先作变换  $T_2$ ! 记住这种方法的一个办法是注意到  $T_1T_2\mathbf{x}$  中, 变换  $T_2$  首先与  $\mathbf{x}$  相遇。

**注记** 除了“按行列规则”的矩阵乘法之外, 还有另一种核验矩阵乘积维数的方法: 对于复合  $T_1T_2$  的定义, 必须使得  $T_2\mathbf{x}$  属于  $T_1$  的定义域。如果  $T_2$  作用于某个空间, 比如  $\mathbb{F}^r$  到  $\mathbb{F}^n$ , 那么  $T_1$  必须从  $\mathbb{F}^n$  作用到某个空间 (比如  $\mathbb{F}^m$ )。因此, 为了使  $T_1T_2$  有定义,  $T_1$  和  $T_2$  的矩阵的大小应该分别是  $m \times n$  和  $n \times r$  ——这与从“按行列规则”获得的条件相同。

**例子** 设  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  是关于直线  $x_1 = 3x_2$  的反射。这是一个线性变换, 那么让我们找到它的矩阵。为了找到矩阵, 我们需要计算  $T\mathbf{e}_1$  和  $T\mathbf{e}_2$ 。然而, 直接计算  $T\mathbf{e}_1$  和  $T\mathbf{e}_2$  需要的知识比一个理智的人愿意记住的三角学显然更多。

找到  $T$  的矩阵的一个更简单的方法是将其表示为简单线性变换的复合。也就是说, 设  $\gamma$  是  $x_1$  轴和直线  $x_1 = 3x_2$  之间的夹角, 设  $T_0$  是关于  $x_1$  轴的反射。那么要得到反射  $T$ , 我们可以先将平面旋转  $-\gamma$  角, 将直线  $x_1 = 3x_2$  移动到  $x_1$  轴, 然后将所得在  $x_1$  轴上反射, 最后将平面旋转  $\gamma$  角, 将所有东西移回原位。形式上可以写成

$$T = R_\gamma T_0 R_{-\gamma}$$

(注意乘积的顺序!), 其中  $R_\gamma$  是绕原点逆时针旋转  $\gamma$  角的变换矩阵。 $T_0$  的矩阵很容易计算, 是

$$T_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

旋转矩阵是已知的

$$R_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$$

$$R_{-\gamma} = \begin{pmatrix} \cos(-\gamma) & -\sin(-\gamma) \\ \sin(-\gamma) & \cos(-\gamma) \end{pmatrix} = \begin{pmatrix} \cos \gamma & \sin \gamma \\ -\sin \gamma & \cos \gamma \end{pmatrix}$$

<sup>12</sup>注意变换的顺序!



为了计算  $\sin \gamma$  和  $\cos \gamma$ ，取直线  $x_1 = 3x_2$  上的一个向量，例如向量  $(3, 1)^T$ 。那么

$$\cos \gamma = \frac{\text{向量的第一个坐标}}{\text{向量长度}} = \frac{3}{\sqrt{3^2 + 1^2}} = \frac{3}{\sqrt{10}},$$

类似地

$$\sin \gamma = \frac{\text{向量的第二个坐标}}{\text{向量长度}} = \frac{1}{\sqrt{3^2 + 1^2}} = \frac{1}{\sqrt{10}}.$$

将此前的所有内容合并起来，我们得到

$$\begin{aligned} T &= R_\gamma T_0 R_{-\gamma} = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \\ &= \frac{1}{10} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \end{aligned}$$

最后一步是进行矩阵乘法以得到最终结果，此处从略。

### 5.3. 矩阵乘法的性质

矩阵乘法享有许多性质，它们和我们从高中代数中学到的知识相类似：

1. 结合律： $A(BC) = (AB)C$ ，只要等式的一侧或另一侧有定义；因此我们可以（并且将）简单地写为  $ABC$ 。

2. 分配律： $A(B + C) = AB + AC$ ,  $(A + B)C = AC + BC$ ，只要每个等式的任一侧有定义。

3. 系数（标量）可以提取： $A(\alpha B) = (\alpha A)B = \alpha(AB) = \alpha AB$ 。

这些性质很容易证明。可以先证明它们对应于线性变换的性质，然后它们将几乎自然地定义中得出。因为线性变换的性质蕴含了矩阵乘法的性质。

这里新的特点是交换律失败了：

矩阵乘法是不可交换的，即通常  $AB \neq BA$ 。

容易看出，期望矩阵乘法满足交换律是不合理的。确实，如果  $A$  和  $B$  是  $m \times n$  和  $n \times r$  大小的矩阵，那么乘积  $AB$  有定义，但如果  $m \neq r$ ，则  $BA$  不被定义。

即使两个乘积都有定义，例如，当  $A$  和  $B$  是  $n \times n$ （方阵）时，乘法仍然是非交换的。如果我们随机选择矩阵  $A$  和  $B$ ，那么  $AB \neq BA$  的概率很高：我们非常幸运时才能得到  $AB = BA$  的结果。

### 5.4. 转置矩阵与乘法

转置的定义已在 1.2 节中给出。

一个简单的分析表明

$$(AB)^T = B^T A^T,$$

即当你取矩阵乘积的转置时, 你需要改变矩阵的顺序。

## 5.5. 迹与矩阵乘法

对于一个  $(n \times n)$  方阵  $A = (a_{j,k})$ , 它的迹(trace) (记作  $\text{trace } A$ ) 是其对角线上所有项的和:

$$\text{trace } A = \sum_{k=1}^n a_{k,k}$$

**定理 5.1.** 设  $A$  和  $B$  是大小分别为  $m \times n$  和  $n \times m$  的矩阵 (因此两个乘积  $AB$  和  $BA$  都有定义)。那么

$$\text{trace}(AB) = \text{trace}(BA).$$

我们将此定理的证明留作练习, 见下文问题 5.6. 证明此定理大体上有两种方法。一种方法是计算  $AB$  和  $BA$  的对角线项并比较它们的和。这种方法需要一些熟练处理带  $\sum$  符号求和的技巧。

如果你不熟悉代数运算, 还有另一种方法。我们可以考虑两个线性变换  $T$  和  $T_1$ , 它们作用于  $M_{n \times m}$  到  $\mathbb{F} = \mathbb{F}^1$ , 由

$$T(X) = \text{trace}(AX), \quad T_1(X) = \text{trace}(XA)$$

定义。为了证明该定理, 只需检验  $T = T_1$  即可; 当  $X = B$  时, 等式即给出该定理。

由于线性变换完全由其在生成系统上的值决定, 我们只需在一些简单的矩阵上检验等式是否成立, 之后即可推广至一般情况。例如矩阵  $X_{j,k}$ , 该矩阵除了在第  $j$  列和第  $k$  行的交汇处有一个 1 之外, 其余所有元素都是 0。

### 练习

5.1. 设  $A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & -2 \end{pmatrix}, C = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 1 & -1 \end{pmatrix}, D = \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}.$

- a) 标记所有有定义的乘积, 并给出结果的维数:  $AB, BA, ABC, ABD, BC, BC^T, B^T C, DC, D^T C^T$ .  
b) 计算  $AB, A(3B + C), B^T A, A(BD), (AB)D$ .

5.2. 设  $T_\gamma$  是  $\mathbb{R}^2$  中绕  $\gamma$  角旋转的矩阵。通过矩阵乘法验证  $T_\gamma T_{-\gamma} = T_{-\gamma} T_\gamma = I$ .

5.3. 乘以两个旋转矩阵  $T_\alpha$  和  $T_\beta$  (这是罕见的乘法可交换的情况, 即  $T_\alpha T_\beta = T_\beta T_\alpha$ , 所以顺序不重要)。从中推导出  $\sin(\alpha + \beta)$  和  $\cos(\alpha + \beta)$  的公式。

5.4. 找到  $\mathbb{R}^2$  中关于直线  $x_1 = -2x_2$  的正交投影矩阵。

**提示:**  $x_1$  轴上的投影矩阵是什么?

5.5. 找到  $A, B: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  的线性变换, 使得  $AB = 0$  但  $BA \neq 0$ .

5.6. 证明定理 5.1, 即证明  $\text{trace}(AB) = \text{trace}(BA)$ .

5.7. 构建一个非零矩阵  $A$  使得  $A^2 = 0$ .

5.8. 找到直线  $y = -2x/3$  的反射矩阵, 并用乘法化至最简。

## 6. 可逆变换与矩阵 同构

### 6.1. 恒等变换与单位矩阵

在所有线性变换中，有一个特殊的变换，即**恒等变换** (identity transformation) (算子)  $I$ ,  $I\mathbf{x} = \mathbf{x}$ ,  $\forall \mathbf{x}$ .

准确地说，存在无数个恒等变换：对于任何向量空间  $V$ ，存在恒等变换  $I = I_V : V \rightarrow V$ ,  $I_V \mathbf{x} = \mathbf{x}$ ,  $\forall \mathbf{x} \in V$ .<sup>13</sup>但是，当不引起混淆时，我们也将使用相同的符号  $I$  来表示所有恒等操作（变换）。只有当我们想强调变换在哪个空间中作用时，我们才会使用符号  $I_V$ .<sup>14</sup>

显然，如果  $I : \mathbb{F}^n \rightarrow \mathbb{F}^n$  是  $\mathbb{F}^n$  中的恒等变换，它的矩阵是

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

(主对角线上的项为 1，其他地方为 0.) 当我们要强调矩阵的大小时，我们使用符号  $I_n$ ；否则，我们只使用  $I$ .

显然，对于任意线性变换  $A$ ，等式

$$AI = A, \quad IA = A$$

成立（只要乘积有定义）。

### 6.2. 可逆变换

**定义** 设  $A : V \rightarrow W$  是一个线性变换。我们说变换  $A$  是**左可逆** (left invertible) 的，如果存在一个线性变换  $B : W \rightarrow V$  使得

$$BA = I \quad (\text{此处 } I = I_V) .$$

变换  $A$  称为**右可逆**的，如果存在一个线性变换  $C : W \rightarrow V$  使得

$$AC = I \quad (\text{此处 } I = I_W) .$$

变换  $B$  和  $C$  分别称为  $A$  的**左逆** (left inverse) 和**右逆** (right inverse)。注意，我们没有假定  $B$  或  $C$  的唯一性，并且通常情况下，左逆和右逆不是唯一的。

<sup>13</sup>译者注：这一表述隐含了，对于一个给定的向量空间  $V$ ，恒等变换  $I_V$  是唯一的，且对应的矩阵必然为方阵。

<sup>14</sup>符号  $E$  经常在线性代数教科书中用来表示单位矩阵，但我更喜欢  $I$ ，因为它也用于算子理论中的相应表示。

**定义** 线性变换  $A: V \rightarrow W$  称为**可逆的** (invertible), 如果它既是右可逆又是左可逆的。

**定理 6.1.** 如果线性变换  $A: V \rightarrow W$  是可逆的, 那么它的左逆  $B$  和右逆  $C$  是唯一且相等的。

**推论**<sup>15</sup> 变换  $A: V \rightarrow W$  是可逆的, 当且仅当存在唯一一个线性变换 (记作  $A^{-1}$ ),  $A^{-1}: W \rightarrow V$ , 使得

$$A^{-1}A = I_V, \quad AA^{-1} = I_W.$$

变换  $A^{-1}$  称为  $A$  的**逆** (inverse)。<sup>16</sup>

**定理 6.1 的证明** 设  $BA = I$  且  $AC = I$ . 那么

$$BAC = B(AC) = BI = B.$$

另外,

$$BAC = (BA)C = IC = C,$$

因此  $B = C$ .

假设对于某个变换  $B_1$  有  $B_1A = I$ , 重复上面的推理, 用  $B_1$  而不是  $B$ , 我们得到  $B_1 = C$ . 因此左逆  $B$  是唯一的。右逆  $C$  的唯一性同理可证。□

**推论** 矩阵被称为**可逆的** (分别地, 左可逆、右可逆), 如果相应的线性变换是可逆的 (分别地, 左可逆、右可逆)。

定理 6.1 断言, 如果存在唯一的矩阵  $A^{-1}$  使得  $A^{-1}A = I, AA^{-1} = I$ , 那么矩阵  $A$  是可逆的。这个矩阵  $A^{-1}$  称为 (惊喜!)  $A$  的**逆**。

### 例子

1. 恒等变换 (矩阵) 是可逆的,  $I^{-1} = I$ ;

2. 旋转变换  $R_\gamma$

$$R_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$$

是可逆的, 并且其逆由  $(R_\gamma)^{-1} = R_{-\gamma}$  给出。这个等式从  $R_\gamma$  的几何描述中就清楚了, 但也可以通过矩阵乘法来验证;

<sup>15</sup>更为常见的是, 这个性质被用于对可逆变换的定义。

<sup>16</sup>译者注: 一旦选定一组基并用矩阵来表示线性变换, 这里从矩阵乘法的定义的层面就已经意味着逆矩阵  $A^{-1}$  是方阵, 否则无法通过矩阵乘法同时得到前面这两个单位矩阵。进一步, 矩阵  $A$  也为方阵。

但如果要严格形式化证明, 即不使用线性变换的矩阵表示, 只针对线性变换本身, 需要依赖可逆变换是双射, 从而维数相等。本书到这里还没有引入这些概念。

3. 列向量  $(1, 1)^T$  是左可逆但不是右可逆的。可能的左逆之一是行向量  $(1/2, 1/2)$ 。要证明这个矩阵不是右可逆的，我们只需注意到它有不只一个左逆。**练习：**描述这个矩阵的所有左逆。

4. 行向量  $(1, 1)$  是右可逆但不是左可逆的。列向量  $(1/2, 1/2)^T$  是一个可能的右逆。

**注记 6.2** 可逆矩阵**必须**是方阵（之后会证明）。而且，如果一个方阵  $A$  既有左逆又有右逆，那么它就是可逆的。所以，只需检验  $AA^{-1} = I$  或  $A^{-1}A = I$  中的一个即可。

重申，这个事实将在后面证明<sup>17</sup>。在此之前，我们不会使用它。我在这里呈现它只是为了阻止学生尝试错误的证明方向。

### 6.2.1. 逆变换的性质

**定理 6.3.**（乘积的逆）如果线性变换  $A$  和  $B$  是可逆的（并且乘积  $AB$  有定义），那么乘积  $AB$  也是可逆的，并且

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

注意它们顺序的变化！

**证明：**直接计算表明：

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

同理

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

□

**注记 6.4** 乘积  $AB$  的可逆性并不意味着因子  $A$  和  $B$  的可逆性（你能想到一个例子吗？）。然而，如果其中一个因子（无论是  $A$  还是  $B$ ）以及乘积  $AB$  都是可逆的，那么第二个因子也是可逆的。

我们将此事实的证明留作练习。

**定理 6.5.**（ $A^T$  的逆）如果矩阵  $A$  是可逆的，那么  $A^T$  也是可逆的，并且

$$(A^T)^{-1} = (A^{-1})^T.$$

**证明：**使用  $(AB)^T = B^T A^T$  我们得到

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I,$$

<sup>17</sup>译者注：见于本书第二章第3节

同理

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I.$$

最后, 如果  $A$  是可逆的, 那么  $A^{-1}$  也是可逆的,  $(A^{-1})^{-1} = A$ .  $\square$

所以, 让我们总结一下逆的三个主要性质:

1. 如果  $A$  可逆, 那么  $A^{-1}$  也可逆,  $(A^{-1})^{-1} = A$ ;
2. 如果  $A$  和  $B$  均可逆且乘积  $AB$  有定义, 那么  $AB$  可逆且  $(AB)^{-1} = B^{-1}A^{-1}$ ;
3. 如果  $A$  可逆, 那么  $A^T$  也可逆且  $(A^T)^{-1} = (A^{-1})^T$ .

### 6.3. 同构 同构空间

**可逆线性变换**  $A: V \rightarrow W$  称为**同构** (isomorphism)。我们这里没有引入任何新东西, 这只是我们已经研究过的对象的另一个名称。

两个向量空间  $V$  和  $W$  称为**同构** (记作  $V \cong W$ ), 如果存在一个同构  $A: V \rightarrow W$ .

同构空间可以视为同一个空间的不同表示, 这意味着所有涉及向量空间运算的性质和构造在同构下都得到保留。

下面的定理说明了这一点。

**定理 6.6.** 设  $A: V \rightarrow W$  是一个同构, 并设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组基。那么向量系统  $A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n$  是  $W$  中的一组基。

我们把这个定理的证明留作练习。

**注记** 在上面的定理中, 我们可以用“线性无关”、“生成”或“线性相关”来替换原本写着“基”的地方——所有这些性质在同构下都被保留。

**注记** 如果  $A$  是同构, 那么  $A^{-1}$  也是同构。因此, 在上面的定理中, 我们可以说  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是基当且仅当  $A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n$  是基。

定理 6.6 的逆命题也成立。

**定理 6.7.** 设  $A: V \rightarrow W$  是线性映射, 并设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  和  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$  分别是  $V$  和  $W$  中的基。如果  $A\mathbf{v}_k = \mathbf{w}_k, k = 1, 2, \dots, n$ , 那么  $A$  是一个同构。

**证明:** 定义逆变换  $A^{-1}$  为  $A^{-1}\mathbf{w}_k = \mathbf{v}_k, k = 1, 2, \dots, n$  (正如我们所知, 线性变换由其在基上的值定义)。  $\square$

**例子**

1.  $A: \mathbb{F}^{n+1} \rightarrow \mathbb{P}_n^{\mathbb{F}}$  ( $\mathbb{P}_n^{\mathbb{F}}$  是  $\sum_{k=0}^n a_k t^k, a_k \in \mathbb{F}$  的形式的  $n$  次多项式集) 定义为

$$A\mathbf{e}_1 = 1, \quad A\mathbf{e}_2 = t, \quad \dots, \quad A\mathbf{e}_n = t^{n-1}, \quad A\mathbf{e}_{n+1} = t^n.$$

根据定理 6.7,  $A$  是同构, 所以  $\mathbb{P}_n^{\mathbb{F}} \cong \mathbb{F}^{n+1}$ .

2. 设  $V$  是一个在  $\mathbb{F}$  上的向量空间,<sup>18</sup>它有一组基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ . 定义变换  $A: \mathbb{F}^n \rightarrow V$  为

$$A\mathbf{e}_k = \mathbf{v}_k, \quad k = 1, 2, \dots, n,$$

其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{F}^n$  的标准基. 根据定理 6.7,  $A$  是同构, 所以  $V \cong \mathbb{F}^n$ .

3.  $M_{2 \times 3}^{\mathbb{F}}$  空间 ( $\mathbb{F}$  中的  $2 \times 3$  矩阵) 同构于  $\mathbb{R}^6$ .

4. 更一般地,  $M_{m \times n}^{\mathbb{F}} \cong \mathbb{F}^{m \cdot n}$ .

## 6.4. 可逆性与方程

**定理 6.8.** <sup>19</sup>设  $A: X \rightarrow Y$  是一个线性变换. 那么  $A$  是可逆的, 当且仅当对于任意右侧  $\mathbf{b} \in Y$ , 方程

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

有唯一解  $\mathbf{x} \in X$ .

**证明:** 假设  $A$  是可逆的. 那么  $\mathbf{x} = A^{-1}\mathbf{b}$  就是方程  $A\mathbf{x} = \mathbf{b}$  的解. 为了证明解是唯一的, 假设对于另一个向量  $\mathbf{x}_1 \in X$ ,

$$A\mathbf{x}_1 = \mathbf{b}.$$

将这个恒等式从左边乘以  $A^{-1}$ , 我们得到

$$A^{-1}A\mathbf{x}_1 = A^{-1}\mathbf{b},$$

因此  $\mathbf{x}_1 = A^{-1}\mathbf{b} = \mathbf{x}$ . 注意, 这里使用了两个恒等式,  $AA^{-1} = I$  和  $A^{-1}A = I$ .

现在假设方程  $A\mathbf{x} = \mathbf{b}$  对于任意  $\mathbf{b} \in Y$  都有唯一解  $\mathbf{x} \in X$ . 让我们用  $\mathbf{y}$  代替  $\mathbf{b}$ . 我们知道, 对于给定的  $\mathbf{y} \in Y$ , 方程

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

有唯一解  $\mathbf{x} \in X$ . 让我们称这个解为  $B(\mathbf{y})$ .

注意,  $B(\mathbf{y})$  对所有  $\mathbf{y} \in Y$  都有定义, 因此我们定义了一个变换  $B: Y \rightarrow X$ .

让我们检验  $B$  是否是线性变换. 我们需要证明  $B(\alpha\mathbf{y}_1 + \beta\mathbf{y}_2) = \alpha B(\mathbf{y}_1) + \beta B(\mathbf{y}_2)$ . 设  $\mathbf{x}_k := B(\mathbf{y}_k)$ ,  $k = 1, 2$ , 即  $A\mathbf{x}_k = \mathbf{y}_k$ ,  $k = 1, 2$ . 那么

$$A(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \alpha A\mathbf{x}_1 + \beta A\mathbf{x}_2 = \alpha\mathbf{y}_1 + \beta\mathbf{y}_2,$$

这意味着

$$B(\alpha\mathbf{y}_1 + \beta\mathbf{y}_2) = \alpha B(\mathbf{y}_1) + \beta B(\mathbf{y}_2).$$

<sup>18</sup>任何带有基的实向量空间都同构于  $\mathbb{R}^n$  (对于某个  $n$ ). 类似地, 任何带有基的复向量空间都同构于  $\mathbb{C}^n$ .

<sup>19</sup>这难道不让你联想到基的定义吗?

最后, 让我们证明  $B$  确实是  $A$  的逆。取  $\mathbf{x} \in X$ , 设  $\mathbf{y} = A\mathbf{x}$ , 所以根据  $B$  的定义, 我们有  $\mathbf{x} = B\mathbf{y}$ . 那么对于所有  $\mathbf{x} \in X$ ,

$$BA\mathbf{x} = B\mathbf{y} = \mathbf{x},$$

所以  $BA = I$ . 类似地, 对于任意  $\mathbf{y} \in Y$ , 设  $\mathbf{x} = B\mathbf{y}$ , 所以  $\mathbf{y} = A\mathbf{x}$ . 那么对于所有  $\mathbf{y} \in Y$ ,

$$AB\mathbf{y} = A\mathbf{x} = \mathbf{y},$$

所以  $AB = I$ . □

回忆基的定义, 我们得到以下定理 6.6 和 6.7 的推论。

**推论 6.9** 一个  $m \times n$  矩阵可逆, 当且仅当它的列在  $\mathbb{F}^m$  中构成一组基。

### 练习

- 6.1. 证明, 如果  $A: V \rightarrow W$  是一个同构 (即一个可逆线性变换), 并且  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组基, 那么  $A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n$  是  $W$  中的一组基。
- 6.2. 找到行向量  $A = (1, 1)$  的所有右逆。由此得出行向量  $A$  不是左可逆的。
- 6.3. 找到列向量  $(1, 2, 3)^T$  的所有左逆。
- 6.4. 列向量  $(1, 2, 3)^T$  是右可逆的吗? 请给出理由。
- 6.5. 找到两个矩阵  $A$  和  $B$ , 使得  $AB$  是可逆的, 但  $A$  和  $B$  都不是可逆的。  
**提示:** 若  $A$  和  $B$  是方阵, 那你不可能找到。  
**注记:** 即使  $AB$  是  $1 \times 1$  矩阵 (标量), 也很容易构造这样的  $A$  和  $B$ . 但是你能得到  $2 \times 2$  矩阵  $AB$  吗?  $3 \times 3$  呢?  $n \times n$  呢?
- 6.6. 假设乘积  $AB$  是可逆的。证明  $A$  是右可逆的,  $B$  是左可逆的。提示: 你可以直接写出右逆和左逆的公式。
- 6.7. 假设  $A$  和  $AB$  都是可逆的 (假设乘积  $AB$  有定义)。证明  $B$  是可逆的。
- 6.8. 设  $A$  是一个  $n \times n$  矩阵。证明如果  $A^2 = \mathbf{0}$ , 则  $A$  不可逆。
- 6.9. 假设  $AB = \mathbf{0}$  对某个非零矩阵  $B$  成立。  $A$  能是可逆的吗? 请给出理由。
- 6.10. 在  $\mathbb{F}^5$  中找到代表如下变换的矩阵  $T_1$  和  $T_2$ :  $T_1$  交换向量  $x$  的坐标  $x_2$  和  $x_4$ , 而  $T_2$  只是将  $x_4$  的  $a$  倍加到坐标  $x_2$  上, 而不改变其他坐标, 即

$$T_1 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_4 \\ x_3 \\ x_2 \\ x_5 \end{pmatrix}, \quad T_2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 + ax_4 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix};$$

这里  $a$  是某个固定的常数。

证明  $T_1$  和  $T_2$  是可逆变换, 并写出它们的逆矩阵。**提示:** 先描述逆变换, 然后再求它的矩阵, 可能比猜测 (或计算)  $T_1, T_2$  的逆矩阵要简单。

- 6.11. 找到  $\mathbb{R}^3$  中绕以向量  $(1, 2, 3)^T$  所在直线为轴  $\alpha$  角旋转的变换的矩阵。我们假设旋转是从向量的尖端看向原点时逆时针旋转的。  
 读者可以将答案表示为几个矩阵的乘积: 你不必用乘法化简得到一个矩阵。
- 6.12. 举出一些  $2 \times 2$  矩阵, 使得:
  - a)  $A + B$  不可逆, 尽管  $A$  和  $B$  都可逆;
  - b)  $A + B$  可逆, 尽管  $A$  和  $B$  都不可逆;
  - c)  $A, B$  和  $A + B$  都可逆。
- 6.13. 设  $A$  是一个可逆的对称矩阵 ( $A^T = A$ )。  $A$  的逆是否对称? 请给出理由。



## 7. 子空间

向量空间  $V$  的一个子空间 (subspace) 是  $V$  的一个非空子集  $V_0 \subset V$ , 它对向量加法和标量乘法是封闭的, 即

1. 如果  $\mathbf{v} \in V_0$ , 则  $\alpha\mathbf{v} \in V_0$  对所有标量  $\alpha$ ;
2. 对于任何  $\mathbf{u}, \mathbf{v} \in V_0$ , 它们的和  $\mathbf{u} + \mathbf{v} \in V_0$ ;

再次, 条件 1 和 2 可以被以下一个条件替换:

$\alpha\mathbf{u} + \beta\mathbf{v} \in V_0$  对所有  $\mathbf{u}, \mathbf{v} \in V_0$ , 以及所有标量  $\alpha, \beta$ .

请注意, 子空间  $V_0 \subset V$  连同从  $V$  中继承的运算 (向量加法和标量乘法), 是一个向量空间。实际上, 由于  $V$  非空, 它至少包含 1 个向量  $\mathbf{v}$ , 并且由于  $\mathbf{0} = 0\mathbf{v}$ , 所以上述条件 1 意味着零向量  $\mathbf{0}$  在  $V$  中。此外, 对于任何  $\mathbf{v} \in V$ , 它的加法逆元  $-\mathbf{v}$  由  $-\mathbf{v} = (-1)\mathbf{v}$  给出, 所以再次根据性质 1,  $-\mathbf{v} \in V$ . 向量空间的其余公理之所以成立, 是因为所有运算都源于向量空间  $V$ . 唯一可能出错的是某个运算的结果不属于  $V_0$ , 但子空间的定义防止了这一点!

现在来看一些例子:

1. 空间  $V$  的平凡 (trivial) 子空间, 即  $V$  本身和  $\{\mathbf{0}\}$  (仅包含零向量的子空间)。注意, 空集  $\emptyset$  不是向量空间, 因为它不包含零向量, 所以它不是子空间。

任何线性变换  $A: V \rightarrow W$  都可以关联以下两个子空间:

2.  $A$  的零空间 (null space) 或核 (kernel), 记作  $\text{Null } A$  或  $\text{Ker } A$ , 由所有满足  $A\mathbf{v} = \mathbf{0}$  的向量  $\mathbf{v} \in V$  组成。

3. 像空间 (range)  $\text{Ran } A$  定义为所有可以表示为  $\mathbf{w} = A\mathbf{v}$  的向量  $\mathbf{w} \in W$  的集合, 其中某个  $\mathbf{v} \in V$ .

如果  $A$  是一个矩阵, 即  $A: \mathbb{F}^m \rightarrow \mathbb{F}^n$ , 那么回忆矩阵-向量乘法的“按列坐标规则”, 我们可以看到任何向量  $\mathbf{w} \in \text{Ran } A$  都可以表示为  $A$  的列的线性组合。这解释了为什么列空间 (表示为  $\text{Col } A$ ) 这个术语经常用来表示矩阵的像空间。因此, 对于矩阵  $A$ , 符号  $\text{Col } A$  通常用于代替  $\text{Ran } A$ .

还有最后一个例子。

4. 给定向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in V$ , 它的线性张成 (linear span) (有时简单称为张成)  $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  是  $V$  中所有可以表示为向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  的线性组合  $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \dots + \alpha_r\mathbf{v}_r$  的向量的集合。符号  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  也用于代替  $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ .

容易检验, 在所有这些例子中, 我们确实得到了子空间。我们把检验部分交给读者, 作为练习。其中一些陈述将在本书后面证明。

### 练习

- 7.1. 设  $X$  和  $Y$  是向量空间  $V$  的子空间。证明  $X \cap Y$  是  $V$  的子空间。
- 7.2. 设  $V$  是一个向量空间。对于  $X, Y \subset V$ , “和空间”  $X + Y$  是所有可以表示为  $\mathbf{v} = \mathbf{x} + \mathbf{y}$ ,  $\mathbf{x} \in X, \mathbf{y} \in Y$  的向量的集合。证明如果  $X$  和  $Y$  是  $V$  的子空间, 那么  $X + Y$  也是  $V$  的子空间。
- 7.3. 设  $X$  是向量空间  $V$  的子空间, 设  $\mathbf{v} \in V, \mathbf{v} \notin X$ . 证明如果  $\mathbf{x} \in X$ , 则  $\mathbf{x} + \mathbf{v} \notin X$ .
- 7.4. 设  $X$  和  $Y$  是向量空间  $V$  的子空间。利用上一道练习, 证明  $X \cup Y$  是子空间当且仅当  $X \subset Y$  或  $Y \subset X$ .
- 7.5. 包含所有上三角矩阵 ( $a_{j,k} = 0 \forall j > k$ ) 和所有对称矩阵 ( $A = A^T$ ) 的  $4 \times 4$  矩阵空间中最小的子空间是什么? 包含在这两个子空间中的最大子空间是什么?

## 8. 应用于计算机图形学

在本节中, 我们将介绍一些线性代数在计算机图形学中的应用。我们不会深入细节, 只是解释一些思想。特别是我们将解释为什么对三维图像的操作会简化为  $4 \times 4$  矩阵的乘法。

### 8.1. 二维操作

$x - y$  平面 (更准确地说, 平面上的一个矩形) 是表示计算机显示器的一个很好的模型。显示器上的任何对象都表示为**像素 (pixel)** 的集合, 每个像素被分配一个特定的颜色。每个像素的位置由其列和行确定, 它们充当平面上的  $x$  和  $y$  坐标。所以, 一个平面上具有  $x - y$  坐标的矩形是计算机屏幕的良好模型: 而图形对象只是点的集合。

**注记** 有两种类型的图形对象: 位图对象, 其中描述了对象的每个像素, 以及矢量对象, 其中我们只描述**关键点 (critical points)**, 然后图形引擎将它们连接起来以重建对象。照片是位图对象的一个好例子: 它的每个像素都被描述。位图对象可能包含很多点, 所以处理位图需要大量的计算能力。任何使用过位图处理程序 (如 Adobe Photoshop) 的人都知道, 你需要一台相当强大的计算机, 即使是现代和强大的计算机, 操作也可能需要一些时间。

这就是为什么出现在计算机屏幕上的大多数对象都是矢量对象的原因: 计算机只需要记住关键点。例如, 要描述一个多边形, 你只需要给出其顶点的坐标, 以及哪些顶点连接到哪个顶点。当然, 并非屏幕上的所有对象都可以表示为多边形, 有些对象, 如字母, 具有平滑弯曲的边界。但是, 存在标准方法可以允许我们通过一组点绘制平滑曲线, 例如贝塞尔曲线, 在 PostScript 和 Adobe PDF (以及许多其他格式) 中使用。

无论如何, 这是另一本书的主题, 我们在这里不讨论它。对我们来说, 一个图形对象将是一组点 (无论是线框模型 (wireframe model) 还是位图), 我们想展示如

何对这些对象执行一些操作。

最简单的变换是**平移** (translation) (移动) (shift), 其中每个点 (向量)  $\mathbf{v}$  被平移  $\mathbf{a}$ , 即向量  $\mathbf{v}$  被替换为  $\mathbf{v} + \mathbf{a}$  (表示  $\mathbf{v} \mapsto \mathbf{v} + \mathbf{a}$ )。计算机非常适合计算向量加法, 因此平移很容易实现。注意, 平移不是线性变换 (如果  $\mathbf{a} \neq \mathbf{0}$ ): 虽然它保留了直线, 但它不保留  $\mathbf{0}$ 。

计算机图形学中使用的所有其他变换都是线性的。第一个想到的就是旋转。绕原点  $\mathbf{0}$  的  $\gamma$  角旋转由我们上面讨论过的旋转矩阵  $R_\gamma$  给出,

$$R_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}.$$

如果我们想绕一个点  $\mathbf{a}$  旋转, 我们首先需要平移图像  $-\mathbf{a}$ , 将点  $\mathbf{a}$  移动到  $\mathbf{0}$ , 然后绕  $\mathbf{0}$  旋转 (乘以  $R_\gamma$ ), 然后将所有内容平移回  $\mathbf{a}$ 。

另一个非常有用的变换是**缩放** (scaling), 由矩阵

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

给出,  $a, b \geq 0$ . 如果  $a = b$  它是**均匀缩放** (uniform scaling), 它放大 (缩小) 对象, 保持其形状。如果  $a \neq b$  则  $x$  和  $y$  坐标缩放不同; 对象变得“更高”或“更宽”。

另一个经常使用的变换是**反射** (reflection): 例如矩阵

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

定义了关于  $x$  轴的反射。

我们将在本书后面证明,  $\mathbb{R}^2$  中的任何线性变换都可以表示为缩放、旋转和反射的复合。然而, 有时考虑一些不同的变换, 如**剪切变换** (shear transformation), 由矩阵

$$\begin{pmatrix} 1 & \tan \phi \\ 0 & 1 \end{pmatrix}$$

给出。这个变换使得所有对象倾斜, 水平线保持水平, 但垂直线变成与水平线成  $\phi$  角的倾斜线。

## 8.2. 三维图形

三维图形更复杂。首先我们需要能够操作三维物体, 然后需要将其表示在二维平面 (显示器) 上。

三维物体的操作非常直接, 我们有相同的基本变换: 平移、平面反射、缩放、

旋转。这些变换的矩阵与其二维对应物的矩阵非常相似。例如，矩阵

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}, \quad \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

分别代表了关于  $x-y$  平面的反射、缩放和绕  $z$  轴的旋转。

请注意，上述旋转本质上是二维变换，它不改变  $z$  坐标。类似地，可以为绕  $x$  轴和绕  $y$  轴的其他两组基本旋转写出矩阵。后面将表明，任意轴上的旋转可以表示为基本旋转的复合。因此，我们知道了应该如何操作三维物体。

现在让我们讨论如何将三维物体表示在二维平面上。最简单的方法是将其投影到平面，比如  $x-y$  平面上。要执行这种投影，只需将  $z$  坐标替换为 0，这个投影 (projection) 的矩阵是

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

这种方法通常用于技术插图。旋转一个物体并对其进行投影相当于从不同的点看它。然而，这种方法没有给出非常逼真的图像，因为它没有考虑透视，即远处的物体看起来更小的事实。

为了获得更逼真的图像，我们需要使用所谓的**透视投影** (perspective projection)。要定义透视投影，我们需要选择一个点 (投影中心或焦点) 和一个要投影到的平面。然后， $\mathbb{R}^3$  中的每个点被投影到一个平面上的点，使得该点、它的像以及**投影中心** (center of the projection) 位于同一条线上，见图0.2。

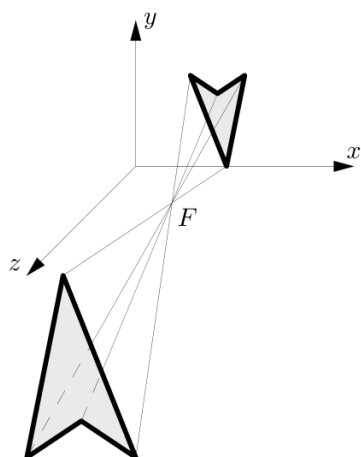


图 0.2 透视投影到  $x-y$  平面： $F$  是投影中心 (焦点)

这正是相机的工作方式，也是我们眼睛工作方式的一个合理初步近似。

让我们得到投影的公式。假设焦点是  $(0, 0, d)^T$ ，并且我们投影到  $x - y$  平面，见图0.3a)。考虑一个点  $\mathbf{v} = (x, y, z)^T$ ，让  $\mathbf{v}^* = (x^*, y^*, 0)^T$  是它的投影。分析相似三角形，见图0.3b)，我们得到

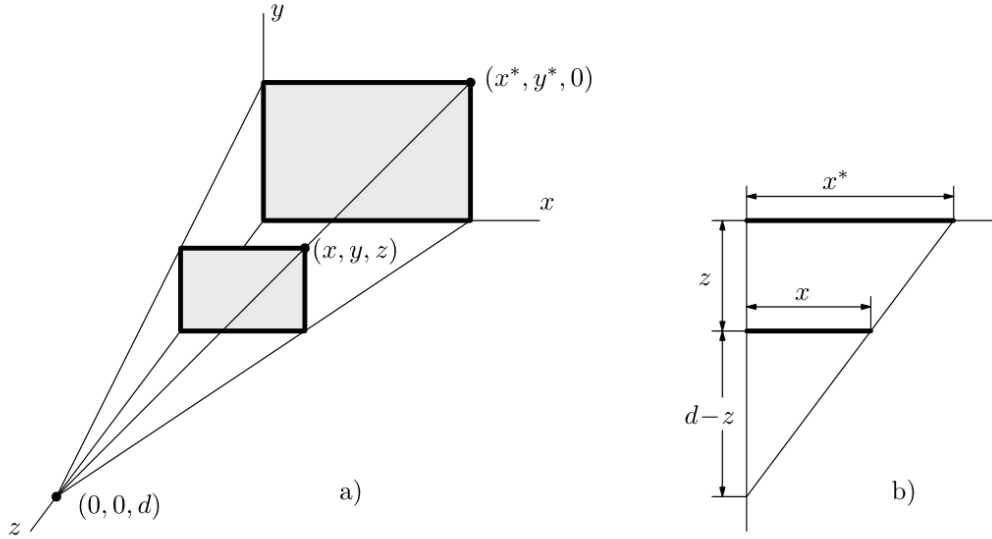


图 0.3 图 3. 找到点  $(x, y, z)^T$  的透视投影的坐标  $x^*, y^*$

$$\frac{x^*}{d} = \frac{x}{d - z},$$

所以

$$x^* = \frac{xd}{d - z} = \frac{x}{1 - z/d},$$

类似地

$$y^* = \frac{y}{1 - z/d}.$$

注意，这个公式在  $z > d$  和  $z < 0$  时也有效：你可以画出相应的相似三角形来验证它。

因此，透视投影将点  $(x, y, z)^T$  映射到点  $(\frac{x}{1-z/d}, \frac{y}{1-z/d}, 0)^T$ 。

这个变换肯定不是线性的（由于分母中的  $z$ ）。然而，通过引入所谓的**齐次坐标** (homogeneous coordinates)，仍然可以将其表示为线性变换。

在齐次坐标中， $\mathbb{R}^3$  中的每个点都由 4 个坐标表示，最后一个（第四个）坐标起到了缩放系数的作用。因此，要从  $\mathbf{v} = (x, y, z)^T$  的齐次坐标  $\mathbf{v} = (x_1, x_2, x_3, x_4)^T$  得到其通常的三维坐标，需要将所有项除以最后一个坐标  $x_4$ ，然后取前 3 个坐标<sup>20</sup>（如果  $x_4 = 0$ ，则此方法不适用，因此我们假设  $x_4 \neq 0$  的情况对应于无穷远点）。

<sup>20</sup>如果我们对齐次坐标下一个在  $\mathbb{R}^2$  中的点乘上一个非零标量，我们没有改变这个点。换言之，在齐次坐标下，一个在  $\mathbb{R}^3$  中的点被表示为在  $\mathbb{R}^4$  中有一行为 0 的点。

因为在齐次坐标中, 向量  $\mathbf{v}^*$  可以被表示为  $x, y, 0, 1 - z/d)^T$ , 因此, 在齐次坐标中, 透视投影是一个线性变换:

$$\begin{pmatrix} x \\ y \\ 0 \\ 1 - z/d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

请注意, 在齐次坐标中, 平移也是一个线性变换:

$$\begin{pmatrix} x + a_1 \\ y + a_2 \\ z + a_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

但是, 如果投影中心不是  $(0, 0, d)^T$  这样的点, 而是任意点  $(d_1, d_2, d_3)^T$  呢? 那么我们首先需要应用  $-(d_1, d_2, 0)^T$  的平移来将中心移到  $(0, 0, d_3)^T$ , 同时保持  $x - y$  平面不变, 应用投影, 然后通过  $(d_1, d_2, 0)^T$  的平移将所有内容移回。类似地, 如果投影平面不是  $x - y$  平面, 我们通过使用旋转和平移将其移到  $x - y$  平面, 等等。

所有这些操作只是  $4 \times 4$  矩阵的乘法。这就解释了为什么现代显卡将  $4 \times 4$  矩阵运算嵌入到处理器中。

当然, 这里我们只触及了三维图形背后的数学, 还有更多内容等待我们学习。例如, 如何确定物体的哪些部分可见, 哪些部分被隐藏, 如何制作逼真的照明、阴影等等。

### 练习

8.1.  $\mathbb{R}^3$  中齐次坐标为  $(10, 20, 30, 5)^T$  的向量是什么?

8.2. 证明  $\gamma$  角的旋转可以表示为两次剪切-缩放变换的复合

$$T_1 = \begin{pmatrix} 1 & 0 \\ \sin \gamma & \cos \gamma \end{pmatrix}, \quad T_2 = \begin{pmatrix} \sec \gamma & -\tan \gamma \\ 0 & 1 \end{pmatrix}.$$

应该以什么顺序进行变换?

8.3. 一个 2 维向量乘以一个任意的  $2 \times 2$  矩阵通常需要 4 次乘法。

假设  $2 \times 1000$  矩阵  $D$  包含  $\mathbb{R}^2$  中 1000 个点的坐标。使用两个任意  $2 \times 2$  矩阵  $A$  和  $B$  来变换这些点需要多少次乘法? 比较两种可能性,  $A(BD)$  和  $(AB)D$ 。

8.4. 写出符合要求的  $4 \times 4$  矩阵, 来进行到  $x - y$  平面的透视投影, 其投影中心为  $(d_1, d_2, d_3)^T$ 。

8.5. 变换  $T$  在  $\mathbb{R}^3$  中是  $x - y$  平面中直线  $y = 2x + 3$  绕  $\gamma$  角的旋转。写出与此变换对应的  $4 \times 4$  矩阵。你可以将结果表示为矩阵的乘积。

## 第二章 线性方程组

### 1. 线性方程组的不同表示

关于线性方程组，或者简而言之**线性系统** (linear system)，存在几种观点。第一种，朴素的观点是，它仅仅是含  $n$  个未知数  $x_1, x_2, \dots, x_n$  的  $m$  个线性方程的集合：

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \cdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m. \end{cases}$$

求解该系统是指找到所有满足这  $m$  个方程的  $n$  元数组  $x_1, x_2, \dots, x_n$ .

如果我们记  $\mathbf{x} := (x_1, x_2, \dots, x_n)^T \in \mathbb{F}^n$ ,  $\mathbf{b} = (b_1, b_2, \dots, b_m)^T \in \mathbb{F}^m$ ，以及

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix},$$

那么上述线性系统可以用**矩阵形式** (matrix form) (作为**矩阵-向量方程** (matrix-vector equation)) 写成

$$A\mathbf{x} = \mathbf{b}.$$

求解上述方程是指找到所有满足  $A\mathbf{x} = \mathbf{b}$  的向量  $\mathbf{x} \in \mathbb{F}^n$ .

最后，回忆矩阵-向量乘法的“按列坐标规则”，我们可以将系统写成一个**向量方程** (vector equation)：

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

其中  $\mathbf{a}_k$  是矩阵  $A$  的第  $k$  列， $\mathbf{a}_k = (a_{1,k}, a_{2,k}, \dots, a_{m,k})^T$ ,  $k = 1, 2, \dots, n$ .

注意，这三个例子本质上只是同一个数学对象的不同表示。

在解释如何求解线性系统之前，让我们注意到，无论我们如何称呼未知数，例如  $x_k, y_k$ ，或其他名称，这都不重要。因此，所有求解系统所需的信息都包含在矩阵  $A$  中，该矩阵称为系统的**系数矩阵** (coefficient matrix)，以及向量（处在右侧的）

b. 因此，我们所需的所有信息都包含在以下矩阵中：

$$\left( \begin{array}{cccc|c} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & b_m \end{array} \right).$$

该矩阵是通过将列  $\mathbf{b}$  连接到矩阵  $A$  上形成的。这个矩阵称为系统的**增广矩阵** (augmented matrix)。我们通常会放一条垂直线来分隔  $A$  和  $\mathbf{b}$ ，以区分增广矩阵和系数矩阵。

## 2. 线性方程组的求解 阶梯形与简化阶梯形

线性系统可以通过**高斯-若尔当消元法** (Gauss-Jordan elimination) (有时称为**行约简** (row reduction)) 求解。通过对系统增广矩阵的行 (即方程) 执行运算，我们将它简化为一种简单的形式，即所谓的**阶梯形** (echelon form)。当系统处于阶梯形时，我们可以轻松地写出解。

### 2.1. 行运算

我们使用的行运算有三种类型：

1. 行交换：交换矩阵的任意两行；
2. 缩放：用一个非零标量  $a$  乘以某一行；
3. 行替换：用第  $j$  行的常数倍加上第  $k$  行来整体替换第  $k$  行；其余行保持不变。

可以清楚地看出，运算 1 和 2 不会改变系统的解集；它们本质上不改变系统。

至于运算 3，可以很容易地看出它不会丢失解。也就是说，设一个“新”系统是通过类型 3 的行运算从“旧”系统中得到的，那么“旧”系统的任何解也都是“新”系统的解。

为了证明我们没有得到任何额外的东西，即“新”系统的任何解也是“旧”系统的解，我们只需注意到类型 3 的行运算是**可逆**的，也就是说，“旧”系统也可以通过应用类型 3 的行运算从“新”系统中获得。(你能说出是哪一种吗?)

#### 2.1.1. 行运算与初等矩阵的乘法

还有另一种更“高级”的解释来说明为什么上述行运算是合法的。也就是说，每个行运算都相当于从左边乘以一个特殊的初等矩阵。



也即，乘以矩阵

$$\begin{matrix} & & j & & k & & \\ & & & & & & \\ & & & & & & \\ j & \left( \begin{array}{cccccc} 1 & & \vdots & & \vdots & & \mathbf{0} \\ & \ddots & \vdots & & \vdots & & \\ \cdots & \cdots & 0 & \cdots & 1 & \cdots \\ & & \vdots & \ddots & \vdots & & \\ k & \cdots & \cdots & 1 & \cdots & 0 & \cdots \\ & & & & & \ddots & \\ \mathbf{0} & & & & & & 1 \end{array} \right) \end{matrix}$$

其中第  $j$  行和第  $k$  行可以看作是对单位矩阵  $I$  的第  $j$  行和第  $k$  行的交换。乘以这个矩阵

$$\begin{matrix} & & & & & & \\ & & & & & & \\ & & & & & & \\ k & \left( \begin{array}{cccccc} 1 & 0 & \vdots & & \mathbf{0} \\ 0 & \ddots & \vdots & & \\ & & 1 & 0 & & \\ \cdots & \cdots & 0 & a & 0 & \cdots \\ & & 0 & 1 & & \\ & & & & \ddots & \vdots \\ \mathbf{0} & & & & \cdots & 1 \end{array} \right) \end{matrix}$$

其中第  $k$  行乘以  $a$ . 最后，乘以这个矩阵

$$\begin{matrix} & & & & & & \\ & & & & & & \\ j & \left( \begin{array}{cccccc} 1 & & \vdots & & \vdots & & \mathbf{0} \\ & \ddots & \vdots & & \vdots & & \\ \cdots & \cdots & 1 & \cdots & 0 & \cdots \\ & & \vdots & \ddots & \vdots & & \\ k & \cdots & \cdots & a & \cdots & 1 \\ & & & & \ddots & \\ \mathbf{0} & & & & & & 1 \end{array} \right) \end{matrix}$$

其中第  $k$  行加上第  $j$  行的  $a$  倍，而其他行不变。

如果要看到这些初等矩阵的乘法确实是按照预期执行的，你可以简单地看看它们如何作用于向量（列）。

注意，所有这些矩阵都是可逆的（与行运算的可逆性进行比较）。第一个矩阵的逆是它本身。要得到第二个矩阵的逆，只需将  $a$  替换为  $1/a$ . 最后，第三个矩阵的逆可以通过将  $a$  替换为  $-a$  来获得。要看到逆确实是这样获得的，我们（再次）

可以简单地验证它们如何作用于列。

因此，对系统  $A\mathbf{x} = \mathbf{b}$  的增广矩阵执行行运算，相当于将系统（从左边）乘以一个特殊的初等矩阵  $E$ 。将等式  $A\mathbf{x} = \mathbf{b}$  从左边乘以  $E$ ，我们得到

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

的任何解，也是

$$EA\mathbf{x} = E\mathbf{b}$$

的解。将这个方程从左边乘以  $E^{-1}$ ，我们得到它的任何解也是

$$E^{-1}EA\mathbf{x} = E^{-1}E\mathbf{b}$$

的解，也就是原始方程  $A\mathbf{x} = \mathbf{b}$ 。所以，行运算不改变系统的解集。

## 2.2. 行约简

行约简的主步骤包括三个子步骤：

1. 找到矩阵中最左边的非零列；
2. 通过使用类型 1 的行运算（即在必要时进行行交换），确保该列的第一个（最上面的）项非零。这个项将被称为主元项 (pivot entry) 或简称为主元 (pivot)；
3. 通过从第 2, 3, ...,  $m$  行减去第一行的适当倍数来“消去” (Kill) 主元下的所有非零项（即让其为 0）。

我们将主步骤应用于一个矩阵，然后将第一行单独处理，并对第 2, ...,  $m$  行应用主步骤，然后对第 3, ...,  $m$  行应用主步骤，等等。

需要记住的一点是，在将一行的适当倍数减去此行所有下面的行（步骤 3）之后，我们就需要将该行抛诸脑后，不再操作它，甚至不与其他行交换。

在应用主步骤有限次（最多  $m$  次）之后，我们得到所谓的矩阵的**阶梯形**。

### 2.2.1. 行约简的一个例子

让我们考虑以下线性系统：

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ 3x_1 + 2x_2 + x_3 = 7 \\ 2x_1 + x_2 + 2x_3 = 1 \end{cases}$$

系统的增广矩阵是

$$\left( \begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 2 & 1 & 7 \\ 2 & 1 & 2 & 1 \end{array} \right)$$

将第二行减去第一行的 3 倍，并将第三行减去第一行的 2 倍，我们得到：

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 2 & 1 & 7 \\ 2 & 1 & 2 & 1 \end{array}\right) \xrightarrow[R_3-2R_1]{R_2-3R_1} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -4 & -8 & 4 \\ 0 & -3 & -4 & -1 \end{array}\right)$$

将第二行乘以  $-1/4$  得到：

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

将第三行加上第二行的 3 倍得到：

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & -3 & -4 & -1 \end{array}\right) \xrightarrow{R_3+3R_2} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 2 & -4 \end{array}\right)$$

现在我们可以使用所谓的**反代人** (back substitution) 来求解系统。即，从最后一行 (方程) 我们得到  $x_3 = -2$ 。然后从第二个方程我们得到

$$x_2 = -1 - 2x_3 = -1 - 2(-2) = 3,$$

最后，从第一行 (方程)

$$x_1 = 1 - 2x_2 - 3x_3 = 1 - 6 + 6 = 1.$$

所以，解是

$$\begin{cases} x_1 = 1 \\ x_2 = 3 \\ x_3 = -2, \end{cases}$$

或者向量形式

$$\mathbf{x} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}.$$

或  $\mathbf{x} = (1, 3, -2)^T$ 。我们可以通过乘以系数矩阵  $A$  来验算解。

换一种思路，与其使用反代人，不如从下到上进行行约简，消去系数矩阵主对角线以上的所有项。我们从把最后一行乘以  $1/2$  开始，其余的都很直观：

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -2 \end{array}\right) \xrightarrow[R_2-2R_3]{R_1-3R_3} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{array}\right) \xrightarrow{R_1-2R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{array}\right)$$

我们只需从简化阶梯形矩阵中读出解  $\mathbf{x} = (1, 3, -2)^T$ .

我们把阐述从下到上阶段的行约简算法留给读者作练习。

### 2.3. 阶梯形

一个矩阵被称为**阶梯形** (echelon form), 如果它满足以下两个条件:

1. 所有零行 (zero rows) (即所有项都等于 0 的行), 如果存在的话, 都位于所有非零行的下方。

对于非零行, 称最左边的非零项为**前导项** (leading entry)。那么阶梯形的第二性质可以表述如下:

2. 对于任何非零行, 其前导项严格位于前一行前导项的右侧。

阶梯形中的每一行的前导项也称为**主元项**, 或简称为**主元**, 因为这些项正是我们在行约简中使用的主元。

我们上面得到的例子中的一个特殊情况是所谓的**三角形** (triangular) 形式。在该形式中, 系数矩阵是方阵 ( $n \times n$ ), 其主对角线上的所有项都非零, 并且主对角线下的所有项都为零。右侧, 即增广矩阵的最右边一列, 可以是任意的。

在行约简的向后阶段也完成之后, 我们就得到了所谓的矩阵的**简化阶梯形** (reduced echelon form)。上例中系数矩阵等于  $I$ , 这是简化阶梯形的一个特例。

一般定义如下: 我们说一个矩阵处于**简化阶梯形**, 如果它处于阶梯形并且

3. 所有主元项都等于 1;

4. 主元上方的所有项都为 0。注意, 由于阶梯形的原因, 主元下方的所有项也为 0。

为了从阶梯形得到简化阶梯形, 我们从下往上, 从右往左操作, 使用行替换来消去主元上方的所有项。

简化阶梯形的一个例子是系数矩阵等于  $I$  的系统。在这种情况下, 只需从简化阶梯形中读出解。通常情况下, 从简化阶梯形中观察到解也很容易。例如, 设系统 (增广矩阵) 的简化阶梯形是

$$\left( \begin{array}{ccccc|c} \boxed{1} & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & \boxed{1} & 5 & 0 & 2 \\ 0 & 0 & 0 & 0 & \boxed{1} & 3 \end{array} \right)$$

这里我们框出了主元。这个想法是, 将与没有主元的列对应的变量 (所谓的**自由**

变量 (free variables)) 移到右侧, 这样我们就可以直接把解写出来。

$$\begin{cases} x_1 = 1 - 2x_2 \\ x_2 \text{ 是自由变量} \\ x_3 = 2 - 5x_4 \\ x_4 \text{ 是自由变量} \\ x_5 = 3 \end{cases}$$

或者, 在向量形式下:

$$\mathbf{x} = \begin{pmatrix} 1 - 2x_2 \\ x_2 \\ 2 - 5x_4 \\ x_4 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{pmatrix}, \quad x_2, x_4 \in \mathbb{F}$$

也可以通过反代入从阶梯形得到解: 其思想是从下往上操作, 将所有自由变量移到右侧。

### 练习

2.1. 将以下方程组分别写成矩阵形式和向量方程形式; 而后求解这些系统, 并以向量形式写出解。

$$\text{a) } \begin{cases} x_1 + 2x_2 - x_3 = -1 \\ 2x_1 + 2x_2 + x_3 = 1 \\ 3x_1 + 5x_2 - 2x_3 = -1 \end{cases}$$

$$\text{b) } \begin{cases} x_1 - 2x_2 - x_3 = 1 \\ 2x_1 - 3x_2 + x_3 = 6 \\ 3x_1 - 5x_2 = 7 \\ x_1 + 5x_3 = 9 \end{cases}$$

$$\text{c) } \begin{cases} x_1 + 2x_2 + 2x_4 = 6 \\ 3x_1 + 5x_2 - x_3 + 6x_4 = 17 \\ 2x_1 + 4x_2 + x_3 + 2x_4 = 12 \\ 2x_1 - 7x_3 + 11x_4 = 7 \end{cases}$$

$$\text{d) } \begin{cases} x_1 - 4x_2 - x_3 + x_4 = 3 \\ 2x_1 - 8x_2 + x_3 - 4x_4 = 9 \\ -x_1 + 4x_2 - 2x_3 + 5x_4 = -6 \end{cases}$$

$$\text{e) } \begin{cases} x_1 + 2x_2 - x_3 + 3x_4 = 2 \\ 2x_1 + 4x_2 - x_3 + 6x_4 = 5 \\ x_2 + 2x_4 = 3 \end{cases}$$

$$\text{f) } \begin{cases} 2x_1 - 2x_2 - x_3 + 6x_4 - 2x_5 = 1 \\ x_1 - x_2 + x_3 + 2x_4 - x_5 = 2 \\ 4x_1 - 4x_2 + 5x_3 + 7x_4 - x_5 = 6 \end{cases}$$

$$\text{g)} \begin{cases} 3x_1 - x_2 + x_3 - x_4 + 2x_5 = 5 \\ x_1 - x_2 - x_3 - 2x_4 - x_5 = 2 \\ 5x_1 - 2x_2 + x_3 - 3x_4 + 3x_5 = 10 \\ 2x_1 - x_2 - 2x_4 + x_5 = 5 \end{cases}$$

## 2.2. 找到向量方程

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + x_3 \mathbf{v}_3 = \mathbf{0}$$

的所有解，其中  $\mathbf{v}_1 = (1, 1, 0)^T$ ,  $\mathbf{v}_2 = (0, 1, 1)^T$  和  $\mathbf{v}_3 = (1, 0, 1)^T$ . 你能从中得出关于向量系统  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  线性无关 (或相关) 的什么结论?

## 3. 主元的分析

关于解的存在性和唯一性的所有问题都可以通过分析增广矩阵在阶梯形 (或简化阶梯形) 中的主元来回答。

首先，让我们研究一下方程  $A\mathbf{x} = \mathbf{b}$  **不一致 (inconsistent)** (即它无解) 的情况。如果我们稍微思考一下，答案立即得出：

当增广矩阵的阶梯形中最后一个列有一个主元时，线性方程组才是不一致的 (无解)，即增广矩阵的阶梯形包含一行  $(0 \ 0 \ \dots \ 0 \ | \ b)$ ，其中  $b \neq 0$ .

实际上，这样的行对应于方程  $0x_1 + 0x_2 + \dots + 0x_n = b \neq 0$ ，它确实无解。如果不含这样的行，我们只需将其化为简化阶梯形，然后从中读出解。

现在，还有三个陈述。所有这些陈述都只涉及**系数矩阵**，而不是系统的增广矩阵。

1. 解 (如果它存在) 是唯一的，当且仅当没有自由变量，也就是说，当且仅当系数矩阵的阶梯形在每一列都有一个主元；

2. 方程  $A\mathbf{x} = \mathbf{b}$  对于所有右侧  $\mathbf{b}$  都是一**致 (consistent)** 的，当且仅当系数矩阵的阶梯形在每一行都有一个主元。

3. 方程  $A\mathbf{x} = \mathbf{b}$  对于任意右侧  $\mathbf{b}$  都有**唯一解 (unique solution)**，当且仅当系数矩阵  $A$  的阶梯形在每一列和每一行都有一个主元。

第一个陈述是显然的，因为自由变量的存在导致了所有的不唯一性。我应该只强调这个陈述**并不说明**关于存在性的任何信息。

第二个陈述稍微复杂一些。如果我们有一个系数矩阵  $A$ ，其阶梯形在每一行都有一个主元，那么我们不能在**增广**矩阵的最后一列中有一个主元，所以系统总是有一致的解，无论右侧  $\mathbf{b}$  是什么。

让我们证明，如果我们有一个系数矩阵  $A$  的阶梯形中的零行，那么我们可以选择一个右侧  $\mathbf{b}$  使得系统  $A\mathbf{x} = \mathbf{b}$  不一致。设  $A_e$  是系数矩阵  $A$  的阶梯形。那么

$$A_e = EA,$$

其中  $E$  是对应于行运算的初等矩阵的乘积,  $E = E_N \dots E_2 E_1$ . 如果  $A_e$  有一个零行, 那么最后一行也是零. 因此, 如果我们取  $\mathbf{b}_e = (0, \dots, 0, 1)^T$  (所有项都是 0, 除了最后一个是 1), 那么方程

$$A_e \mathbf{x} = \mathbf{b}_e$$

无解. 从左边乘以  $E^{-1}$  并回忆  $E^{-1}A_e = A$ , 我们得到方程

$$A\mathbf{x} = E^{-1}\mathbf{b}_e$$

无解。

最后, 陈述 3 可以直接由陈述 1 和 2 得出。

上述对主元的分析带来了几个非常重要的推论. 我们在观察中使用的主要事实是:

在阶梯形中, 每一行和每一列最多有一个主元 (也可以没有主元) .

### 3.1. 关于线性无关和基的推论 维数

关于向量系统在  $\mathbb{F}^n$  中是否是基、线性无关或生成系统的问题, 都可以通过行约简轻松回答。

**命题 3.1** 假设我们有一个向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}^n$ , 并且令  $A = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$  是以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  为列的  $n \times m$  矩阵. 那么

1. 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  线性无关, 当且仅当  $A$  的阶梯形在每一列都有一个主元;
2. 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  是  $\mathbb{F}^n$  中的完备 (生成) 集, 当且仅当  $A$  的阶梯形在每一行都有一个主元;
3. 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  是  $\mathbb{F}^n$  中的基, 当且仅当  $A$  的阶梯形在每一列和每一行都有一个主元。

**证明:** 向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}^n$  线性无关, 当且仅当方程

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \dots + x_m \mathbf{v}_m = \mathbf{0}$$

只有唯一的 (平凡) 解  $x_1 = x_2 = \dots = x_m = 0$ , 或者等价地说, 方程  $A\mathbf{x} = \mathbf{0}$  有唯一解  $\mathbf{x} = \mathbf{0}$ . 根据上面的陈述 1, 这发生在当且仅当矩阵的主元在每一列时。

类似地, 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}^n$  是  $\mathbb{F}^n$  中的完备集, 当且仅当方程

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \dots + x_m \mathbf{v}_m = \mathbf{b}$$

对任何右侧  $\mathbf{b} \in \mathbb{F}^n$  都有解. 根据上面的陈述 2, 这发生在当且仅当  $A$  的矩阵的阶梯形在每一行都有一个主元。

最后, 系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}^n$  是  $\mathbb{F}^n$  中的基, 当且仅当方程

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_m \mathbf{v}_m = \mathbf{b}$$

对任何右侧  $\mathbf{b} \in \mathbb{F}^n$  都有唯一解。根据陈述 3, 这发生在当且仅当  $A$  的阶梯形在每一列和每一行都有一个主元时。□

**命题 3.2**  $\mathbb{F}^n$  中的任何线性无系统不能包含超过  $n$  个向量。

**证明:** 设系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}^n$  是线性无关的, 并且设  $A = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$  是以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  为列的  $n \times m$  矩阵。根据命题 3.1,  $A$  的阶梯形必须在每一列都有一个主元, 这在  $m > n$  时是不可能的 (主元数量不能超过行数)。□

**命题 3.3** 向量空间  $V$  中的任何两组基具有相同数量的向量。

**证明:** 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  和  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  是  $V$  中的两组不同的基。不失一般性, 我们假设  $n \leq m$ . 考虑一个同构  $A: \mathbb{F}^n \rightarrow V$ , 定义为

$$A\mathbf{e}_k = \mathbf{v}_k, \quad k = 1, 2, \dots, n,$$

其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{R}^n$  的标准基。

由于  $A^{-1}$  也是一个同构, 系统

$$A^{-1}\mathbf{w}_1, A^{-1}\mathbf{w}_2, \dots, A^{-1}\mathbf{w}_m$$

是一组基 (见第一章定理 6.6)。所以它是线性无关的, 根据命题 3.2,  $m \leq n$ . 结合假设  $n \leq m$ , 我们得到  $m = n$ . □

上述命题的一个特例是以下命题。

**命题 3.4**  $\mathbb{F}^n$  中的任何基必须恰好有  $n$  个向量。

**证明:** 这个事实直接源于前面的命题, 但也有一个直接的证明。设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  是  $\mathbb{F}^n$  中的一组基, 令  $A$  为以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  为列的  $n \times m$  矩阵。系统是基的事实意味着方程

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

对任何 (所有可能的) 右侧  $\mathbf{b}$  都有唯一解。存在性意味着在 (简化) 阶梯形矩阵的每一行中都有一个主元, 因此主元数量恰好是  $n$ . 唯一性意味着在系数矩阵 (的阶梯形) 的每一列中都有一个主元, 所以

$$m = \text{列数} = \text{主元数} = n.$$

□

**命题 3.5**  $\mathbb{F}^n$  中的任何生成集必须至少有  $n$  个向量。



**证明：** 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  是  $\mathbb{F}^n$  中的完备集，令  $A$  为以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  为列的  $n \times m$  矩阵。命题 3.1 的陈述 2 暗示  $A$  的阶梯形在每一行都有一个主元。由于主元数量不能超过列的数量，所以  $n \leq m$ .  $\square$

### 3.2. 可逆矩阵的推论

**命题 3.6** 一个矩阵  $A$  是可逆的，当且仅当它的阶梯形在每一列和每一行都有一个主元。

**证明：** 正如我们在本节开头所讨论的，方程  $A\mathbf{x} = \mathbf{b}$  对于任何右侧  $\mathbf{b}$  都有唯一解，当且仅当  $A$  的阶梯形在每一行和每一列都有一个主元。但是，我们知道（见第一章定理 6.8），矩阵  $A$  是可逆的，当且仅当方程  $A\mathbf{x} = \mathbf{b}$  对右侧  $\mathbf{b}$  的每一项有唯一解。

也存在一个备选的证明。我们知道，一个矩阵是可逆的，当且仅当它的列（见第一章第 6.4 节的推论 6.9）构成一组基。前面的命题 3.4 表明了，这发生在当且仅当在每一行和每一列都有一个主元时。  $\square$

上述命题立即蕴含了以下推论。

**推论 3.7** 可逆矩阵必须是方阵 ( $n \times n$ )。

**命题 3.8** 如果一个  $n \times n$  方阵  $A$  是左可逆的，或者它是右可逆的，那么它就是可逆的。换句话说，要核验方阵  $A$  的可逆性，只需核验  $AA^{-1} = I$  或  $A^{-1}A = I$  其中一个条件即可。

注意，这个命题仅适用于方阵！

**证明：** 我们知道，矩阵  $A$  是可逆的，当且仅当方程  $A\mathbf{x} = \mathbf{b}$  对于任何右侧  $\mathbf{b}$  都有唯一解。这发生在当且仅当  $A$  的阶梯形在每一行和每一列都有一个主元。

如果矩阵  $A$  是左可逆的，那么方程  $A\mathbf{x} = \mathbf{0}$  有唯一解  $\mathbf{x} = \mathbf{0}$ 。实际上，如果  $B$  是  $A$  的左逆（即  $BA = I$ ），并且  $\mathbf{x}$  满足

$$A\mathbf{x} = \mathbf{0},$$

那么从左边将这个恒等式乘以  $B$ ，我们得到  $\mathbf{x} = \mathbf{0}$ ，所以解是唯一的。因此， $A$  的阶梯形在每一列都有一个主元（没有自由变量）。如果矩阵  $A$  是方阵 ( $n \times n$ )，那么阶梯形也在每一行都有一个主元（ $n$  个主元，而一行最多有一个主元），所以矩阵是可逆的。

如果矩阵  $A$  是右可逆的，并且  $C$  是它的右逆 ( $AC = I$ )，那么对于  $\mathbf{x} = C\mathbf{b}$ ， $\mathbf{b} \in \mathbb{F}^n$ ，

$$A\mathbf{x} = AC\mathbf{b} = I\mathbf{b} = \mathbf{b}.$$

因此, 对于任何右侧  $\mathbf{b}$ , 方程  $A\mathbf{x} = \mathbf{b}$  都有解  $\mathbf{x} = C\mathbf{b}$ . 因此,  $A$  的阶梯形在每一行都有一个主元。如果  $A$  是方阵, 那么它也在每一列都有一个主元。所以  $A$  是可逆的。  $\square$

### 练习

3.1. 对于  $b$  的哪个值, 系统

$$\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 4 \\ b \end{pmatrix}$$

有解? 对于该值  $b$ , 找到系统的通解。

3.2. 确定向量

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

是否线性无关或相关。

这四个向量是否张成  $\mathbb{R}^4$ ? (换句话说, 它们是生成系统吗?) 对于  $\mathbb{C}^4$  呢?

3.3. 确定以下向量系统是否是  $\mathbb{R}^3$  的基:

a)  $(1, 2, -1)^T, (1, 0, 2)^T, (2, 1, 1)^T$ ;

b)  $(-1, 3, 2)^T, (-3, 1, 3)^T, (2, 10, 2)^T$ ;

c)  $(67, 13, -47)^T, (\pi, -7.84, 0)^T, (3, 0, 0)^T$ .

哪个系统是  $\mathbb{C}^3$  的基?

3.4. 多项式  $t^3 + 2t, t^2 + t + 1, t^3 + 5$  是否生成 (张成)  $\mathbb{P}_3$ ? 给出你的理由。

3.5.  $\mathbb{F}^4$  中的 5 个向量可能线性无关吗? 给出你的理由。

3.6. 证明或证伪: 如果一个  $(n \times n)$  方阵  $A$  的列是线性无关的, 那么  $A^2 = AA$  的列也是线性无关的。

3.7. 证明或证伪: 如果一个  $(n \times n)$  方阵  $A$  的列是线性无关的, 那么  $A^3 = AAA$  的行也是线性无关的。

3.8. 证明, 如果方程  $A\mathbf{x} = \mathbf{0}$  只有唯一解 (即, 如果  $A$  的阶梯形在每一列都有一个主元), 那么  $A$  是左可逆的。提示: 想想初等矩阵可能会有帮助。

注记: 这可能是一个非常难的问题, 因为它需要对主题有深入的理解。但是, 当你理解了该做什么之后, 问题就变得几乎显然了。

3.9. 一个矩阵的简化阶梯形是唯一的吗? 给出你的结论和理由。

也就是说, 假设通过执行某些行运算 (不一定遵循任何算法) 我们得到了一个简化阶梯形矩阵。那么我们总是得到相同的矩阵, 还是有可能得到不同的矩阵? 请注意, 我们只允许执行行运算, “列运算” 是被禁止的。

提示: 如果以可逆矩阵开始, 会发生什么? 此外, 主元是否总是在相同的列中, 还是这取决于你执行的行运算? 如果你能在不诉诸行运算的情况下知道主元列是什么, 那么主元列的位置就不依赖于它们。

## 4. 通过行约简求 $A^{-1}$

正如我们在上面讨论的, 可逆矩阵必须是方阵, 并且其阶梯形在每一行和每一列都必须有主元。故而, 可逆矩阵的简化阶梯形是单位矩阵  $I$ . 因此,

任何可逆矩阵都可通过行约简 (即通过行运算) 化为单位矩阵。

下面有一个简单的算法，可以让我们来找到一个  $n \times n$  矩阵的逆：

1. 通过在  $A$  的右侧拼接  $n \times n$  单位矩阵来形成一个  $n \times 2n$  的增广矩阵  $(A|I)$ ；
2. 对增广矩阵执行行运算，将  $A$  转化为单位矩阵  $I$ ；
3. 原本对应  $I$  的地方将被自动转化为  $A^{-1}$ ；
4. 如果通过行运算无法将  $A$  转化为  $n \times n$  单位矩阵，则  $A$  是不可逆的。

对于以上算法，我们有几个解释。第一个，朴素的解释：我们知道（对于可逆矩阵  $A$ ），向量  $A^{-1}\mathbf{b}$  是方程  $A\mathbf{x} = \mathbf{b}$  的解。所以，要找到  $A^{-1}$  的第  $k$  列，我们需要找到  $A\mathbf{x} = \mathbf{e}_k$  的解，其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{R}^n$  的标准基。上述算法只是同时求解方程

$$A\mathbf{x} = \mathbf{e}_k, \quad k = 1, 2, \dots, n.$$

我们也提供另一个看起来更“高级”的解释：正如我们上面讨论的，每个行运算都可以通过左乘一个初等矩阵来实现。设  $E_1, E_2, \dots, E_N$  是对应于我们执行的行运算的初等矩阵，令  $E = E_N \dots E_2 E_1$  为它们的乘积<sup>21</sup>。我们知道行运算会将  $A$  转化为单位矩阵，即  $EA = I$ ，所以  $E = A^{-1}$ 。那么，相同的行运算就将增广矩阵  $(A|I)$  转化为  $(EA|E) = (I|A^{-1})$ 。

这个“高级”解释利用初等矩阵表述了一个重要的命题，该命题将在以后经常被使用。

**定理 4.1.** 任何可逆矩阵都可以表示为初等矩阵的乘积。

**证明：**正如我们在上一段中所讨论的， $A^{-1} = E_N \dots E_2 E_1$ ，所以

$$A = (A^{-1})^{-1} = (E_N \dots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \dots E_N^{-1}$$

（初等矩阵的逆也是初等矩阵）。□

**一个例子** 假设我们想找到矩阵

$$\begin{pmatrix} 1 & 4 & -2 \\ -2 & -7 & 7 \\ 3 & 11 & -6 \end{pmatrix}$$

的逆。将其与单位矩阵拼接为增广矩阵，并进行行约简，我们得到

$$\left( \begin{array}{ccc|ccc} 1 & 4 & -2 & 1 & 0 & 0 \\ -2 & -7 & 7 & 0 & 1 & 0 \\ 3 & 11 & -6 & 0 & 0 & 1 \end{array} \right) \xrightarrow[R_3-3R_1]{R_2+2R_1} \left( \begin{array}{ccc|ccc} 1 & 4 & -2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 2 & 1 & 0 \\ 0 & -1 & 0 & -3 & 0 & 1 \end{array} \right) \xrightarrow{R_3+R_2}$$

<sup>21</sup>虽然在这里并不重要，但请注意，如果行运算  $E_1$  最先得到执行，那么  $E_1$  必须是乘积中最右边的项。

$$\left( \begin{array}{ccc|ccc} 1 & 4 & -2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 2 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right) \xrightarrow{R_1 \times 3} \left( \begin{array}{ccc|ccc} 3 & 12 & -6 & 3 & 0 & 0 \\ 0 & 1 & 3 & 2 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right) \xrightarrow{\begin{array}{l} R_1 + 2R_3 \\ R_2 - 2R_3 \end{array}}$$

在最后一步行运算中，我们将第一行乘以 3 以避免在向后阶段的行约简中出现分数。再次进行行约简，我们得到

$$\left( \begin{array}{ccc|ccc} 3 & 12 & 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 3 & 0 & -1 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right) \xrightarrow{R_1 - 12R_2} \left( \begin{array}{ccc|ccc} 3 & 0 & 0 & -35 & 2 & 14 \\ 0 & 1 & 0 & 3 & 0 & -1 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right)$$

将第一行和最后一行除以 3，我们得到逆矩阵

$$\begin{pmatrix} -35/3 & 2/3 & 14/3 \\ 3 & 0 & -1 \\ -1/3 & 1/3 & 1/3 \end{pmatrix}$$

### 练习

4.1. 找到以下矩阵的逆：

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

写出所有步骤。

## 5. 维数 有限维空间

**定义** 向量空间  $V$  的**维数**  $\dim V$  是基中向量的数量。

对于仅由零向量  $\mathbf{0}$  组成的向量空间，我们设  $\dim V = 0$ . 如果  $V$  不存在（有限）基，我们设  $\dim V = \infty$ .

如果  $\dim V$  是有限的，我们称空间  $V$  为**有限维的** (finite-dimensional)；否则，我们称其为**无限维的** (infinite-dimensional)。

命题 3.3 表明维数是良好定义的，即它不依赖于基的选择。

第一章的命题 2.8 表明，有限维向量空间中的任何有限生成集都包含一组基。这直接蕴含了以下命题。

**命题 5.1** 向量空间  $V$  是有限维的，当且仅当它有一个有限生成集。

假设我们有一个有限维向量空间中的向量系统，并且我们想核验它是否是基（或者它是否线性无关，或者是否完备），最简单的方法可能就是使用同构  $A: V \rightarrow \mathbb{R}^n, n = \dim V$  将问题转移到  $\mathbb{R}^n$ ，在  $\mathbb{R}^n$  中，所有这些问题都可以通过行约简（研究主元）来回答。

请注意, 如果  $\dim V = n$ , 那么总存在一个同构  $A: V \rightarrow \mathbb{R}^n$ . 实际上, 如果  $\dim V = n$ , 则存在一组基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$ , 并且可以定义一个同构  $A: V \rightarrow \mathbb{R}^n$  为

$$A\mathbf{v}_k = \mathbf{e}_k, \quad k = 1, 2, \dots, n.$$

例如, 让我们给出命题 3.2 和 3.5 的两个推论如下:

**命题 5.2** 有限维向量空间  $V$  中的任何线性无系统不能包含超过  $\dim V$  个向量。

**证明:** 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in V$  是  $V$  中线性无系统, 令  $A: V \rightarrow \mathbb{R}^n$  为一个同构。那么  $A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_m$  是  $\mathbb{R}^n$  中的线性无系统, 根据命题 3.2,  $m \leq n$ .  $\square$

**命题 5.3** 有限维向量空间  $V$  中的任何生成系统必须至少有  $\dim V$  个向量。

**证明:** 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in V$  是  $V$  中的完备系统, 令  $A: V \rightarrow \mathbb{R}^n$  为一个同构。那么  $A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_m$  是  $\mathbb{R}^n$  中的完备系统, 根据命题 3.5,  $m \geq n$ .  $\square$

## 5.1. 将线性无系统补全为基

以下陈述将在后面扮演重要角色。

**命题 5.4 (补全为基)** 有限维空间中线性无系统的向量可以补全为基, 即, 给定有限维向量空间  $V$  中的线性无关向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ , 可以找到向量  $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n$  使得系统中向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组基。

**证明:** 设  $\dim V = n$ . 选择一个不属于  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  的向量并称之为  $\mathbf{v}_{r+1}$  (由于系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  不是生成的, 总能做到这一点)。根据第一章练习 2.5, 系统  $\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}$  是线性无关的 (注意在这种情况下  $r < n$ , 根据命题 5.2)。用新向量  $\mathbf{v}_{r+2}$  重复这个过程, 依此类推。

直到得到一个生成系统后, 这个过程才会停止。注意, 这个过程不能无限进行, 因为向量空间  $V$  中的线性无关向量系统不能包含超过  $n = \dim V$  个向量。  $\square$

## 5.2. 有限维空间的子空间

**定理 5.5.** 设  $V$  是  $W$  的一个子空间, 且  $\dim W < \infty$ . 那么  $V$  是有限维的, 并且  $\dim V \leq \dim W$ .

此外, 如果  $\dim V = \dim W$ , 则  $V = W$  (这里我们仍然假设  $V$  是  $W$  的子空间)。

**注记** 这个定理看起来像是一个平凡的论断, 就像是命题 5.2 的一个简单的推论。但是, 我们只能在已知  $V$  的基的情况下才能应用命题 5.2。现在, 我们只知道

$W$  的基, 而无法知道这组基中有多少向量属于  $V$ ; 实际上, 很容易构造一个例子, 其中  $W$  的基向量中没有属于  $V$ .

**定理 5.5 的证明** 如果  $V = \{0\}$ , 那么定理是平凡的, 所以我们假设不为此情况。

我们想在  $V$  中找到一组基。取一个非零向量  $\mathbf{v}_1 \in V$ . 如果  $V = \text{span}\{\mathbf{v}_1\}$ , 我们就找到了基 (由单个向量  $\mathbf{v}_1$  组成)。

否则, 我们继续归纳。假设我们已经构造了  $r$  个线性无关向量  $\mathbf{v}_1, \dots, \mathbf{v}_r \in V$ . 如果  $V = \text{span}\{\mathbf{v}_k : 1 \leq k \leq r\}$ , 那么我们已经找到了  $V$  中的一组基。如果不是, 则存在一个向量  $\mathbf{v}_{r+1} \in V, \mathbf{v}_{r+1} \notin \text{span}\{\mathbf{v}_k : 1 \leq k \leq r\}$ . 根据第一章练习 2.5, 系统  $\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}$  是线性无关的。□

### 练习

#### 5.1. 判断正误:

- 任何由有限集生成的向量空间都有基;
- 任何向量空间都有 (有限) 基;
- 一个向量空间不能有多组基;
- 如果一个向量空间有有限基, 那么所有基中的向量数量是相同的;
- $\mathbb{P}_n$  的维数是  $n$ ;
- $M_{m \times n}$  的维数是  $m + n$ ;
- 如果向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  生成 (张成) 向量空间  $V$ , 那么  $V$  中的每个向量都可以唯一地表示为向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的线性组合;
- 任何有限维空间的子空间都是有限维的;
- 如果向量空间  $V$  的维数是  $n$ , 那么  $V$  只有一个零维子空间和一个  $n$  维子空间。

5.2. 证明: 如果  $V$  是一个  $n$  维向量空间, 那么  $V$  中的向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是线性无关的, 当且仅当它张成  $V$ .

5.3. 证明  $V$  中的线性无关向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是基当且仅当  $n = \dim V$ .

5.4. (重温一个旧问题, 第一章题 2.6: 现在这个问题应该很容易回答了) 是否可能, 向量  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  是线性相关的, 而向量  $\mathbf{w}_1 = \mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_2 = \mathbf{v}_2 + \mathbf{v}_3$  和  $\mathbf{w}_3 = \mathbf{v}_3 + \mathbf{v}_1$  是线性无关的?

**提示:** 向量空间  $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$  可以是什么维数?

5.5. 设  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  是  $V$  中的一组基。证明  $\mathbf{u} + \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}, \mathbf{w}$  也是  $V$  中的一组基。

5.6. 在  $\mathbb{R}^5$  空间中考虑向量  $\mathbf{v}_1 = (2, -1, 1, 5, -3)^T, \mathbf{v}_2 = (3, -2, 0, 0, 0)^T, \mathbf{v}_3 = (1, 1, 50, -921, 0)^T$ .

a) 证明这些向量是线性无关的。

b) 将此向量系统补全为基。

(如果你先做了 b) 部分, 你可以完全不进行计算。)

## 6. 线性系统的通解

在本节中, 我们将讨论线性系统的通解 (所有解, 即解集) 的结构。

我们称一个系统  $A\mathbf{x} = \mathbf{b}$  为**齐次** (homogeneous) 的, 如果右侧  $\mathbf{b} = \mathbf{0}$ , 即齐次系统是  $A\mathbf{x} = \mathbf{0}$  的形式。

并且对于每个系统

$$A\mathbf{x} = \mathbf{b},$$

我们可以关联一个齐次系统，只需令  $\mathbf{b}$  为  $\mathbf{0}$ 。

**定理 6.1.** (线性方程的通解) 设向量  $\mathbf{x}_1$  满足方程  $A\mathbf{x} = \mathbf{b}$ ，并设  $H$  是相关齐次系统

$$A\mathbf{x} = \mathbf{0}$$

的所有解的集合。那么集合

$$\{\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_h : \mathbf{x}_h \in H\}$$

是方程  $A\mathbf{x} = \mathbf{b}$  的所有解的集合。

换句话说，这个定理可以陈述为：

$$\boxed{A\mathbf{x} = \mathbf{b} \text{ 的通解}} = \boxed{A\mathbf{x} = \mathbf{b} \text{ 的一个特解}} + \boxed{A\mathbf{x} = \mathbf{0} \text{ 的通解.}}$$

**证明：** 固定一个满足  $A\mathbf{x}_1 = \mathbf{b}$  的向量  $\mathbf{x}_1$ 。设向量  $\mathbf{x}_h$  满足  $A\mathbf{x}_h = \mathbf{0}$ 。那么对于  $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_h$ ，我们有

$$A\mathbf{x} = A(\mathbf{x}_1 + \mathbf{x}_h) = A\mathbf{x}_1 + A\mathbf{x}_h = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

所以任何形式为

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_h, \quad \mathbf{x}_h \in H$$

的  $\mathbf{x}$  都是  $A\mathbf{x} = \mathbf{b}$  的解。

现在设  $\mathbf{x}$  满足  $A\mathbf{x} = \mathbf{b}$ 。那么对于  $\mathbf{x}_h := \mathbf{x} - \mathbf{x}_1$ ，我们得到

$$A\mathbf{x}_h = A(\mathbf{x} - \mathbf{x}_1) = A\mathbf{x} - A\mathbf{x}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

所以  $\mathbf{x}_h \in H$ 。因此， $A\mathbf{x} = \mathbf{b}$  的任何解都可以表示为  $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_h$  的形式，其中某个  $\mathbf{x}_h \in H$ 。□

这个定理的威力在于它的普适性。它适用于所有线性方程，而不必假设向量空间是有限维的。你将在微分方程、积分方程、偏微分方程等领域遇到这个定理。

除了展示解集结构之外，这个定理还允许我们将唯一性与存在性的研究分开。也就是说，为了研究唯一性，我们只需要分析齐次方程  $A\mathbf{x} = \mathbf{0}$  的唯一性，而它总是有一个解。

在本节中，我们立即可以应用它：这个定理让我们能核验系统  $A\mathbf{x} = \mathbf{b}$  的解。

例如, 考虑系统

$$\begin{pmatrix} 2 & 3 & 1 & 4 & -9 \\ 1 & 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & 2 & -5 \\ 2 & 2 & 2 & 3 & -8 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 17 \\ 6 \\ 8 \\ 14 \end{pmatrix}$$

通过行约简, 可以找到这个系统的解

$$(6.1) \quad \mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \quad x_3, x_5 \in \mathbb{F}$$

参数  $x_3, x_5$  在这里可以记作其他字母, 例如  $t$  和  $s$ ; 我们在这里保留符号  $x_3$  和  $x_5$  仅仅是为了提醒我们, 参数来自相应的自由变量。

现在, 假设我们只是得到了这个解, 并且我们想核验它是否正确。当然, 我们可以重复行运算, 但这太耗时了。而且, 如果解是通过某种非标准行运算方法得到的, 它可能看起来与我们从行约简得到的结果不同。例如, 通解

$$(6.2) \quad \mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}, \quad s, t \in \mathbb{F}$$

给出与 (6.1) 相同的解集 (你能说明为什么吗?); 这里我们只是将 (6.1) 中的最后一个向量替换为它自身加上第二个向量的和。所以, 这个式子看起来与我们从行约简得到的解不同, 但它仍然是正确的。

核验 (6.1) 和 (6.2) 是否给出正确的解的最简单方法是核验第一个向量  $(3, 1, 0, 2, 0)^T$  是否满足方程  $A\mathbf{x} = \mathbf{b}$ , 而其他两个向量 (带参数的向量,  $x_3$  和  $x_5$  或  $s$  和  $t$  在它们前面) 应该满足相应的齐次方程  $A\mathbf{x} = \mathbf{0}$ 。

如果核验通过了, 我们就能确保由 (6.1) 或 (6.2) 定义的任何向量  $\mathbf{x}$  确实是一个解。

请注意, 这种核验解的方法并不能保证 (6.1) (或 (6.2)) 给出所有解。例如, 如果我们只是不知怎么就遗漏了  $x_3$  相关的项, 上述方法仍然能正常进行。

因此, 我们如何保证我们没有遗漏任何自由变量, 并且在 (6.1) 中不需要额外的项?



这时浮现在我们脑海中的办法，就是再次计算主元的数量。在这个例子中，如果进行行运算，主元的数量是 3。所以确实应该有 2 个自由变量，而且看起来我们没有遗漏 (6.1) 中的任何内容。

为了能够证明这一点成立，我们需要引入基本子空间和矩阵秩的新概念。我还得指出，在上面的例子中，我们不必进行所有行运算，就能核验出只有 2 个自由变量，并且公式 (6.1) 和 (6.2) 都给出了正确的通解。

### 练习

#### 6.1. 判断正误：

- a) 任何线性方程组都有至少一个解；
- b) 任何线性方程组最多有一个解；
- c) 任何齐次线性方程组至少有一个解；
- d) 含  $n$  个未知数  $n$  个方程的线性方程组至少有一个解；
- e) 含  $n$  个未知数  $n$  个方程的线性方程组最多有一个解；
- f) 如果与给定线性方程组相对应的齐次方程组有解，则给定方程组有解；
- g) 如果含  $n$  个未知数  $n$  个方程的齐次线性方程组的系数矩阵是可逆的，那么该系统没有非零解；
- h) 任何含  $n$  个未知数  $m$  个方程的线性方程组的解集是  $\mathbb{R}^n$  中的一个子空间；
- i) 任何含  $n$  个未知数  $m$  个方程的齐次线性方程组的解集是  $\mathbb{R}^n$  中的一个子空间。

#### 6.2. 找到一个 $2 \times 3$ 系统（含有 3 个未知数的 2 个方程），使得其通解具有形式

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad s \in \mathbb{R}.$$

## 7. 矩阵的基本子空间 秩

正如我们在第一章第 7 节中所讨论的，任何线性变换  $A: V \rightarrow W$  都可以关联两个子空间，即它的核或零空间

$$\text{Ker } A = \text{Null } A := \{\mathbf{v} \in V : A\mathbf{v} = \mathbf{0}\} \subset V,$$

以及它的像空间

$$\text{Ran } A = \{\mathbf{w} \in W : \mathbf{w} = A\mathbf{v} \text{ 对于某个 } \mathbf{v} \in V\} \subset W.$$

换句话说，核  $\text{Ker } A$  是齐次方程  $A\mathbf{x} = \mathbf{0}$  的解集，而像空间  $\text{Ran } A$  正是方程  $A\mathbf{x} = \mathbf{b}$  有解的所有右侧  $\mathbf{b} \in W$  的集合。

如果  $A$  是一个  $m \times n$  矩阵，即从  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的一个线性变换，那么回忆矩阵乘法的“按列坐标规则”，我们可以看到任何向量  $\mathbf{w} \in \text{Ran } A$  都可以表示为  $A$  的列的线性组合。这解释了为什么列空间（表示为  $\text{Col } A$ ）这个术语经常用来表示矩阵的像空间。因此，对于矩阵  $A$ ，符号  $\text{Col } A$  通常用于代替  $\text{Ran } A$ 。

如果  $A$  是一个矩阵，那么除了  $\text{Ran } A$  和  $\text{Ker } A$  之外，我们还可以考虑转置矩阵

$A^T$  的像空间和核空间。通常**行空间**(row space)用于表示  $\text{Ran } A^T$ , 而**左零空间**(left null space)用于表示  $\text{Ker } A^T$  (但通常没有特殊的符号来表示)。

四个子空间  $\text{Ran } A, \text{Ker } A, \text{Ran } A^T, \text{Ker } A^T$  称为矩阵  $A$  的**基本子空间**(fundamental subspaces)。在本节中, 我们将研究它们的维数之间重要的关系。

我们需要以下定义, 这是线性代数的基本概念之一。

**定义** 给定一个线性变换 (矩阵)  $A$ , 它的**秩**,  $\text{rank } A$ , <sup>22</sup>是它的像空间的维数:

$$\text{rank } A := \dim \text{Ran } A$$

### 7.1. 计算基本子空间和秩

要计算矩阵的基本子空间和秩, 就需进行行约简。也就是说, 设  $A$  是矩阵, 且  $A_e$  是其阶梯形:

1. 原始矩阵  $A$  的**主元列** (即行约简后将有主元的列) 给出了  $\text{Ran } A$  的一组基 (它是众多可能的基之一)。

2. 阶梯形  $A_e$  的**主元行**给出了行空间的基。当然, 也可以简单地转置矩阵, 然后进行行约简。但是, 如果我们已经得到了  $A$  的阶梯形, 例如在计算  $\text{Ran } A$  时, 那么我们就能自然得到  $\text{Ran } A^T$ 。

3. 要找到零空间  $\text{Ker } A$  的基, 需求解齐次方程  $A\mathbf{x} = \mathbf{0}$ : 具体细节将从下面的例子中看出。

**例子** 考虑矩阵

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

进行行运算我们得到阶梯形

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

<sup>22</sup>译者注: 或表示为  $r(A)$ , 后者在国内教科书中更常见。

(这里主元已被框出)。因此，**原始矩阵**的第 1 列和第 3 列，即列向量

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 1 \\ 3 \\ -1 \end{pmatrix}$$

给出了  $\text{Ran } A$  的一组基。我们也自然得到了行空间  $\text{Ran } A^T$  的基： $A$  的**阶梯形**的第一行和第二行，即向量

$$\begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ -3 \\ -3 \\ -1 \end{pmatrix}$$

(这里我们将向量垂直放置。放在这里的向量到底是作为列，还是作为行，这真的只是一个约定问题。我们将其垂直放置的原因是，尽管我们称  $\text{Ran } A^T$  为**行空间**，但我们将其定义为  $A^T$  的列空间)。

为了计算零空间  $\text{Ker } A$  的基，我们需要求解方程  $A\mathbf{x} = \mathbf{0}$ 。为此计算  $A$  的**简化**阶梯形，在这个例子中得到的是

$$\begin{pmatrix} \boxed{1} & 1 & 0 & 0 & 1/3 \\ 0 & 0 & \boxed{1} & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

注意，在求解齐次方程  $A\mathbf{x} = \mathbf{0}$  时，不必写出整个增广矩阵，只处理系数矩阵就足够了。实际上，在这种情况下，增广矩阵的最后一列是零列，它在行运算下不会改变。所以，我们可以不实际写出它以节省笔墨，只在心中记住有这一列即可。在心中记住它时，我们可以从上面的简化阶梯形中读出解：

$$\begin{cases} x_1 = -x_2 - \frac{1}{3}x_5, \\ x_2 \text{ 是自由变量} \\ x_3 = -x_4 - \frac{1}{3}x_5, \\ x_4 \text{ 是自由变量} \\ x_5 \text{ 是自由变量} \end{cases}$$

或者，在向量形式下：

$$(7.1) \quad \mathbf{x} = \begin{pmatrix} -x_2 - \frac{1}{3}x_5 \\ x_2 \\ -x_4 - \frac{1}{3}x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -1/3 \\ 0 \\ -1/3 \\ 0 \\ 1 \end{pmatrix}.$$

向量在每个自由变量处，即在我们的例子中，向量

$$\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -1/3 \\ 0 \\ -1/3 \\ 0 \\ 1 \end{pmatrix}$$

构成  $\text{Ker } A$  的一组基。

不幸的是，想找到  $\text{Ker } A^T$  的基没有捷径，必须求解方程  $A^T \mathbf{x} = \mathbf{0}$ 。知道  $A$  的阶梯形在此处无济于事。

## 7.2. 基本子空间基的计算方法的解释

那么，为什么上述方法确实给出了基本子空间的基呢？

### 7.2.1. 零空间 $\text{Ker } A$

零空间  $\text{Ker } A$  的情况可能是最简单的：由于我们求解了方程  $A\mathbf{x} = \mathbf{0}$ ，即找到了所有解，那么  $\text{Ker } A$  中的任何向量都是我们得到的向量的线性组合。因此，我们得到的向量构成了  $\text{Ker } A$  中的一个生成系统。要看到系统是线性无关的，我们可以让每个向量乘以相应的自由变量并将所有向量相加，见 (7.1)。那么对于每个自由变量  $x_k$ ，结果向量的第  $k$  个分量恰好是  $x_k$ ，再次见 (7.1)，所以这个向量（线性组合）是  $\mathbf{0}$  的唯一方式是当所有自由变量都为  $0$  时。

### 7.2.2. 列空间 $\text{Ran } A$

让我们现在解释为什么用于找到列空间  $\text{Ran } A$  的基的方法有效。首先，注意到  $A$  的简化阶梯形， $A_{re}$  的**主元列**构成了  $\text{Ran } A_{re}$  的基（不是原始矩阵的列空间，而是其简化阶梯形列空间的基！）。由于行运算只是可逆矩阵的左乘，它们不改变线性无关性。因此，**原始矩阵**  $A$  的主元列是线性无关的。

让我们现在证明  $A$  的主元列张成  $A$  的列空间。设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是  $A$  的主元列，设  $\mathbf{v}$  是  $A$  的任意一列。我们想证明  $\mathbf{v}$  可以表示为主元列  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  的线性

组合,

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r.$$

简化阶梯形  $A_{re}$  是通过左乘

$$A_{re} = EA$$

从  $A$  得到的, 其中  $E$  是初等矩阵的乘积, 所以  $E$  是可逆矩阵。向量  $E\mathbf{v}_1, E\mathbf{v}_2, \dots, E\mathbf{v}_r$  是  $A_{re}$  的主元列, 而  $A$  的列  $\mathbf{v}$  被变换为  $A_{re}$  的列  $E\mathbf{v}$ 。由于  $A_{re}$  的主元列构成了  $\text{Ran } A_{re}$  的基, 向量  $E\mathbf{v}$  可以表示为线性组合

$$E\mathbf{v} = \alpha_1 E\mathbf{v}_1 + \alpha_2 E\mathbf{v}_2 + \cdots + \alpha_r E\mathbf{v}_r.$$

将这个等式两边同时左乘  $E^{-1}$ , 我们得到表示

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r,$$

因此  $A$  的主元列确实张成了  $\text{Ran } A$ 。

### 7.2.3. 行空间 $\text{Ran } A^T$

可以很容易地看出,  $A$  的阶梯形  $A_e$  的主元行是线性无关的。实际上, 设  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$  是  $A_e$  的转置 (因为我们公认总是将向量垂直放置) 的主元行。假设

$$\alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 + \cdots + \alpha_r \mathbf{w}_r = \mathbf{0}.$$

考虑  $\mathbf{w}_1$  的第一个非零项。由于对于所有其他向量  $\mathbf{w}_2, \mathbf{w}_3, \dots, \mathbf{w}_r$ , 相应的项等于 0 (根据阶梯形的定义), 我们可以得出  $\alpha_1 = 0$ 。所以我们可以消去这一项。

现在考虑  $\mathbf{w}_2$  的第一个非零项。向量  $\mathbf{w}_3, \dots, \mathbf{w}_r$  的相应项为 0, 所以  $\alpha_2 = 0$ 。重复这个过程, 我们得到  $\alpha_k = 0 \forall k = 1, 2, \dots, r$ 。

要证明向量  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$  张成了行空间, 我们注意到“行运算不改变行空间”。这可以从直接分析行运算得到, 但我们在这里提供一种更正式的方式来演示这个事实。

对于变换  $A$  和集合  $X$ , 我们用  $A(X)$  来表示所有可以表示为  $y = A(x), x \in X$  的元素  $y$  的集合, 即

$$A(X) := \{y = A(x) : x \in X\}.$$

如果  $A$  是一个  $m \times n$  矩阵,  $A_e$  是它的阶梯形,  $A_e$  是通过左乘

$$A_e = EA$$

得到的, 其中  $E$  是一个  $m \times m$  可逆矩阵 (对应初等矩阵的乘积)。那么

$$\text{Ran } A_e^T = \text{Ran}(A^T E^T) = A^T(\text{Ran } E^T) = A^T(\mathbb{R}^m) = \text{Ran } A^T,$$

所以确实  $\text{Ran } A^T = \text{Ran } A_e^T$ 。

### 7.3. 秩定理 基本子空间的维数

在许多应用中, 我们需要找到列空间或零空间的基。例如, 正如前面所示, 求解齐次方程  $A\mathbf{x} = \mathbf{0}$  等价于找到零空间  $\text{Ker } A$  的基。找到列空间的基意味着通过移除不必要的 (列) 向量, 来从生成集中提取基。

然而, 基本子空间计算方法最重要的应用是得到它们维数之间的关系。

**定理 7.1.** (秩定理) 对于矩阵  $A$ ,

$$\text{rank } A = \text{rank } A^T.$$

这个定理通常表述为:

矩阵的**列秩**等于它的**行秩**。

这个定理的证明是显然的, 因为  $\text{Ran } A$  和  $\text{Ran } A^T$  的维数都等于  $A$  的阶梯形中的主元数量。

以下定理为我们提供了基本子空间维数之间重要的关系。它通常也称为秩定理。

**定理 7.2.** 设  $A$  是一个  $m \times n$  矩阵, 即从  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性变换。那么

1.  $\dim \text{Ker } A + \dim \text{Ran } A = \dim \text{Ker } A + \text{rank } A = n$  ( $A$  的定义域的维数);
2.  $\dim \text{Ker } A^T + \dim \text{Ran } A^T = \dim \text{Ker } A^T + \text{rank } A^T = \dim \text{Ker } A^T + \text{rank } A = m$  ( $A$  的目标空间的维数)

**证明:** 这个证明, 在上述计算基本子空间基的方法所传达的意思中, 几乎是显然的。第一个陈述仅仅是自由变量的数量 ( $\dim \text{Ker } A$ ) 加上基本变量的数量 (即主元数量, 即  $\text{rank } A$ ) 等于列的数量 (即等于  $n$ )。

第二个陈述, 考虑到  $\text{rank } A = \text{rank } A^T$ , 仅仅是将第一个陈述应用于  $A^T$ .  $\square$

作为上述定理的一个应用, 让我们回顾一下第 6 节的例子。在那里, 我们考虑了系统

$$\begin{pmatrix} 2 & 3 & 1 & 4 & -9 \\ 1 & 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & 2 & -5 \\ 2 & 2 & 2 & 3 & -8 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 17 \\ 6 \\ 8 \\ 14 \end{pmatrix},$$

并且我们声称它的通解由

$$\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \quad x_3, x_5 \in \mathbb{F}$$

或者由

$$\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}, \quad s, t \in \mathbb{F}$$

给出。我们在第 6 节中核验了由任一公式给出的向量  $\mathbf{x}$  确实是方程的解。但是，我们如何保证 (6.1) 或 (6.2) 中的任何一个公式都给出了所有的解？

首先，我们知道在任一公式中，最后两个向量（被参数乘的向量）都属于  $\text{Ker } A$ 。很容易看出，在任一情况下，这两个向量都是线性无关的（两个向量线性相关当且仅当其中一个为另一个的某标量倍）。

现在，让我们计算维数：将第一行和第二行交换，并进行第一轮行运算

$$\begin{pmatrix} 1 & 1 & 1 & 1 & -3 \\ 2 & 3 & 1 & 4 & -9 \\ 1 & 1 & 1 & 2 & -5 \\ 2 & 2 & 2 & 3 & -8 \end{pmatrix} \xrightarrow[R_3 - R_1; R_4 - 2R_1]{R_2 - 2R_1} \begin{pmatrix} 1 & 1 & 1 & 1 & -3 \\ 0 & 1 & -1 & 2 & -3 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix}$$

我们看到已经有三个主元了，所以  $\text{rank } A \geq 3$ 。（实际上，我们可以已经看出秩是 3，但这里只需估计就足够了）。根据定理 7.2， $\text{rank } A + \dim \text{Ker } A = 5$ ，因此  $\dim \text{Ker } A \leq 2$ ，所以  $\text{Ker } A$  中不能有超过 2 个线性无关向量。因此，任一公式中的最后 2 个向量构成了  $\text{Ker } A$  的基，所以任一公式都给出了方程的所有解。

秩定理的一个重要推论，是以下将存在性和唯一性联系起来的关于线性方程的定理。

**定理 7.3.** 设  $A$  是一个  $m \times n$  矩阵。那么方程

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

对于每一个  $\mathbf{b} \in \mathbb{R}^m$  都有解，当且仅当对偶方程

$$A^T \mathbf{x} = \mathbf{0}$$

只有唯一的（仅平凡的）解。（注意，在第二个方程中的是  $A^T$ ，而不是  $A$ ）。

**证明：**证明直接从定理 7.2 得出，只需计算维数即可。我们将具体细节留给读者作为练习。  $\square$

上述定理有一个很好的几何解释。也就是说，陈述 1 表明，如果一个变换  $A: \mathbb{F}^n \rightarrow \mathbb{F}^m$  具有平凡核 ( $\text{Ker } A = \{0\}$ )，那么定义域  $\mathbb{F}^n$  和像空间  $\text{Ran } A$  的维数相等。如果核不是平凡的，那么变换“消去” (kills) 了  $\dim \text{Ker } A$  个维数，所以  $\dim \text{Ran } A = n - \dim \text{Ker } A$ 。

## 7.4. 将线性无关系统补全为基

正如上面第 5 节的命题 5.4 所断言的，任何线性无关系统都可以补全为基，即，给定有限维向量空间  $V$  中的线性无关向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ ，可以找到向量  $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n$  使得向量系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组基。

理论上，这个命题的证明为我们提供了寻找向量  $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n$  的算法，但该算法看起来不太实用。

本节的思想为我们提供了一种更实用的补全为基的方法。

首先，注意到如果一个  $m \times n$  矩阵处于阶梯形，那么它的非零行（它们显然是线性无关的）可以很容易地补全为  $\mathbb{F}^n$  中的基：我们只需要在适当的位置添加一些行，使得结果矩阵仍然是阶梯形并且在每一列都有主元。

然后，新矩阵的非零行构成一组基。我们可以按任何我们想要的顺序排列它们，因为作为基的性质不依赖于顺序。

假设现在我们有线性无关向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_k \in \mathbb{F}^n$ 。考虑以  $\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_r^T$  为行的矩阵  $A$ ，并执行行运算得到阶梯形  $A_e$ 。正如我们上面所讨论的， $A_e$  的行可以很容易地补全为  $\mathbb{R}^n$  中的基；结果是，能够补全  $A_e$  的行使其成为一组基的向量，同样能够补全原始向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  为一组基。

为了证明这一点，设向量  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$  补全了  $A_e$  的行，使其在  $\mathbb{F}^n$  中成为一组基。然后，如果我们向矩阵  $A_e$  添加行  $\mathbf{v}_{r+1}^T, \dots, \mathbf{v}_n^T$ ，我们会得到一个可逆矩阵。称此矩阵为  $\tilde{A}_e$ ，并设  $\tilde{A}$  是通过添加行  $\mathbf{v}_{r+1}^T, \dots, \mathbf{v}_n^T$  从  $A$  得到的矩阵。矩阵  $\tilde{A}_e$  可以通过行运算从  $\tilde{A}$  得到，所以

$$\tilde{A}_e = E\tilde{A},$$

其中  $E$  是相应初等矩阵的乘积。那么  $\tilde{A} = E^{-1}\tilde{A}_e$ ，并且  $\tilde{A}$  作为可逆矩阵的乘积，也是可逆的。

但这表示  $\tilde{A}$  的行在  $\mathbb{F}^n$  中构成一组基，这正是我们所需要的。

**注记** 上面描述的补全为基的方法可能不是最简单的方法，但其主要优点之一是它适用于任意域上的向量空间。



## 练习

7.1. 判断正误:

- a) 矩阵的秩等于其非零列的数量;
- b)  $m \times n$  零矩阵是唯一的秩为 0 的  $m \times n$  矩阵;
- c) 初等行运算保持秩;
- d) 初等列运算不一定保持秩;
- e) 矩阵的秩等于矩阵中线性无关列的最大数量;
- f) 矩阵的秩等于矩阵中线性无关行的最大数量;
- g)  $n \times n$  矩阵的秩最多为  $n$ ;
- h) 秩为  $n$  的  $n \times n$  矩阵是可逆的。

7.2. 一个  $54 \times 37$  的矩阵秩为 31. 所有 (4 个) 基本子空间的维数是多少?

7.3. 计算矩阵

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

的秩和所有四组基本子空间的基。

7.4. 证明, 如果  $A: X \rightarrow Y$ , 且  $V$  是  $X$  的子空间, 那么  $\dim AV \leq \text{rank } A$ . (这里  $AV$  表示进行了  $A$  变换后的子空间  $V$ , 即  $AV$  中的任何向量都可以表示为  $A\mathbf{v}$ ,  $\mathbf{v} \in V$ ). 由此推导出  $\text{rank}(AB) \leq \text{rank } A$ .

**注记:** 这里你可以利用  $V \subset W$  则  $\dim V \leq \dim W$  的事实. 你知道它为什么是对的吗?

7.5. 证明, 如果  $A: X \rightarrow Y$ , 且  $V$  是  $X$  的子空间, 那么  $\dim AV \leq \dim V$ . 由此推导出  $\text{rank}(AB) \leq \text{rank } B$ .

7.6. 证明, 如果两个  $n \times n$  矩阵  $A$  和  $B$  的乘积  $AB$  是可逆的, 那么  $A$  和  $B$  都是可逆的. (即使你知道行列式解法, 也请不要使用, 因为我们现在还没有引入它.) **提示:** 使用前两问的结论。

7.7. 证明, 如果  $A\mathbf{x} = \mathbf{0}$  只有唯一解, 那么方程  $A^T\mathbf{x} = \mathbf{b}$  对于每个右侧  $\mathbf{b}$  都有解. **提示:** 计算主元。

7.8. 构造一个具有所需性质的矩阵, 或解释为什么不存在这样的矩阵:

- a) 列空间包含  $(1, 0, 0)^T, (0, 0, 1)^T$ , 行空间包含  $(1, 1)^T, (1, 2)^T$ ;
- b) 列空间由  $(1, 1, 1)^T$  张成, 零空间由  $(1, 2, 3)^T$  张成;
- c) 列空间是  $\mathbb{R}^4$ , 行空间是  $\mathbb{R}^3$ .

**提示:** 首先核验维数是否匹配。

7.9. 如果  $A$  和  $B$  具有相同的四组基本子空间, 那么  $A = B$  成立吗?7.10. 将以下向量的行补全为  $\mathbb{R}^7$  的基:

$$\begin{pmatrix} e^3 & 3 & 4 & 0 & -\pi & 6 & -2 \\ 0 & 0 & 2 & -1 & \pi^e & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 & -3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

7.11. 对于矩阵

$$\begin{pmatrix} 1 & 2 & -1 & 2 & 3 \\ 2 & 2 & 1 & 5 & 5 \\ 3 & 6 & -3 & 0 & 24 \\ -1 & -4 & 4 & -7 & 11 \end{pmatrix},$$

找到其列空间和行空间的基。

7.12. 对于上题的矩阵, 将行空间的基补全为  $\mathbb{R}^5$  的基。

7.13. 对于矩阵

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix},$$

计算  $\text{Ran } A$  和  $\text{Ker } A$ . 你能看出这些子空间之间的关系吗?

7.14. 对于实数矩阵  $A$ ,  $\text{Ran } A = \text{Ker } A^T$  是否可能? 若对于复数矩阵  $A$ , 是否可能?

7.15. 将向量  $(1, 2, -1, 2, 3)^T, (2, 2, 1, 5, 5)^T, (-1, -4, 4, 7, -11)^T$  补全为  $\mathbb{R}^5$  的基。

## 8. 任意基下线性变换的表示 坐标变换公式

我们在第一章中学到的关于线性变换及其矩阵的知识, 可以很容易地推广到有限基下的抽象向量空间中的变换。在本节中, 我们将区分线性变换  $T$  和它的矩阵, 原因是我们将考虑不同的基, 所以线性变换可以有不同矩阵表示。

### 8.1. 坐标向量

设  $V$  是一个向量空间, 具有基  $\mathcal{B} := \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ . 任何向量  $\mathbf{v} \in V$  都可以唯一地表示为线性组合

$$\mathbf{v} = x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + \cdots + x_n \mathbf{b}_n = \sum_{k=1}^n x_k \mathbf{b}_k.$$

系数  $x_1, x_2, \dots, x_n$  称为向量  $\mathbf{v}$  在基  $\mathcal{B}$  下的**坐标** (coordinates)。可以方便地将这些坐标组合成向量  $\mathbf{v}$  相对于基  $\mathcal{B}$  的所谓的**坐标向量** (coordinate vector), 这是一个列向量

$$[\mathbf{v}]_{\mathcal{B}} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n.$$

注意, 映射

$$\mathbf{v} \mapsto [\mathbf{v}]_{\mathcal{B}}$$

是  $V$  和  $\mathbb{F}^n$  之间的同构。它将基  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  映射到  $\mathbb{F}^n$  中的标准基  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ .

### 8.2. 线性变换的矩阵

设  $T: V \rightarrow W$  是一个线性变换, 设  $\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}, \mathcal{B} := \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\}$  分别是  $V$  和  $W$  中的基。

变换  $T$  在基  $\mathcal{A}$  和  $\mathcal{B}$  下 (或相对于基  $\mathcal{A}$  和  $\mathcal{B}$ ) 的矩阵是  $m \times n$  矩阵, 记作  $[T]_{\mathcal{B}\mathcal{A}}$ , 它关联了坐标向量  $[T\mathbf{v}]_{\mathcal{B}}$  和  $[\mathbf{v}]_{\mathcal{A}}$ :

$$[T\mathbf{v}]_{\mathcal{B}} = [T]_{\mathcal{B}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}}$$

注意这里  $\mathcal{A}$  和  $\mathcal{B}$  下标符号的对应: 这是我们将第一组基  $\mathcal{A}$  置于第二个位置的原因。

矩阵  $[T]_{\mathcal{B}\mathcal{A}}$  很容易找到：它的第  $k$  列就是坐标向量  $[Ta_k]_{\mathcal{B}}$ （与从  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性变换矩阵的寻找方法进行比较！）。

正如在  $\mathbb{F}^n$  空间和标准基情况一样，线性变换的复合等价于它们的矩阵乘法：我们只需要在基方面更小心一些。也就是说，设  $T_1 : X \rightarrow Y$  和  $T_2 : Y \rightarrow Z$  是线性变换，设  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  分别是  $X, Y, Z$  中的基。那么对于复合  $T = T_2 T_1$ ,

$$T : X \rightarrow Z, \quad T\mathbf{x} := T_2(T_1(\mathbf{x})),$$

我们有

$$(8.1) \quad [T]_{\mathcal{C}\mathcal{A}} = [T_2 T_1]_{\mathcal{C}\mathcal{A}} = [T_2]_{\mathcal{C}\mathcal{B}} [T_1]_{\mathcal{B}\mathcal{A}}$$

（再次注意这里的下标对应）。

这个证明与  $\mathbb{F}^n$  空间在标准基下的证明完全相同，所以我们在此不再重复。另一种可能性是，通过坐标同构  $\mathbf{v} \mapsto [\mathbf{v}]_{\mathcal{B}}$  将所有内容传递到  $\mathbb{F}^n$  空间。然后我们就不需要任何证明了，一切都遵循矩阵乘法的相关结果。

### 8.3. 坐标变换矩阵

设我们在向量空间  $V$  中有两组基  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  和  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ . 考虑恒等变换  $I = I_V$  以及它在这些基下的矩阵  $[I]_{\mathcal{B}\mathcal{A}}$ . 根据定义，

$$[\mathbf{v}]_{\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}} [\mathbf{v}]_{\mathcal{A}}, \quad \forall \mathbf{v} \in V,$$

也就是说，对于任何向量  $\mathbf{v} \in V$ ，矩阵  $[I]_{\mathcal{B}\mathcal{A}}$  将其在基  $\mathcal{A}$  下的坐标变换为在基  $\mathcal{B}$  下的坐标。矩阵  $[I]_{\mathcal{B}\mathcal{A}}$  通常称为**坐标变换矩阵** (change of coordinates matrix)（从基  $\mathcal{A}$  到基  $\mathcal{B}$ ）。

矩阵  $[I]_{\mathcal{B}\mathcal{A}}$  很容易计算：根据线性变换矩阵的一般寻找规则，它的第  $k$  列是基  $\mathcal{A}$  下第  $k$  组基元素的坐标表示  $[\mathbf{a}_k]_{\mathcal{B}}$ .

注意

$$[I]_{\mathcal{A}\mathcal{B}} = ([I]_{\mathcal{B}\mathcal{A}})^{-1}$$

（这从矩阵乘法规则 (8.1) 中立即得出），所以任何坐标变换矩阵总是可逆的。

#### 8.3.1. 从标准基进行坐标变换的例子

设我们的空间  $V$  是  $\mathbb{F}^n$ ，并且我们有一组基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ . 我们还有标准基  $\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ . 从  $\mathcal{B}$  到  $\mathcal{S}$  的坐标变换矩阵  $[I]_{\mathcal{S}\mathcal{B}}$  很容易计算：

$$[I]_{\mathcal{S}\mathcal{B}} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n] =: B,$$

即它只是矩阵  $B$ ，其第  $k$  列是（列）向量  $\mathbf{v}_k$ . 反向也成立

$$[I]_{\mathcal{B}\mathcal{S}} = [I]_{\mathcal{S}\mathcal{B}}^{-1} = B^{-1}.$$

例如, 考虑  $\mathbb{F}^2$  中的一组基

$$B = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\},$$

设  $\mathcal{S}$  表示那里的标准基。那么

$$[I]_{\mathcal{S}B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} =: B,$$

并且

$$[I]_{B\mathcal{S}} = [I]_{\mathcal{S}B}^{-1} = B^{-1} = \frac{1}{3} \begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix}$$

(我们现在知道如何计算逆, 并且也很容易验证上述矩阵确实是  $B$  的逆)。

### 8.3.2. 通过标准基进行变换的例子

在最多为 1 次的多项式空间中, 我们还有基

$$\mathcal{A} = \{1, 1+x\}, \quad \text{和} \quad \mathcal{B} = \{1+2x, 1-2x\},$$

并且我们想找到坐标变换矩阵  $[I]_{B\mathcal{A}}$ 。

当然, 我们总是可以从基  $\mathcal{A}$  中取向量并尝试在基  $\mathcal{B}$  中分解它们; 这涉及到求解线性系统, 而我们知道如何做到这一点。

然而, 我认为以下方法更简单。在  $\mathbb{P}_1$  中, 我们还有标准基  $\mathcal{S} = \{1, x\}$ , 对于这组基

$$[I]_{\mathcal{S}\mathcal{A}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} =: A, \quad [I]_{\mathcal{S}B} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} =: B,$$

并且取逆

$$[I]_{\mathcal{A}\mathcal{S}} = A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad [I]_{B\mathcal{S}} = B^{-1} = \frac{1}{4} \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}.$$

那么<sup>23</sup>

$$[I]_{B\mathcal{A}} = [I]_{B\mathcal{S}}[I]_{\mathcal{S}\mathcal{A}} = B^{-1}A = \frac{1}{4} \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

并且

$$[I]_{\mathcal{A}B} = [I]_{\mathcal{A}\mathcal{S}}[I]_{\mathcal{S}B} = A^{-1}B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}.$$

---

<sup>23</sup>注意这里的下标对应。

## 8.4. 变换的矩阵与坐标变换

设  $T: V \rightarrow W$  是一个线性变换, 设  $\mathcal{A}, \tilde{\mathcal{A}}$  是  $V$  中的两组基, 设  $\mathcal{B}, \tilde{\mathcal{B}}$  是  $W$  中的两组基。假设我们知道矩阵  $[T]_{\mathcal{B}\mathcal{A}}$ , 并且我们想找到关于新基  $\tilde{\mathcal{A}}, \tilde{\mathcal{B}}$  的矩阵表示, 即矩阵  $[T]_{\tilde{\mathcal{B}}\tilde{\mathcal{A}}}$ , 规则非常简单:

要得到“新”基下的矩阵, 需要用坐标变换矩阵将“旧”基下的矩阵包围起来。

我在这里没有提及应该将哪个坐标变换矩阵放在哪里, 因为如果我们遵循下标对应规则, 我们别无选择。也就是说, 线性变换的矩阵表示遵循以下公式:

$$[T]_{\tilde{\mathcal{B}}\tilde{\mathcal{A}}} = [I]_{\tilde{\mathcal{B}}\mathcal{B}}[T]_{\mathcal{B}\mathcal{A}}[I]_{\mathcal{A}\tilde{\mathcal{A}}}$$

(请注意这里的下标对应)。证明可以通过分析每个矩阵的作用来完成。

## 8.5. 单组基的情况: 相似矩阵

设  $V$  是一个向量空间, 设  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  是  $V$  中的一组基。考虑一个线性变换  $T: V \rightarrow V$ , 并设  $[T]_{\mathcal{A}\mathcal{A}}$  是它在该基下的矩阵 (我们对“输入”和“输出”使用相同的基)。<sup>24</sup>

当我们将“输入”和“输出”使用相同基时, 这种情况非常重要 (因为此时我们可以将一个矩阵与自身相乘), 所以让我们更仔细地研究一下。请注意, 在这种情况下, 人们常常使用更简洁的记号  $[T]_{\mathcal{A}}$  来代替  $[T]_{\mathcal{A}\mathcal{A}}$ 。然而, 双下标表示法  $[T]_{\mathcal{A}\mathcal{A}}$  更适用于“下标对应规则”, 所以我建议在坐标变换时使用它 (或至少时刻牢记这一点)。

设  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是  $V$  中的另一组基。根据上面的坐标变换规则

$$[T]_{\mathcal{B}\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}}[T]_{\mathcal{A}\mathcal{A}}[I]_{\mathcal{A}\mathcal{B}}.$$

回忆

$$[I]_{\mathcal{B}\mathcal{A}} = [I]_{\mathcal{A}\mathcal{B}}^{-1}$$

并记  $Q := [I]_{\mathcal{A}\mathcal{B}}$ , 我们可以将上述公式改写为

$$[T]_{\mathcal{B}\mathcal{B}} = Q^{-1}[T]_{\mathcal{A}\mathcal{A}}Q.$$

这为以下定义提供了动机:

**定义 8.1.** 我们说矩阵  $A$  与矩阵  $B$  **相似**, 如果存在一个可逆矩阵  $Q$  使得  $A = Q^{-1}BQ$ .

因为可逆矩阵必须是方阵, 且通过计算维数后, 我们可以看出, 相似矩阵  $A$  和  $B$  必须是方阵且大小相同 (同型)。如果  $A$  与  $B$  相似, 即如果  $A = Q^{-1}BQ$ , 那

<sup>24</sup>符号  $[T]_{\mathcal{A}}$  常常被用来代替  $[T]_{\mathcal{A}\mathcal{A}}$ 。它虽然更简洁, 但是双下标表示法能更好地适用于“下标对应规则”。

么

$$B = QAQ^{-1} = (Q^{-1})^{-1}A(Q^{-1})$$

(因为  $Q^{-1}$  是可逆的), 因此  $B$  与  $A$  相似。所以, 我们可以简单地说  $A$  和  $B$  是**相似的** (similar)。

上述推理表明,  $Q$  和  $Q^{-1}$  哪个放在左哪个放在右并不重要: 人们也可以使用公式  $A = QBQ^{-1}$  来定义相似性。

上述讨论表明, 我们可以将相似的矩阵视为同一线性算子 (变换) 的不同矩阵表示。

### 练习

8.1. 判断正误:

- a) 任何坐标变换矩阵都是方阵;
- b) 任何坐标变换矩阵都是可逆的;
- c) 如果矩阵  $A$  和  $B$  相似, 那么  $B = Q^T A Q$  对于某些矩阵  $Q$  成立;
- d) 如果矩阵  $A$  和  $B$  相似, 那么  $B = Q^{-1} A Q$  对于某些矩阵  $Q$  成立;
- e) 相似矩阵不一定是方阵。

8.2. 考虑向量系统

$$(1, 2, 1, 1)^T, \quad (0, 1, 3, 1)^T, \quad (0, 3, 2, 0)^T, \quad (0, 1, 0, 0)^T.$$

- a) 证明它们是  $\mathbb{F}^4$  中的一组基。尽量少做计算。
- b) 找到将此基下的坐标变为  $\mathbb{F}^4$  中标准坐标 (即标准基  $\mathbf{e}_1, \dots, \mathbf{e}_4$  下的坐标) 的坐标变换矩阵。

8.3. 找到将  $\mathbb{P}_1$  中的基  $1, 1+t$  下的坐标变为基  $1-t, 2t$  下的坐标的坐标变换矩阵。

8.4. 设  $T$  是  $\mathbb{R}^2$  中的线性算子, 定义为 (在标准坐标下)

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3x + y \\ x - 2y \end{pmatrix}.$$

找到  $T$  在标准基下的矩阵, 以及在基

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

下的矩阵。

8.5. 证明, 如果  $A$  和  $B$  相似, 那么  $\text{trace } A = \text{trace } B$ . **提示:** 回忆  $\text{trace}(XY)$  和  $\text{trace}(YX)$  是如何关联的。

8.6. 矩阵

$$\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} 0 & 2 \\ 4 & 2 \end{pmatrix}$$

是否相似? 请给出理由。

## 第三章 行列式

### 1. 引言

读者可能已经遇到过行列式，至少是在微积分或代数学中遇到的  $2 \times 2$  和  $3 \times 3$  矩阵的行列式。对于  $2 \times 2$  矩阵

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

行列式就是  $ad - bc$ ； $3 \times 3$  矩阵的行列式可以通过“大卫之星” (Star of David) 法则<sup>25</sup> 找到。

在本章中，我们想要引入  $n \times n$  矩阵的行列式，而不是仅仅给出一个形式定义。首先我想给出一些动机，然后推导出行列式应具备的一些性质。然后，如果我们想要让这些性质同时成立，我们就别无选择，只能得到行列式的几个等价定义。

从矩阵的行列式开始引入，而不是从向量组的行列式开始，因为前者更为方便：这里没有真正的区别，因为我们可以始终将向量拼接在一起（比如作为列）来构成一个矩阵。

设我们有  $\mathbb{R}^n$  中的  $n$  个向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ （注意向量的数量与维数一致），我们想找到由这些向量确定的平行六面体的  $n$  维体积。

由向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  确定的平行六面体可以定义为所有向量  $\mathbf{v} \in \mathbb{R}^n$  的集合，这些向量可以表示为

$$\mathbf{v} = t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \dots + t_n \mathbf{v}_n, \quad 0 \leq t_k \leq 1 \quad \forall k = 1, 2, \dots, n.$$

当  $n = 2$ （平行四边形）和  $n = 3$ （平行六面体）时，这很容易可视化。那么  $n$  维体积是什么？

在维数为 2 时，它表示面积；在维数为 3 时，它表示体积；在维数为 1 时，它则是长度。

最后，让我们引入一些符号。对于（列）向量组  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ ，我们把它的行列式（我们将要构造的）表示为  $D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ 。如果我们把这些向量拼接成矩阵  $A$ （ $A$  的第  $k$  列是  $\mathbf{v}_k$ ），那么我们将使用符号  $\det A$ ，

$$\det A = D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n).$$

---

<sup>25</sup> 译者注：大卫之心，即六芒星，可以将两个三角形中心重合地以反方向覆盖而得到。其名称与犹太历史上的大卫王有关。

对于矩阵

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix},$$

它的行列式也常常表示为

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{vmatrix}.$$

## 2. 行列式应具备的性质

我们知道, 对于维度 2 和 3, 平行六面体的“体积”由“底乘以高”法则确定: 如果我们选择一个向量, 那么高是该向量到由其余向量张成的子空间的距离, 底是其余向量确定的平行六面体的 ( $n-1$  维) 体积。

现在让我们将这个想法推广到更高维度。我们暂时不关心如何精确地确定高和底。我们将表明, 如果我们假设高和底满足某些自然性质, 那么我们就别无选择, 行列式就被唯一确定了。

### 2.1. 关于列向量的线性性质

首先, 如果我们把向量  $\mathbf{v}_1$  乘以一个正数  $a$ , 那么, 高 (即到线性张成  $\mathcal{L}(\mathbf{v}_2, \dots, \mathbf{v}_n)$  的距离) 就会乘以  $a$ . 如果我们允许负高度 (和负体积), 那么这个性质对所有标量  $a$  都成立, 因此向量组  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的行列式  $D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$  应该满足

$$D(\alpha \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \alpha D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

当然, 向量  $\mathbf{v}_1$  没有什么特别之处, 所以对于任何下标  $k$

$$(2.1) \quad D(\mathbf{v}_1, \dots, \underbrace{\alpha \mathbf{v}_k}_k, \dots, \mathbf{v}_n) = \alpha D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k}_k, \dots, \mathbf{v}_n)$$

为了得到下一个性质, 让我们注意到, 如果我们将两个向量相加, 那么结果的“高度”应该等于被加数“高度”的总和, 即

$$(2.2) \quad D(\mathbf{v}_1, \dots, \underbrace{\mathbf{u}_k + \mathbf{v}_k}_k, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \underbrace{\mathbf{u}_k}_k, \dots, \mathbf{v}_n) + D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k}_k, \dots, \mathbf{v}_n)$$

换句话说, 上述两个性质表明, 行列式是**每个参数 (向量) 的线性**, 这意味着如果我们固定  $n-1$  个向量, 并将剩余向量看作一个变量 (参数), 我们就会得到一个线性函数。



**注记** 我们已经知道**线性性质** (linearity) 是一个非常好的性质, 在许多情况下对我们都有帮助。因此, 允许负高度 (以及由此而来的负体积) 是获得线性的一个非常小的代价, 因为我们之后总是可以取绝对值。

实际上, 通过允许负高度, 我们并没有牺牲任何东西! 相反, 我们甚至获得了一些东西, 因为行列式的符号包含了有关向量系统 (方向) 的一些信息。

## 2.2. “列替换” 下保持不变性

下一个性质也看起来很自然。也就是说, 如果我们取一个向量, 比如  $\mathbf{v}_j$ , 并将其加上另一个向量  $\mathbf{v}_k$  的倍数, “高度” 不会改变, 所以

$$(2.3) \quad D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j + \alpha \mathbf{v}_k}_{j}, \dots, \underbrace{\mathbf{v}_k}_{k}, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j}_{j}, \dots, \underbrace{\mathbf{v}_k}_{k}, \dots, \mathbf{v}_n)$$

换句话说, 如果我们应用第三种类型的**列运算**, 行列式不会改变。

**注记** 虽然在此并非必需, 但让我们注意到第二个部分的线性性质 (性质 (2.2)) 不是独立的: 它可以从性质 (2.1) 和 (2.3) 推导出来。

我们将证明留作读者的练习。

## 2.3. 反对称性

行列式应该具备的另外一个性质, 是如果我们交换了两个向量, 行列式的符号改变一次。<sup>26</sup>

也就是说, 如果我们交换两个向量, 行列式会变号:

$$(2.4) \quad D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k}_{j}, \dots, \underbrace{\mathbf{v}_j}_{k}, \dots, \mathbf{v}_n) = -D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j}_{j}, \dots, \underbrace{\mathbf{v}_k}_{k}, \dots, \mathbf{v}_n)$$

第一次看到时, 这个性质看起来并不自然, 但它可以从前面的性质推导出来。也就是说, 三次应用性质 (2.3), 然后使用 (2.1), 我们得到

$$\begin{aligned} & D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j}_{j}, \dots, \underbrace{\mathbf{v}_k}_{k}, \dots, \mathbf{v}_n) \\ &= D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j}_{j}, \dots, \underbrace{\mathbf{v}_k - \mathbf{v}_j}_{k}, \dots, \mathbf{v}_n) \\ &= D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_j + (\mathbf{v}_k - \mathbf{v}_j)}_{j}, \dots, \underbrace{\mathbf{v}_k - \mathbf{v}_j}_{k}, \dots, \mathbf{v}_n) \\ &= D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k}_{j}, \dots, \underbrace{\mathbf{v}_k - \mathbf{v}_j}_{k}, \dots, \mathbf{v}_n) \\ &= D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k}_{j}, \dots, \underbrace{(\mathbf{v}_k - \mathbf{v}_j) - \mathbf{v}_k}_{k}, \dots, \mathbf{v}_n) \end{aligned}$$

<sup>26</sup>具有多个变量的一些函数的性质, 是在交换任意两个参数时会变号, 这类函数称为反对称函数。

$$\begin{aligned}
&= D(\mathbf{v}_1, \dots, \underset{j}{\mathbf{v}_k}, \dots, -\underset{k}{\mathbf{v}_j}, \dots, \mathbf{v}_n) \\
&= -D(\mathbf{v}_1, \dots, \underset{j}{\mathbf{v}_k}, \dots, \underset{k}{\mathbf{v}_j}, \dots, \mathbf{v}_n).
\end{aligned}$$

## 2.4. 归一化

最后个性质是最简单的。对于  $\mathbb{R}^n$  中的标准基  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ ，对应的平行六面体是  $n$  维单位立方体，所以

$$(2.5) \quad D(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1.$$

在矩阵表示中，这可以写成

$$\det(I) = 1.$$

## 3. 行列式的构造

我们现在的计划是：利用我们从第 2 节认为的行列式应该具有的性质，推导出行列式的其他性质，其中一些性质非常不平凡。我们将展示如何使用这些性质通过我们熟悉的朋友——行约简来计算行列式。

稍后，在第 4 节，我们将展示行列式，即具有所需性质的函数，是存在且唯一的。毕竟，我们必须确信我们正在计算和研究的对象是存在的。

虽然我们引入行列式及其性质的初始几何动机来自于考虑  $\mathbb{R}^n$  中的向量，因此它们只与实数项的矩阵相关，但以下所有构造只使用代数运算（加法、乘法、除法）并且适用于具有复数项的矩阵，甚至适用于具有任意域项的矩阵。

因此，在以下内容中，我们不仅为实数矩阵构造行列式，也为复数矩阵（以及具有任意域项的矩阵）构造行列式。虽然我们最初的几何动机仅适用于实数情况，但在我们确定了行列式的性质（见本节的性质 1—3）之后，所有内容都适用于一般情况。

### 3.1. 基本性质

在这一节中，我们将使用以下行列式性质：

1. 行列式在每一列中都是线性的，即，在向量表示中，对于每个下标  $k$ ，

$$D(\mathbf{v}_1, \dots, \underbrace{\alpha \mathbf{u}_k + \beta \mathbf{v}_k}_k, \dots, \mathbf{v}_n) = \alpha D(\mathbf{v}_1, \dots, \underset{k}{\mathbf{u}_k}, \dots, \mathbf{v}_n) + \beta D(\mathbf{v}_1, \dots, \underset{k}{\mathbf{v}_k}, \dots, \mathbf{v}_n)$$

对所有标量  $\alpha, \beta$  成立。

2. 行列式是**反对称** (antisymmetric) 的，即，如果我们交换两列，行列式前正负符号改变一次。

3. 归一化性质:  $\det I = 1$ .

所有这些性质在第 2 节中都已讨论过。第一个性质只是 (2.1) 和 (2.2) 的组合。第二个是 (2.4)，最后一个是归一化性质 (2.5)。注意，我们没有使用性质 (2.3)：它可以从上述三个性质中推导出来。因此，这三个性质就完全定义了行列式！

### 3.2. 从行列式基本性质中推导出的性质

**命题 3.1** 对于方阵  $A$ ，以下陈述成立：

1. 如果  $A$  有一个零列，那么  $\det A = 0$ 。
2. 如果  $A$  有两列相等，那么  $\det A = 0$ ；
3. 如果  $A$  的一列是另一列的倍数，那么  $\det A = 0$ ；
4. 如果  $A$  的列是线性相关的，即如果矩阵不可逆，那么  $\det A = 0$ 。

**证明：** 陈述 1 由线性性质直接得出。如果我们用零乘以零列，我们不会改变矩阵及其行列式。但根据上面的性质 1，我们应该得到 0。

行列式的反对称性蕴含了陈述 2。实际上，如果我们交换两列相等的列，我们什么也没改变，所以行列式保持不变。另一方面，交换两列改变了行列式的符号，所以

$$\det A = -\det A,$$

这只有在  $\det A = 0$  时才可能。

陈述 3 是陈述 2 和线性性质的直接推论。

要证明最后一个陈述，让我们首先假设第一个向量  $\mathbf{v}_1$  是其他向量的线性组合，

$$\mathbf{v}_1 = \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 + \cdots + \alpha_n \mathbf{v}_n = \sum_{k=2}^n \alpha_k \mathbf{v}_k.$$

那么根据线性性质，我们有（在向量表示中）

$$D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = D\left(\left(\sum_{k=2}^n \alpha_k \mathbf{v}_k\right), \mathbf{v}_2, \dots, \mathbf{v}_n\right) = \sum_{k=2}^n \alpha_k D(\mathbf{v}_k, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

并且和中的每个行列式都为零，因为存在两个相等的列。

现在考虑一般情况，即假设系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是线性相关的。那么其中一个向量，比如  $\mathbf{v}_k$ ，可以表示为其他向量的线性组合。将此向量与  $\mathbf{v}_1$  交换，我们得到我们刚刚处理过的情况，所以

$$D(\mathbf{v}_1, \dots, \underset{k}{\mathbf{v}_k}, \dots, \mathbf{v}_n) = -D(\mathbf{v}_k, \dots, \underset{k}{\mathbf{v}_1}, \dots, \mathbf{v}_n) = -0 = 0,$$

所以这种情况下的行列式也为零。

□

下一个命题推广了性质 (2.3)。正如我们上面已经说过的，这个性质可以从我们本节中使用的三个“基本”性质中推导出来。

**命题 3.2** 当我们向一列添加其他列的线性组合时，行列式不会改变（保持其他列不变）。特别地，行列式在“列替换”（第三类列运算）下保持不变。<sup>27</sup>

**证明：** 固定一个向量  $\mathbf{v}_k$ ，令  $\mathbf{u}$  为其他向量的线性组合，

$$\mathbf{u} = \sum_{j \neq k} \alpha_j \mathbf{v}_j.$$

那么根据线性性质

$$D(\mathbf{v}_1, \dots, \underbrace{\mathbf{v}_k + \mathbf{u}}_k, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_n) + D(\mathbf{v}_1, \dots, \mathbf{u}, \dots, \mathbf{v}_n),$$

并且根据命题 3.1，最后一项为零。  $\square$

### 3.3. 对角和三角矩阵的行列式

现在我们准备计算一些重要的特殊矩阵类别的行列式。第一类是所谓的**对角 (diagonal)** 矩阵。让我们回顾一下，一个方阵  $A = \{a_{j,k}\}_{j,k=1}^n$  称为**对角矩阵**，如果**主对角线之外**的所有项都为零，即如果  $a_{j,k} = 0 \ \forall j \neq k$ 。我们将经常使用  $\text{diag}\{a_1, a_2, \dots, a_n\}$  来表示对角矩阵：

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix}.$$

由于对角矩阵  $\text{diag}\{a_1, a_2, \dots, a_n\}$  可以通过将第  $k$  列乘以  $a_k$  从单位矩阵  $I$  得到，所以

$$\text{对角矩阵的行列式等于所有对角项的乘积, } \det(\text{diag}\{a_1, a_2, \dots, a_n\}) = a_1 a_2 \dots a_n.$$

下一个重要类别是所谓的**三角 (triangular)** 矩阵。一个方阵  $A = \{a_{j,k}\}_{j,k=1}^n$  称为**上三角矩阵**，如果主对角线以下的所有项都为零，即如果  $a_{j,k} = 0 \ \forall k < j$ 。一个方阵称为**下三角矩阵**，如果主对角线以上的所有项都为零，即如果  $a_{j,k} = 0 \ \forall j < k$ 。我们称矩阵为**三角矩阵**，如果它是下三角或上三角矩阵。

很容易看出，

$$\text{三角矩阵的行列式等于所有对角项的乘积, } \det A = a_{1,1} a_{2,2} \dots a_{n,n}.$$

实际上，如果一个三角矩阵的主对角线上有零出现，那么它就是不可逆的（这可以通过列运算很容易地核验出来），因此两边都等于零。如果所有对角项都非零，

<sup>27</sup>把某列加上自己本身的某个倍数在这里是被禁止的。我们只能在其余非自身列上相加。

那么使用列替换（第三类列运算）可以将矩阵转化为具有相同对角项的对角矩阵：对于上三角矩阵，首先应该从第  $2, 3, \dots, n$  列减去第 1 列的适当倍数，“消去”第 1 行中的所有项，然后从第  $3, \dots, n$  列减去第 2 列的适当倍数，依此类推。

为了处理下三角矩阵的情况，需要从左到右进行“列约简”，即首先从最后一列开始，将适当倍数的最后一列减去第  $n-1, \dots, 2, 1$  列，依此类推。

### 3.4. 计算行列式

现在我们知道如何通过使用行列式的性质来计算行列式：只需进行列约简（即对  $A^T$  进行行约简），并注意可能会改变行列式符号的列运算。幸运的是，最常使用的运算——行替换，即第三类行运算，也不会改变行列式（见下一小节）。所以我们只需要留心列的交换和用标量乘以列。

如果  $A^T$  的阶梯形不在每一列（和每一行）都有主元，那么  $A$  是不可逆的，因此  $\det A = 0$ 。如果  $A$  是可逆的，我们得到一个三角矩阵，而  $\det A$  是对角项的乘积，乘以来自列交换和乘法的校正因子 (correction)。

上述算法暗示  $\det A$  仅在矩阵  $A$  不可逆时才可能为零。结合命题 3.1 的最后一个陈述，我们得到：

**命题 3.3**  $\det A = 0$  当且仅当  $A$  不可逆，或者等价地说： $\det A \neq 0$  当且仅当  $A$  可逆。

注意，虽然我们现在知道如何计算行列式，但行列式仍然没有被定义。有人可能会问：为什么我们不将其定义为通过上述算法得到的结果？问题在于，从形式上看，这个结果并非被良好定义：我们没有证明不同的列运算顺序会得到相同的结果。

### 3.5. 矩阵转置和乘积的行列式 初等矩阵的行列式

在本节中，我们将证明两个重要定理。

**定理 3.4.**（矩阵转置的行列式）对于方阵  $A$ ,

$$\det A = \det(A^T).$$

这个定理意味着，我们之前讨论过的所有关于列的陈述，相应关于行的陈述也都是正确的。特别地，行列式在行运算下的行为与在列运算下的行为相同。因此，我们可以使用行运算来计算行列式。

**定理 3.5.**（矩阵乘积的行列式）对于  $n \times n$  矩阵  $A$  和  $B$ :

$$\det(AB) = (\det A)(\det B).$$

换句话说,

矩阵乘积的行列式等于矩阵行列式的乘积。

为了证明这两个定理,我们需要先证以下引理。

**引理 3.6.** 对于方阵  $A$  和初等矩阵  $E$  (要求大小相同,即同型):

$$\det(AE) = (\det A)(\det E)$$

**证明:** 证明可以通过直接检验等式两边各自的计算结果来完成: 特殊矩阵的行列式很容易计算; 右乘初等矩阵对应进行列运算, 而列运算对行列式的作用是众所周知的。

这可能看起来像一个幸运的巧合, 即初等矩阵的行列式与其相应的列运算一致, 但这并非巧合。

我们知道, 每个列运算对应的初等矩阵, 可以由对单位矩阵  $I$  的该列加以相应列运算得到。所以, 它的行列式是 1 ( $I$  的行列式) 乘以这一列运算的作用。

要论证的一切就是这样子了! 读者一开始可能很难意识到, 但上述段落就是对引理完整且严谨的证明! □

应用引理 3.6  $N$  次, 我们得到以下推论。

**推论 3.7** 对于任何矩阵  $A$  和任何初等矩阵序列  $E_1, E_2, \dots, E_N$  (所有矩阵均为  $n \times n$ ):

$$\det(AE_1E_2 \dots E_N) = (\det A)(\det E_1)(\det E_2) \dots (\det E_N)$$

**引理 3.8.** 任何可逆矩阵都可以表示为初等矩阵的乘积。

**证明:** 我们知道任何可逆矩阵都可以通过行运算化为单位矩阵 (其简化阶梯形和单位矩阵行等价)。所以

$$I = E_N E_{N-1} \dots E_2 E_1 A,$$

因此任何可逆矩阵都可以表示为初等矩阵的乘积,

$$A = (E_N E_{N-1} \dots E_2 E_1)^{-1} I = E_1^{-1} E_2^{-1} \dots E_{N-1}^{-1} E_N^{-1}$$

(初等矩阵的逆也是初等矩阵)。□

**定理 3.4 的证明** 首先, 容易验证, 对于初等矩阵  $E$ ,  $\det E = \det(E^T)$ . 请注意, 只需为可逆矩阵  $A$  证明该定理即可, 因为如果  $A$  不可逆, 那么  $A^T$  也不可逆, 并且它们两个的行列式都为零。根据引理 3.8, 矩阵  $A$  可以表示为初等矩阵的乘积,

$$A = E_1 E_2 \dots E_N,$$

并且根据推论 3.7,  $A$  的行列式是初等矩阵行列式的乘积。由于取转置只是转置每个初等矩阵并反转它们的顺序, 因此推论 3.7 蕴含了  $\det A = \det A^T$ .  $\square$

**定理 3.5 的证明** 首先让我们假设矩阵  $B$  是可逆的。那么引理 3.8 蕴含了  $B$  可以表示为初等矩阵的乘积

$$B = E_1 E_2 \dots E_N,$$

因此根据推论 3.7

$$\det(AB) = (\det A) [(\det E_1)(\det E_2) \dots (\det E_N)] = (\det A)(\det B).$$

如果  $B$  不可逆, 那么乘积  $AB$  也不可逆, 而定理仅仅说明  $0 = 0$ .

要验证上面的乘积  $AB =: C$  是不可逆的, 就让我们假设  $C$  是可逆的。那么将恒等式  $AB = C$  从左边乘以  $C^{-1}$ , 我们得到  $C^{-1}AB = I$ , 所以  $C^{-1}A$  是  $B$  的左逆。因此  $B$  是左可逆的, 并且由于它是方阵, 所以  $B$  是可逆的。我们得到了一个矛盾。  $\square$

### 3.6. 行列式的性质总结

首先, 让我们再说一遍, **行列式仅对方阵有定义!** 由于我们现在知道  $\det A = \det(A^T)$ , 我们之前关于列的所有陈述也对行成立。

1. 行列式在每个行 (列) 中是线性的, 当其他行 (列) 固定不变时。
2. 如果我们交换矩阵  $A$  的两行 (列), 行列式改变符号 (正负号)。
3. 对于三角矩阵 (特别地, 对角矩阵), 其行列式是对角项的乘积。特别地,  $\det I = 1$ .
4. 如果矩阵  $A$  有一个零行 (或零列), 则  $\det A = 0$ .
5. 如果矩阵  $A$  有两行 (列) 相等, 则  $\det A = 0$ .
6. 如果  $A$  的某一行 (列) 是其他行 (列) 的线性组合, 即如果矩阵不可逆, 则  $\det A = 0$ ; 更一般地,
7.  $\det A = 0$  当且仅当  $A$  不可逆, 或者等价地说
8.  $\det A \neq 0$  当且仅当  $A$  可逆。
9. 如果我们将行 (列) 的线性组合加到某个行 (列) 上, 行列式不改变。特别地, 行列式在行 (列) 替换, 即第三类行 (列) 运算下保持不变。
10.  $\det A^T = \det A$ .
11.  $\det(AB) = (\det A)(\det B)$ . 最后,
12. 如果  $A$  是一个  $n \times n$  矩阵, 那么  $\det(\alpha A) = \alpha^n \det A$ .

最后一个性质是从行列式的线性性质得出的, 如果我们回忆起, 若要将矩阵  $A$  乘以  $\alpha$ , 我们必须将每一行乘以  $\alpha$ , 并且每次乘法都会将行列式乘以  $\alpha$ .

### 练习

3.1. 如果  $A$  是一个  $n \times n$  矩阵,  $\det(3A)$  与  $\det A$  有何关系?

注记:  $\det(3A) = 3 \det A$  仅在  $1 \times 1$  矩阵的平凡情况下成立。

3.2. 下面  $A$  和  $B$  各自的行列式之间有什么关系?

$$\text{a) } A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, \quad B = \begin{pmatrix} 2a_1 & 3a_2 & 5a_3 \\ 2b_1 & 3b_2 & 5b_3 \\ 2c_1 & 3c_2 & 5c_3 \end{pmatrix};$$

$$\text{b) } A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, \quad B = \begin{pmatrix} 3a_1 & 4a_2 + 5a_1 & 5a_3 \\ 3b_1 & 4b_2 + 5b_1 & 5b_3 \\ 3c_1 & 4c_2 + 5c_1 & 5c_3 \end{pmatrix}.$$

3.3. 使用列或行运算计算行列式:

$$\begin{vmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{vmatrix}, \quad \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}, \quad \begin{vmatrix} 1 & 0 & -2 & 3 \\ -3 & 1 & 1 & 2 \\ 0 & 4 & -1 & 1 \\ 2 & 3 & 0 & 1 \end{vmatrix}, \quad \begin{vmatrix} 1 & x \\ 1 & y \end{vmatrix}.$$

3.4. 一个方阵 ( $n \times n$ ) 称为**反对称** (skew-symmetric) (或**反交换**) 矩阵, 如果  $A^T = -A$ . 证明如果  $A$  是反对称的且  $n$  是奇数, 则  $\det A = 0$ . 这对偶数  $n$  是否成立?

3.5. 一个方阵称为**幂零** (nilpotent) 矩阵, 如果  $\exists k \in \mathbb{N}_+$ , 使得  $A^k = \mathbf{0}$  成立. 证明如果  $A$  是幂零的, 则  $\det A = 0$ .

3.6. 证明如果矩阵  $A$  和  $B$  相似, 则  $\det A = \det B$ .

3.7. 一个实方阵  $Q$  称为**正交**的, 如果  $Q^T Q = I$ . 证明如果  $Q$  是正交矩阵, 那么  $\det Q = \pm 1$ .

3.8. 证明

$$\begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix} = (z-x)(z-y)(y-x).$$

这是所谓的范德蒙德 (Vandermonde) 行列式的特例。

3.9. 设平面  $\mathbb{R}^2$  中的点  $A, B, C$  的坐标分别为  $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ . 证明三角形  $ABC$  的面积是

$$\frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}$$

的绝对值. **提示:** 使用行运算和  $2 \times 2$  行列式的几何解释 (面积)。

3.10. 设  $A$  和  $C$  是方阵, 证明分块三角矩阵

$$\begin{pmatrix} I & * \\ \mathbf{0} & A \end{pmatrix}, \quad \begin{pmatrix} A & * \\ \mathbf{0} & I \end{pmatrix}, \quad \begin{pmatrix} I & \mathbf{0} \\ * & A \end{pmatrix}, \quad \begin{pmatrix} A & \mathbf{0} \\ * & I \end{pmatrix}$$

的行列式都等于  $\det A$ . 这里  $*$  可以是任何东西. 以下问题说明了分块矩阵表示的力量。

3.11. 使用上一个问题证明, 如果  $A$  和  $C$  是方阵, 那么

$$\det \begin{pmatrix} A & B \\ \mathbf{0} & C \end{pmatrix} = (\det A)(\det C).$$

**提示:**  $\begin{pmatrix} A & B \\ \mathbf{0} & C \end{pmatrix} = \begin{pmatrix} I & B \\ \mathbf{0} & C \end{pmatrix} \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}.$



3.12. 设  $A$  是  $m \times n$  矩阵,  $B$  是  $n \times m$  矩阵. 证明

$$\det \begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix} = \det(AB).$$

**提示:** 虽然可以通过对矩阵进行行运算得到行列式易于计算的形式, 但最简单的方法是右乘矩阵  $\begin{pmatrix} I & \mathbf{0} \\ B & I \end{pmatrix}$ .

## 4. 行列式的正式定义 存在性与唯一性

在本节中, 我们将得到行列式的正式定义。我们表明了, 确实有函数满足第 3 节中的基本性质 1, 2, 3 的存在性, 而且, 这样的函数是唯一的, 也就是说, 在构造行列式时我们别无选择。

考虑一个  $n \times n$  矩阵  $A = \{a_{j,k}\}_{j,k=1}^n$ , 并设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是它的列, 即

$$\mathbf{v}_k = \begin{pmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{n,k} \end{pmatrix} = a_{1,k}\mathbf{e}_1 + a_{2,k}\mathbf{e}_2 + \cdots + a_{n,k}\mathbf{e}_n = \sum_{j=1}^n a_{j,k}\mathbf{e}_j$$

使用行列式的线性性质, 我们在第 1 列展开:

$$(4.1) \quad D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = D\left(\sum_{j=1}^n a_{j,1}\mathbf{e}_j, \mathbf{v}_2, \dots, \mathbf{v}_n\right) = \sum_{j=1}^n a_{j,1}D(\mathbf{e}_j, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

然后我们在第 2 列展开, 然后是第 3 列, 依此类推。我们得到

$$D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n a_{j_1,1}a_{j_2,2} \cdots a_{j_n,n} D(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \dots, \mathbf{e}_{j_n})$$

注意, 我们必须为每一列使用不同的求和下标(哑指标): 我们称它们为  $j_1, j_2, \dots, j_n$ ; 这里  $j_1$  的下标与 (4.1) 中的下标  $j$  相同。

这是一个巨大的求和, 包含  $n^n$  项。幸运的是, 其中一些项为零。也就是说, 如果  $j_1, j_2, \dots, j_n$  中有任何两个下标相同, 则行列式  $D(\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \dots, \mathbf{e}_{j_n})$  为零, 因为这里有两个相等的列。

因此, 让我们重写求和, 省略所有零项。最方便的方式是使用**排列**(permutation)的概念。非正式地说, 一个有序集  $\{1, 2, \dots, n\}$  的**排列**是其元素的重新排列。一种方便表示这种重新排列的形式是通过使用一个函数

$$\sigma: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\},$$

其中  $\sigma(1), \sigma(2), \dots, \sigma(n)$  给出了集合  $1, 2, \dots, n$  的新顺序。换句话说, 排列  $\sigma$  将有序集  $1, 2, \dots, n$  重排为  $\sigma(1), \sigma(2), \dots, \sigma(n)$ 。

这样的函数  $\sigma$  必须是**单射**（对不同的自变量取不同的值）和**满射**（取到目标空间的所有可能值）。既是单射又是满射的函数称为**双射** (bijection)，它们在定义域和目标空间之间建立了一一对应关系。<sup>28</sup>

尽管这在此处不直接相关，但让我们注意到，在组合学中，众所周知，集合  $\{1, 2, \dots, n\}$  的不同排列的数量恰好是  $n!$ 。所有  $n$  的排列的集合将被记为  $\text{Perm}(n)$ 。

使用排列的概念，我们可以将行列式重写为：

$$D(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sum_{\sigma \in \text{Perm}(n)} a_{\sigma(1),1} a_{\sigma(2),2} \dots a_{\sigma(n),n} D(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)})$$

矩阵  $\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)}$  的列可以从单位矩阵通过有限次数的列交换得到，所以行列式

$$D(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \dots, \mathbf{e}_{\sigma(n)})$$

是 1 或  $-1$ ，取决于列交换的次数。

为了将这一点形式化，我们（非正式地）定义排列  $\sigma$  的**符号**（记作  $\text{sign } \sigma$ ）为，如果将  $n$  元组  $1, 2, \dots, n$  重排为  $\sigma(1), \sigma(2), \dots, \sigma(n)$  所需的交换次数是偶数，则  $\text{sign}(\sigma) = 1$ ，如果交换次数是奇数，则  $\text{sign}(\sigma) = -1$ 。

这是组合学中的一个事实，符号是良好定义的，即虽然有无数种方法可以从  $1, 2, \dots, n$  得到  $n$  元组  $\sigma(1), \sigma(2), \dots, \sigma(n)$ ，但交换次数要么总是奇数，要么总是偶数。

一种证明这一点的方法是引入另一种定义。设  $K(\sigma)$  为  $\sigma$  的**逆序对** (disorder) 的数量，即满足  $\sigma(j) > \sigma(k)$  的整数对  $(j, k)$  的数量，其中  $j, k \in \{1, 2, \dots, n\}, j < k$ ，然后检查该数量是偶数还是奇数。我们将排列  $\sigma$  称为**奇排列**如果  $K$  是奇数，称为**偶排列**如果  $K$  是偶数。然后定义  $\text{sign } \sigma := (-1)^{K(\sigma)}$ ；注意通过这种方式定义的  $\text{sign } \sigma$  是明确的。

我们现在想要证明  $\text{sign } \sigma = (-1)^{K(\sigma)}$  可以通过将  $n$  元组  $1, 2, \dots, n$  重排为  $\sigma(1), \sigma(2), \dots, \sigma(n)$  并计算交换次数来得到，如上所述。

如果  $\sigma(k) = k \ \forall k$ ，那么**逆序对**的数量  $K(\sigma) = 0$ ，所以这种**恒等排列**的符号是 1。还请注意，任何两个相邻元素的**置换**（仅交换两个相邻元素）会改变排列的符号，因为它会改变逆序对的数量（增加或减少 1）。因此，要从一个排列得到另

<sup>28</sup>还有一种常用的方式来表示交换，即用一个双射  $\sigma$ ，在这个表示中， $\sigma(k)$  给出元素编号  $k$  在排列中的新位置。在这个表示中， $\sigma$  会将  $\sigma(1), \sigma(2), \dots, \sigma(n)$  排成  $1, 2, \dots, n$  的顺序。

虽然在第一种表示中写出函数比较容易（如果你知道排列的重排），但第二种更适合排列的组成：它与函数的组成是相符的。具体来说，如果我们先执行对应于函数  $\sigma$  的重排，然后再执行对应于  $\tau$  的重排，得到的排列就等于  $\tau \circ \sigma$ 。

一个排列，当排列具有相同的符号时，总是需要偶数次初等置换，而当符号不同时，则需要奇数次。

最后，任何两个元素的交换都可以通过奇数次初等置换来实现。这意味着当两个元素被交换时，符号会改变。因此，要从  $1, 2, \dots, n$  得到偶排列（正的符号）总是需要偶数次交换，而得到奇排列（负的符号）需要奇数次交换。

因此，如果我们希望行列式满足第 3 节中的基本性质 1—3，我们必须将其定义为：

$$(4.2) \quad \det A = \sum_{\sigma \in \text{Perm}(n)} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \text{sign}(\sigma)$$

其中求和遍历集合  $\{1, 2, \dots, n\}$  的所有排列。

如果我们这样定义行列式，可以很容易地验证它满足第 3 节中的基本性质 1—3。实际上，因为每个乘积项在每一列中恰好有一个因子，并且对于任何两个相邻的列交换，我们得到的符号会改变，所以满足线性性质和反对称性。

而且，对于单位矩阵  $I$ ，右侧实质上只有一项（对应于恒等排列  $\sigma(k) = k \forall k$ ），(4.2) 中右侧的符号是 1，所以  $D(I) = 1$ 。

### 练习

- 4.1. 假设排列  $\sigma$  将  $(1, 2, 3, 4, 5)$  映射到  $(5, 4, 1, 2, 3)$ .
  - a) 确定  $\sigma$  的符号；
  - b)  $\sigma^2 := \sigma \circ \sigma$  会对  $(1, 2, 3, 4, 5)$  做什么？
  - c) 逆排列  $\sigma^{-1}$  会对  $(1, 2, 3, 4, 5)$  做什么？
  - d)  $\sigma^{-1}$  的符号是什么？
- 4.2. 设  $P$  是一个排列矩阵 (permutation matrix)，即一个仅由 0 和 1 组成的  $n \times n$  矩阵，并且每行每列恰好有一个 1。
  - a) 你能描述相应的线性变换吗？这将会解释它的名称的由来。
  - b) 证明  $P$  是可逆的。你能描述  $P^{-1}$  吗？
  - c) 证明存在  $N > 0$ ,

$$P^N := \underbrace{PP \cdots P}_{N \text{ times}} = I.$$

利用排列只有有限个的事实。

- 4.3. 为什么  $(1, 2, \dots, 9)$  的排列有偶数个，并且其中恰好一半是奇排列？  
提示：这个问题用排列来解决可能很难，但有一个非常简单的行列式解。
- 4.4. 如果  $\sigma$  是一个奇排列，解释为什么  $\sigma^2$  是偶数但  $\sigma^{-1}$  是奇数。
- 4.5. 使用 (4.2) 的行列式形式计算一个  $n \times n$  矩阵的行列式需要多少次乘法和加法？无需计算  $\text{sign } \sigma$  所需的操作。

## 5. 代数余子式展开

对于  $n \times n$  矩阵  $A = \{a_{j,k}\}_{j,k=1}^n$ ，设  $A_{j,k}$  表示通过划掉第  $j$  行和第  $k$  列得到的  $(n-1) \times (n-1)$  矩阵，称为余子式。

**定理 5.1** (行列式的代数余子式展开) (cofactor expansion) 设  $A$  是一个  $n \times n$  矩阵。对于每个  $j, 1 \leq j \leq n$ , 行列式  $A$  可以按第  $j$  行展开为:

$$\begin{aligned} \det A &= a_{j,1}(-1)^{j+1} \det A_{j,1} + a_{j,2}(-1)^{j+2} \det A_{j,2} + \cdots + a_{j,n}(-1)^{j+n} \det A_{j,n} \\ &= \sum_{k=1}^n a_{j,k}(-1)^{j+k} \det A_{j,k}. \end{aligned}$$

类似地, 对于每个  $k, 1 \leq k \leq n$ , 行列式可以按第  $k$  列展开为:

$$\det A = \sum_{j=1}^n a_{j,k}(-1)^{j+k} \det A_{j,k}.$$

**证明:** 我们首先证明第 1 行的展开公式。第 2 行的展开公式可以通过交换第 1 行和第 2 行从它得到。然后交换第 2 行和第 3 行, 得到第 3 行的展开公式, 依此类推。

由于  $\det A = \det A^T$ , 列展开将自动跟上。

让我们首先考虑一个特殊情况, 即第 1 行只有一个非零项  $a_{1,1}$ 。通过对第 2, 3, ...,  $n$  列进行列运算, 我们可以将  $A$  转化为下三角形式。那么  $A$  的行列式可以计算为:

$$\boxed{\text{三角矩阵的所有对角项的乘积}} \times \boxed{\text{来源于列运算的修正因子}}.$$

但是, 除了  $a_{1,1}$  之外的所有对角项的乘积 (即不包括  $a_{1,1}$ ) 乘上修正因子恰好是  $\det A_{1,1}$ , 所以在这种特定情况下  $\det A = a_{1,1} \det A_{1,1}$ 。

现在考虑第 1 行中除了  $a_{1,2}$  外的其余项都为零的情况。这种情况可以通过交换第 1 列和第 2 列来化简到前面的情况, 因此在这种情况下  $\det A = (-1)^{1+2} a_{1,2} \det A_{1,2}$ 。

当  $a_{1,3}$  是第 1 行唯一非零项的情况, 可以通过交换第 2 行和第 3 行来简化到前面情况, 所以在这种情况下  $\det A = a_{1,3} \det A_{1,3}$ 。

重复这个过程, 我们得到, 当  $a_{1,k}$  是第 1 行中的唯一非零项时,  $\det A = (-1)^{1+k} a_{1,k} \det A_{1,k}$ 。<sup>29</sup>

在一般情况下, 行列式在每一行上的线性意味着

$$\det A = \det A^{(1)} + \det A^{(2)} + \cdots + \det A^{(n)} = \sum_{k=1}^n \det A^{(k)},$$

其中矩阵  $A^{(k)}$  是通过将  $A$  的第 1 行中除  $a_{1,k}$  之外的所有项替换为 0 而得到的。正如我们上面所讨论的,

$$\det A^{(k)} = (-1)^{1+k} a_{1,k} \det A_{1,k},$$

<sup>29</sup> 在  $a_{1,k}$  是第 1 行中唯一非零项的情况下, 这可能会诱使我们交换第 1 列和第  $k$  列, 将问题简化为  $a_{1,1} \neq 0$  的情况。然而, 当我们交换第 1 列和第  $k$  列时, 我们会改变其他列的顺序: 如果我们划掉第  $k$  列, 那么第 1 列将是剩余列中的第 1 列。但是, 如果我们交换第 1 列和第  $k$  列, 然后划掉第  $k$  列 (它现在是第 1 列), 那么第 1 列现在将是第  $k-1$  列。为了避免跟踪复杂的列交换, 我们可以像上面那样, 交换第  $k$  列和第  $k-1$  列, 将一切简化为我们在上一步处理的情况。这样的操作不会改变其余列的顺序。

所以

$$\det A = \sum_{k=1}^n (-1)^{1+k} a_{1,k} \det A_{1,k}.$$

为了得到第 2 行的展开式，我们可以交换第 1 行和第 2 行，然后应用上面的公式。行交换改变了符号，所以我们得到

$$\det A = - \sum_{k=1}^n (-1)^{1+k} a_{2,k} \det A_{2,k} = \sum_{k=1}^n (-1)^{2+k} a_{2,k} \det A_{2,k}.$$

通过交换第 3 行和第 2 行并按第 2 行展开，我们得到公式

$$\det A = \sum_{k=1}^n (-1)^{3+k} a_{3,k} \det A_{3,k},$$

依此类推。

要将行列式  $\det A$  按列展开，只需对  $A^T$  应用行展开公式即可。  $\square$

**定义** 这些数

$$C_{j,k} = (-1)^{j+k} \det A_{j,k}$$

称为  $A$  的**代数余子式** (cofactor)。

使用这个符号，在第  $j$  行展开行列式的公式可以重写为

$$\det A = a_{j,1}C_{j,1} + a_{j,2}C_{j,2} + \cdots + a_{j,n}C_{j,n} = \sum_{k=1}^n a_{j,k}C_{j,k}.$$

类似地，在第  $k$  列展开可以写成

$$\det A = a_{1,k}C_{1,k} + a_{2,k}C_{2,k} + \cdots + a_{n,k}C_{n,k} = \sum_{j=1}^n a_{j,k}C_{j,k}.$$

**注记** 代数余子式展开公式经常被用作行列式的定义。不难证明由该公式给出的数满足行列式的基本性质：归一化性质是显然的，反对称性的证明也很容易。然而，线性性质的证明虽然不难，但有点繁琐。

**注记** 虽然它看起来非常不错，但代数余子式展开公式不适用于直接计算大于  $3 \times 3$  的一般矩阵的行列式。

可以计算，它需要超过  $n!$  次乘法（准确地说，需要  $\sum_{k=2}^n n!/k!$  次乘法），而  $n!$  的增长非常快。例如，计算一个  $20 \times 20$  矩阵的代数余子式展开需要超过  $20! \approx 2.4 \times 10^{18}$  次乘法。一台每秒执行十亿次乘法的计算机需要 77 年才能执行  $20!$  次乘法；事实上，计算一个  $20 \times 20$  矩阵的代数余子式展开所需总运算将需要 132 多年的时间来完成。<sup>30</sup>

<sup>30</sup>读者可以自行验证这一数字，比如使用 WolframAlpha 软件。

另一方面, 使用行约简计算  $n \times n$  矩阵的行列式需要  $(n^3 + 2n - 3)/3$  次乘法 (以及大约相同数量的加法)。对于一台每秒执行一百万次运算 (按当前标准非常慢) 的计算机来说, 计算  $100 \times 100$  矩阵的行列式只需要一秒钟时间里的一小部分。

只有当某一行 (或列) 包含很多零项时, 代数余子式展开公式才会显得实用。

然而, 代数余子式展开公式具有重要的理论价值, 正如下一节所示。

## 5.1. 逆矩阵的代数余子式公式

由代数余子式  $C_{j,k} = (-1)^{j+k} \det A_{j,k}$  组成的矩阵  $C = \{C_{j,k}\}_{j,k=1}^n$  称为  $A$  的代数余子式矩阵 (cofactor matrix)。

**定理 5.2.** 设  $A$  是一个可逆矩阵, 并设  $C$  是它的代数余子式矩阵。<sup>31</sup> 那么

$$A^{-1} = \frac{1}{\det A} C^T$$

**证明:** 让我们计算乘积  $AC^T$ . 第  $j$  个对角项是通过将  $A$  的第  $j$  行与  $C$  的第  $j$  列 (即  $C^T$  的第  $j$  行) 相乘得到的, 所以根据代数余子式展开公式,

$$(AC^T)_{j,j} = a_{j,1}C_{j,1} + a_{j,2}C_{j,2} + \cdots + a_{j,n}C_{j,n} = \det A,$$

为了得到非对角项, 我们需要将  $A$  的第  $k$  行与  $C^T$  的第  $j$  列相乘,  $j \neq k$ , 得到

$$a_{k,1}C_{j,1} + a_{k,2}C_{j,2} + \cdots + a_{k,n}C_{j,n}.$$

根据代数余子式展开公式 (在第  $j$  行展开), 这是将  $A$  中第  $j$  行替换为第  $k$  行 (而所有其他行保持不变) 得到的矩阵的行列式。但是, 这个矩阵的第  $j$  行和第  $k$  行是相同的, 所以行列式为 0。因此,  $AC^T$  的所有非对角项都为零 (而所有对角项都等于  $\det A$ ), 所以

$$AC^T = (\det A)I.$$

这意味着矩阵  $\frac{1}{\det A} C^T$  是  $A$  的右逆, 由于  $A$  是方阵, 所以它是逆。□

回忆一下, 对于可逆矩阵  $A$ , 方程  $A\mathbf{x} = \mathbf{b}$  的解是

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A} C^T \mathbf{b},$$

我们得到以下定理的推论。

**推论 5.3 (Cramer 法则)** 对于可逆矩阵  $A$ , 方程  $A\mathbf{x} = \mathbf{b}$  的解的第  $k$  个项由以下公式给出:

$$x_k = \frac{\det B_k}{\det A},$$

其中矩阵  $B_k$  是通过将  $A$  的第  $k$  列替换为向量  $\mathbf{b}$  而得到的。

<sup>31</sup>译者注: 在国内, 一般记  $A^*$  为书中的  $C^T$ , 读作 “ $A$  的伴随”。但本书的  $A^*$  符号已经名花有主: 它在后面的章节 (始于第五章第 5 节) 中用于表示埃尔米特伴随 (Hermitian adjoint)。

## 5.2. 逆矩阵的代数余子式公式的应用

**例子 (求  $2 \times 2$  矩阵的逆)** 代数余子式公式在求  $2 \times 2$  矩阵

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

的逆时, 确实非常有用。代数余子式仅仅是  $A$  中的项 ( $1 \times 1$  矩阵), 代数余子式矩阵是

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

所以逆矩阵  $A^{-1}$  由公式给出:

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

虽然对于维数大于 3 的情况, 逆矩阵的代数余子式公式看起来并不实用, 但它的确具有重要的理论价值, 正如下面的例子所示。

**例子 (整数逆矩阵)** 假设我们想构造一个具有整数项的矩阵  $A$ , 使得其逆也具有整数项 (对这样的矩阵求逆可以出成一个很好的家庭作业: 你无需处理分数)。如果  $\det A = 1$  且其项是整数, 那么逆矩阵的代数余子式公式表明了  $A^{-1}$  也具有整数项。

注意, 构造一个  $\det A = 1$  的整数矩阵很容易: 可以从主对角线上为 1 的三角矩阵开始, 然后应用几次行或列替换 (第三类运算) 来使矩阵看起来是一般的。

**例子 (多项式矩阵的逆)** 另一个例子是考虑一个多项式矩阵 (polynomial matrix)  $A(x)$ , 即其项不是数字而是变量  $x$  的多项式  $a_{j,k}(x)$ . 如果  $\det A(x) \equiv 1$ , 那么逆矩阵  $A^{-1}(x)$  也是一个多项式矩阵。

如果  $\det A(x) = p(x) \neq 0$ , 则从代数余子式展开可知,  $p(x)$  是一个多项式, 因此  $A^{-1}(x)$  具有有理数项: 更重要的是,  $p(x)$  是每个分母的倍数。

### 练习

5.1. 你可以使用任何方法计算行列式:

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & 2 & -5 \\ 6 & 4 & -3 \end{vmatrix}, \quad \begin{vmatrix} 1 & -2 & 3 & -12 \\ -5 & 12 & -14 & 19 \\ -9 & 22 & -20 & 31 \\ -4 & 9 & -14 & 15 \end{vmatrix}.$$

5.2. 使用行 (列) 展开计算以下行列式。注意, 你没有必要从第 1 行 (列) 开始展开: 选择具有更多零的行 (列) 将简化你的计算。

$$\begin{vmatrix} 1 & 2 & 0 \\ 1 & 1 & 5 \\ 1 & -3 & 0 \end{vmatrix}, \quad \begin{vmatrix} 4 & -6 & -4 & 4 \\ 2 & 1 & 0 & 0 \\ 0 & -3 & 1 & 3 \\ -2 & 2 & -3 & -5 \end{vmatrix}.$$

5.3. 对于  $n \times n$  矩阵

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & a_0 \\ -1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & -1 & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & a_{n-2} \\ 0 & 0 & 0 & \cdots & -1 & a_{n-1} \end{pmatrix},$$

计算  $\det(A + tI)$ , 其中  $I$  是  $n \times n$  单位矩阵. 行展开和归纳可能是最好的方法. 这时你将得到一个涉及  $a_0, a_1, \dots, a_{n-1}$  和  $t$  的漂亮表达式.

5.4. 使用代数余子式公式计算下列矩阵的逆.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} 19 & -17 \\ 3 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 3 & 5 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

5.5. 设  $D_n$  是  $n \times n$  三对角矩阵

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

的行列式. 使用代数余子式展开证明  $D_n = D_{n-1} + D_{n-2}$ . 这表明数列  $D_n$  是斐波那契数列  $1, 2, 3, 5, 8, 13, 21, \dots$ .

5.6. 重新回顾范德蒙德行列式. 我们的目标是证明  $(n+1) \times (n+1)$  范德蒙德行列式的公式:

$$\begin{vmatrix} 1 & c_0 & c_0^2 & \cdots & c_0^n \\ 1 & c_1 & c_1^2 & \cdots & c_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_n & c_n^2 & \cdots & c_n^n \end{vmatrix} = \prod_{0 \leq j < k \leq n} (c_k - c_j).$$

我们将使用归纳法. 为此:

a) 验证公式对  $n = 1, n = 2$  成立.

b) 将最后一行中的变量  $c_n$  看作  $x$ , 并证明行列式是一个  $n$  次多项式  $A_0 + A_1x + A_2x^2 + \cdots + A_nx^n$ , 其中系数  $A_k$  由  $c_0, c_1, \dots, c_{n-1}$  决定.

c) 证明该多项式在  $x = c_0, c_1, \dots, c_{n-1}$  处均有零点, 因此可以表示为  $A_n \cdot (x - c_0)(x - c_1) \cdots (x - c_{n-1})$ , 其中  $A_n$  如 b) 中所述.

d) 假设范德蒙德行列式的公式对  $n-1$  成立, 计算  $A_n$  并证明对  $n$  的公式.

5.7. 使用代数余子式展开来计算  $n \times n$  矩阵的行列式需要多少次乘法? 证明这个公式.

## 6. 子式与秩

对于矩阵  $A$ , 让我们考虑它的  $k \times k$  子矩阵(submatrix), 它通过选取原矩阵中的  $k$  行和  $k$  列得到. 该矩阵的行列式称为  $k$  阶子式(minor). 注意, 一个  $m \times n$  矩阵有  $\binom{m}{k} \cdot \binom{n}{k}$  个不同的  $k \times k$  子矩阵, 因此它就有这么多个  $k$  阶子式.

**定理 6.1.** 对于一个非零矩阵  $A$ , 它的秩等于能使  $k$  阶非零子式存在的最大整数  $k$ .



**证明：**首先，让我们证明，如果  $k > \text{rank } A$ ，则所有  $k$  阶子式都为零。实际上，由于  $A$  的列空间  $\text{Ran } A$  的维数是  $\text{rank } A < k$ ，因此  $A$  中任何含个数为  $k$  的列的系统都是线性相关的。因此，对于  $A$  的任何  $k \times k$  子矩阵，它的列都是线性相关的，所以所有  $k$  阶子式都为零。

为了完成证明，我们需要证明存在一个非零的  $k$  阶子式，其中  $k = \text{rank } A$ 。可能存在许多这样的子式，但也许最简单的方法是取主元行和主元列（即原始矩阵中包含主元的行和列）。这个  $k \times k$  子矩阵具有与原始矩阵相同的主元，因此它是可逆的（每一列和每一行都有主元），并且其行列式非零。□

这个定理看起来不是很有用，因为进行行约简比计算所有子式要容易得多。然而，它同样具有重要的理论价值，正如以下推论所示。

**推论 6.2** 设  $A = A(x)$  是一个  $m \times n$  多项式矩阵（即其项是变量  $x$  的多项式）。那么  $\text{rank } A(x)$  在除了有限个可能的点之外的地方是恒定的，而在这些点上秩会变小。

**证明：**设  $r$  是满足至少存在一个  $x$  使得  $\text{rank } A(x) = r$  成立的最大整数。为了证明这样的  $r$  存在，我们首先尝试  $r = \min\{m, n\}$ 。如果确实存在一个  $x$  使得  $\text{rank } A(x) = r$ ，我们就找到了  $r$ 。如果不是，我们则将  $r$  替换为  $r - 1$  并重试。经过有限步操作，我们要么停止，要么得到 0。因此， $r$  是存在的。

设  $x_0$  是使得  $\text{rank } A(x_0) = r$  成立的一个点，并且设  $M$  是一个  $k$  阶子式，使得  $M(x_0) \neq 0$ 。由于  $M(x)$  是一个  $k \times k$  多项式矩阵的行列式，所以可以将  $M(x)$  看作是一个多项式。由于  $M(x_0) \neq 0$ ，它不是恒零的，因此它只能在有限个点处为零。所以，除了可能有限个点之外， $\text{rank } A(x) \geq r$ 。但是根据  $r$  的定义，对于所有的  $x$ ， $\text{rank } A(x) \leq r$ 。□

## 7. 第三章复习题

7.1. 判断正误：

- 行列式只对方阵有定义。
- 如果  $A$  的两行或两列相同，则  $\det A = 0$ 。
- 如果  $B$  是通过交换  $A$  的两行（或两列）得到的矩阵，则  $\det B = \det A$ 。
- 如果  $B$  是通过将  $A$  的某一行（列）乘以一个标量  $\alpha$  得到的矩阵，则  $\det B = \det A$ 。
- 如果  $B$  是通过将  $A$  的某一行乘以一个数加到另一行得到的矩阵，则  $\det B = \det A$ 。
- 三角矩阵的行列式是其对角线元素的乘积。
- $\det(A^T) = -\det(A)$ 。
- $\det(AB) = \det(A) \det(B)$ 。
- 矩阵  $A$  可逆当且仅当  $\det A \neq 0$ 。
- 如果  $A$  是可逆矩阵，则  $\det(A^{-1}) = 1/\det(A)$ 。

7.2. 设  $A$  是一个  $n \times n$  矩阵。 $\det(3A)$ ,  $\det(-A)$  和  $\det(A^2)$  与  $\det A$  的关系是什么？

7.3. 如果  $A$  和  $A^{-1}$  的所有元素都是整数, 那么  $\det A = 3$  是否可能?

**提示:**  $\det(A) \det(A^{-1})$  是什么?

7.4. 设  $\mathbf{v}_1, \mathbf{v}_2$  是  $\mathbb{R}^2$  中的向量, 然后设  $A$  是以  $\mathbf{v}_1, \mathbf{v}_2$  为列的  $2 \times 2$  矩阵。证明  $|\det A|$  是由向量  $\mathbf{v}_1, \mathbf{v}_2$  作为两邻边确定的平行四边形的面积。

首先考虑  $\mathbf{v}_1 = (x_1, 0)^T$  的情况。对于一般情况  $\mathbf{v}_1 = (x_1, y_1)^T$ , 左乘一个旋转矩阵, 将向量  $\mathbf{v}_1$  变换为  $(\tilde{x}_1, 0)^T$  来处理。

**提示:** 旋转矩阵的行列式是什么?

以下问题说明了行列式的符号与向量组的“定向”(orientation)之间的关系。

7.5. 设  $\mathbf{v}_1, \mathbf{v}_2$  是  $\mathbb{R}^2$  中的向量。证明  $D(\mathbf{v}_1, \mathbf{v}_2) > 0$  当且仅当存在一个旋转矩阵  $T_\alpha$  使得向量  $T_\alpha \mathbf{v}_1$  与  $\mathbf{e}_1$  平行 (并且方向相同), 且  $T_\alpha \mathbf{v}_2$  位于上半平面 ( $\mathbf{e}_2$  所在的半平面), 即分量  $x_2 > 0$ 。

**提示:** 同样地, 旋转矩阵的行列式是什么?

## 第四章 谱理论简介（特征值与特征向量）

谱理论是帮助我们理解线性算子结构的主要工具。在本章中，我们只考虑从一个向量空间映射到其自身的算子（或等价地说， $n \times n$  矩阵）。如果我们有一个线性变换  $A: V \rightarrow V$ ，我们可以容易地将其与自身相乘，取其任意幂，或任意多项式。

谱理论的主要思想是将算子分解为简单的块，并分别分析每个块。

为了解释其主要思想，让我们考虑**差分方程** (difference equations)。许多过程可以用以下类型的方程描述：

$$\mathbf{x}_{n+1} = A\mathbf{x}_n, \quad n = 0, 1, 2, \dots,$$

其中  $A: V \rightarrow V$  是一个线性变换，而  $\mathbf{x}_n$  是系统在时刻  $n$  的状态。给定初始状态  $\mathbf{x}_0$ ，我们希望知道时刻  $n$  的状态  $\mathbf{x}_n$ ，分析  $\mathbf{x}_n$  的长期行为等。<sup>32</sup>

乍一看，这个问题似乎很简单：解由公式  $\mathbf{x}_n = A^n \mathbf{x}_0$  给出。但如果  $n$  非常大呢：成千上万，数百万时，又会如何呢？或者如果我们想分析  $\mathbf{x}_n$  当  $n \rightarrow \infty$  时的行为呢？

这时**特征值**和**特征向量**的概念就出现了。假设  $A\mathbf{x}_0 = \lambda\mathbf{x}_0$ ，其中  $\lambda$  是某个标量。那么  $A^2\mathbf{x}_0 = \lambda^2\mathbf{x}_0$ ,  $A^3\mathbf{x}_0 = \lambda^3\mathbf{x}_0$ ,  $\dots$ ,  $A^n\mathbf{x}_0 = \lambda^n\mathbf{x}_0$ ，解的行为就能得到很好的解释。

在本章中，我们只考虑有限维空间中的算子。无限维空间中的谱理论要复杂得多，这里提出的结果大多在无限维情况下不成立。

### 1. 基本定义

#### 1.1. 特征值、特征向量、谱

标量  $\lambda$  被称为算子  $A: V \rightarrow V$  的**特征值**(eigenvalue)，如果存在一个**非零**向量  $\mathbf{v} \in V$  使得

$$A\mathbf{v} = \lambda\mathbf{v}.$$

向量  $\mathbf{v}$  被称为  $A$  的**特征向量**(eigenvector)（对应于特征值  $\lambda$ ）。

---

<sup>32</sup>差分方程是微分方程  $\mathbf{x}'(t) = A\mathbf{x}(t)$  的离散时间类似物。为了求解微分方程，需要计算  $e^{tA} := \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$ ，而谱理论也有助于完成此操作。

如果我们知道了  $\lambda$  是一个特征值, 那么寻找特征向量将很容易: 只需解方程  $A\mathbf{x} = \lambda\mathbf{x}$ , 或者等价地

$$(A - \lambda I)\mathbf{x} = 0.$$

所以, 找到对应于特征值  $\lambda$  的**所有**特征向量, 就是找到  $A - \lambda I$  的零空间。零空间  $\text{Ker}(A - \lambda I)$ , 即所有特征向量和零向量的集合, 被称为**特征子空间(eigenspace)**。

所有算子  $A$  的特征值集合被称为  $A$  的**谱(spectrum)**, 通常记作  $\sigma(A)$ 。

## 1.2. 寻找特征值: 特征多项式

标量  $\lambda$  是特征值当且仅当零空间  $\text{Ker}(A - \lambda I)$  非平凡 (因此方程  $(A - \lambda I)\mathbf{x} = 0$  有非平凡解)。

设  $A$  作用于  $\mathbb{F}^n$  (即  $A: \mathbb{F}^n \rightarrow \mathbb{F}^n$ )。由于  $A$  的矩阵是方阵,  $A - \lambda I$  有非平凡零空间当且仅当它不可逆。我们知道一个方阵不可逆当且仅当它的行列式为 0. 因此

$$\lambda \in \sigma(A), \text{ 即 } \lambda \text{ 是 } A \text{ 的特征值} \Leftrightarrow \det(A - \lambda I) = 0.$$

如果  $A$  是一个  $n \times n$  矩阵, 那么  $\det(A - \lambda I)$  是关于变量  $\lambda$  的  $n$  次多项式。这个多项式被称为  $A$  的**特征多项式 (characteristic polynomial)**。所以, 要找到  $A$  的所有特征值, 只需计算特征多项式并找到它所有的根。

用这种方法寻找算子的谱在更高维度下并不实用。求解高次多项式的根是一个非常困难的问题, 并且对于次数大于 4 的方程, 也不存在求根公式。所以, 在更高维度下, 我们通常使用很多不同的数值方法来寻找特征值和特征向量。

## 1.3. 寻找抽象算子的特征多项式和特征值

因此, 我们已经知道了如何找到矩阵的谱。但如何找到作用在抽象向量空间中的算子的特征值呢? 方法很简单:

$$\text{选取任意一组基, 然后计算该基下算子的矩阵的特征值。}$$

但我们如何知道这个结果不依赖于基的选取呢?

有几种可能的解释。一种是基于**相似矩阵**的概念。让我们回忆一下, 方阵  $A$  和  $B$  被称为相似的, 如果存在一个可逆矩阵  $S$  使得

$$A = SBS^{-1}.$$

注意, 相似矩阵的行列式是相等的。的确

$$\det A = \det(SBS^{-1}) = \det S \det B \det S^{-1} = \det B,$$

因为  $\det S^{-1} = 1/\det S$ . 注意, 如果  $A = SBS^{-1}$ , 那么

$$A - \lambda I = SBS^{-1} - \lambda SIS^{-1} = S(B - \lambda I)S^{-1},$$

所以矩阵  $A - \lambda I$  和  $B - \lambda I$  是相似的。因此

$$\det(A - \lambda I) = \det(B - \lambda I),$$

即

相似矩阵的特征多项式是相同的。

如果  $T: V \rightarrow V$  是一个线性变换，并且  $\mathcal{A}$  和  $\mathcal{B}$  是  $V$  中的两组基，那么

$$[T]_{\mathcal{A}\mathcal{A}} = [I]_{\mathcal{A}\mathcal{B}}[T]_{\mathcal{B}\mathcal{B}}[I]_{\mathcal{B}\mathcal{A}},$$

并且由于  $[I]_{\mathcal{B}\mathcal{A}} = ([I]_{\mathcal{A}\mathcal{B}})^{-1}$ ，所以矩阵  $[T]_{\mathcal{A}\mathcal{A}}$  和  $[T]_{\mathcal{B}\mathcal{B}}$  在不同基下是相似的。

换句话说，线性变换的矩阵在不同基下是相似的。

因此，我们可以将算子的特征多项式定义为它在某组基下的矩阵的特征多项式。正如我们上面讨论的，结果不依赖于基的选择，所以算子的特征多项式是良好定义的。

## 1.4. 复空间与实空间

代数基本定理断言，任何（至少一次）多项式都有一个复根。这意味着有限维复向量空间中的算子至少有一个特征值，因此它的谱是非空的。

另一方面，很容易在实向量空间中构造一个没有**实数**特征值的线性变换，例如在  $\mathbb{R}^2$  中旋转  $R_\alpha$ ,  $\alpha \neq k\pi$  ( $k \in \mathbb{Z}$ ) 就是一个例子。由于通常假设特征值应该属于标量域（如果一个算子作用在域  $\mathbb{F}$  上的向量空间中，则特征值应该在  $\mathbb{F}$  中），这样的算子具有空谱。

因此，复数情况（即作用在复向量空间中的算子）似乎是谱理论最自然的环境。由于  $\mathbb{R} \subset \mathbb{C}$ ，我们可以始终将实数  $n \times n$  矩阵视为  $\mathbb{C}^n$  中的算子，以允许复数特征值。将实数矩阵视为  $\mathbb{C}^n$  中的算子在谱理论中是很典型的，我们也将遵循这个约定。寻找矩阵的特征值（除非另有说明）将始终意味着寻找所有**复数**特征值，而不是仅限于实数特征值。

注意，抽象实向量空间中的算子也可以解释为复空间中的算子。一种朴素的方法是固定一组基（本章所有空间都是有限维的），然后在该基下使用坐标，允许使用复数坐标：这本质上是从实数矩阵到具有复数特征值的算子的过程。

这种构造描述了所谓的**复化** (complexification)，结果不依赖于基的选择。后面第 5 章第 8.2 节将给出复化的“高明”抽象构造，解释为什么结果不依赖于基的选择。

## 1.5. 特征值的重数

提醒读者, 如果  $p$  是一个多项式, 而  $\lambda$  是它的一个根 (即  $p(\lambda) = 0$ ), 那么  $z - \lambda$  整除  $p(z)$ , 即  $p$  可以表示为  $p(z) = (z - \lambda)q(z)$ , 其中  $q$  是某个多项式。如果  $q(\lambda) = 0$ , 那么  $q$  也可以被  $z - \lambda$  整除, 所以  $(z - \lambda)^2$  整除  $p$  等等。

能够整除  $p(z)$  的  $(z - \lambda)$  的最大正整数  $k$  被称为根  $\lambda$  的**重数**(multiplicity)。

如果  $\lambda$  是算子 (矩阵)  $A$  的一个特征值, 那么它就是特征多项式  $p(z) = \det(A - zI)$  的一个根。这个根的重数被称为特征值  $\lambda$  的**(代数) 重数**。

次数为  $n$  的任何多项式  $p(z) = \sum_{k=0}^n a_k z^k$  恰好有  $n$  个复数根, **计入重数**。计入重数的意思是, 如果一个根的重数是  $d$ , 我们就必须列出 (计数) 它  $d$  次。换句话说,  $p$  可以表示为

$$p(z) = a_n(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n),$$

其中  $\lambda_1, \lambda_2, \dots, \lambda_n$  是它的复数根, 计入重数。

还有另一种关于特征值重数的概念: 特征子空间  $\text{Ker}(A - \lambda I)$  的维数被称为特征值  $\lambda$  的**几何重数**。

几何重数不像代数重数那样被广泛使用。所以, 当人们简单地说“重数”时, 他们通常指的是**代数重数**。

我们在此顺便提一下, 特征值的代数重数和几何重数可能不同。

**命题 1.1** 特征值的几何重数不能超过其代数重数。

**证明:** 见下面的练习 1.9。 □

## 1.6. 迹与行列式

**定理 1.2.** 设  $A$  是一个  $n \times n$  矩阵, 设  $\lambda_1, \lambda_2, \dots, \lambda_n$  是它的 (复数) 特征值 (计入重数)。那么

$$1. \text{trace } A = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

$$2. \det A = \lambda_1 \lambda_2 \cdots \lambda_n.$$

**证明:** 见下面的练习 1.10, 1.11。 □

## 1.7. 三角矩阵的特征值

计算特征值等价于寻找矩阵的特征多项式的根 (或使用某种数值方法), 这可能非常耗时。然而, 有一个特殊情况, 在这种情况下我们可以直接从矩阵中读出特征值。即,

三角矩阵的特征值（计入重数）正是对角线上的元素  $a_{1,1}, a_{2,2}, \dots, a_{n,n}$ .

这里三角矩阵是指上三角或下三角矩阵。由于对角矩阵是三角矩阵的一个特例（它既是上三角也是下三角），所以

对角矩阵的特征值是其对角线元素.

其证明是显然的：我们需要从  $A$  的对角元素中减去  $\lambda$ ，并利用三角矩阵的行列式是其对角线元素乘积这一事实。由此我们得到特征多项式

$$\det(A - \lambda I) = (a_{1,1} - \lambda)(a_{2,2} - \lambda) \dots (a_{n,n} - \lambda),$$

其根正是  $a_{1,1}, a_{2,2}, \dots, a_{n,n}$ .

### 练习

#### 1.1. 判断正误：

- a) 每个  $n$  维向量空间中的线性算子都有  $n$  个不同的特征值；
- b) 如果一个矩阵只有一个特征向量，那么它有无限多个特征向量；
- c) 存在一个实数方阵，它没有实数特征值；
- d) 存在一个方阵，它没有（复数）特征向量；
- e) 相似矩阵总是具有相同的特征值；
- f) 相似矩阵总是具有相同的特征向量；
- g) 矩阵  $A$  的两个特征向量之和（非零）总是  $A$  的特征向量；
- h) 对应于同一特征值  $\lambda$  的矩阵  $A$  的两个特征向量之和总是算子  $A$  的特征向量。

#### 1.2. 找出以下矩阵的特征多项式、特征值和特征向量：

$$\begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}.$$

#### 1.3. 计算旋转矩阵

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

的特征值和特征向量。注意，特征值（和特征向量）不一定必须是实数。

#### 1.4. 计算以下矩阵的特征多项式和特征值：

$$\begin{pmatrix} 1 & 2 & 5 & 67 \\ 0 & 2 & 3 & 6 \\ 0 & 0 & -2 & 5 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 0 & 2 \\ 0 & \pi & 43 & 2 \\ 0 & 0 & 16 & 1 \\ 0 & 0 & 0 & 54 \end{pmatrix}, \quad \begin{pmatrix} 4 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 4 & e & 0 \\ 3 & 3 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 4 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 3 & 3 & 1 & 1 \end{pmatrix}.$$

不要展开特征多项式，将其保留为乘积形式即可。

#### 1.5. 证明三角矩阵的特征值（计入重数）与其对角线元素相等。

#### 1.6. 称算子 $A$ 为**幂零** (nilpotent) 的，如果 $\exists k \in \mathbb{N}_+$ ，使得 $A^k = \mathbf{0}$ 成立。证明如果 $A$ 是幂零的，那么 $\sigma(A) = \{0\}$ （即 0 是 $A$ 的唯一特征值）。

#### 1.7. 证明分块三角矩阵

$$\begin{pmatrix} A & * \\ \mathbf{0} & B \end{pmatrix}$$

的特征多项式（其中  $A$  和  $B$  是方阵）与  $\det(A - \lambda I) \det(B - \lambda I)$  相等。（使用第三章的练习 3.11）。

- 1.8. 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是向量空间  $V$  中的一组基。并假设基的前  $k$  个向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  是算子  $A$  的特征向量, 对应于特征值  $\lambda$  (即  $A\mathbf{v}_j = \lambda\mathbf{v}_j, j = 1, 2, \dots, k$ )。证明在该基下, 算子  $A$  的矩阵具有分块三角形形式

$$\begin{pmatrix} \lambda I_k & * \\ \mathbf{0} & B \end{pmatrix},$$

其中  $I_k$  是  $k \times k$  的单位矩阵,  $B$  是某个  $(n-k) \times (n-k)$  矩阵。

- 1.9. 使用前面两个练习中的结论来证明一个特征值的几何重数不能超过其代数重数。

- 1.10. 证明矩阵  $A$  的行列式是其特征值的乘积 (计重数)。

**提示:** 首先证明  $\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$ , 其中  $\lambda_1, \lambda_2, \dots, \lambda_n$  是特征值 (计重数)。然后比较常数项 (不含  $\lambda$  的项) 或代入  $\lambda = 0$  来得出结论。

- 1.11. 分三步证明矩阵的迹等于特征值之和。

首先, 计算等式

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

右侧  $\lambda^{n-1}$  的系数。然后证明  $\det(A - \lambda I)$  可以表示为

$$\det(A - \lambda I) = (a_{1,1} - \lambda)(a_{2,2} - \lambda) \dots (a_{n,n} - \lambda) + q(\lambda),$$

其中  $q(\lambda)$  是一个最高为  $n-2$  次的多项式。最后, 通过比较  $\lambda^{n-1}$  的系数来得出结论。

## 2. 对角化

谱理论的一个应用是对算子的**对角化** (diagonalization), 即给定一个算子, 可以找到一组基, 使得该算子在该基下的矩阵是对角矩阵。这样的基并非总能找到, 也就是说, 并非所有算子都能对角化 (都是可对角化的)。算子可对角化的重要性在于, 对角矩阵的幂以及更一般的函数很容易计算。因此如果我们对一个算子进行对角化, 我们就可以轻松地计算它的函数。

我们将在本节中解释如何计算可对角化算子的函数。我们还将给出算子可对角化的充要条件, 以及一些简单的充分条件。

此外, 对于  $\mathbb{F}^n$  中的算子 (矩阵),  $A$  的可对角化性质意味着它可以表示为  $A = SDS^{-1}$ , 其中  $D$  是一个对角矩阵,  $S$  是一个可逆矩阵; 我们将在稍后解释这一点。

除非另有说明, 本节中的所有结果对于复数和实数向量空间 (甚至对于任意域  $\mathbb{F}$  上的向量空间) 都成立。

### 2.1. 预备知识

假设向量空间  $V$  中的算子  $A$  具有一组由  $A$  的特征向量组成的基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ , 其中  $\lambda_1, \lambda_2, \dots, \lambda_n$  是相应的特征值。那么  $A$  在此基下的矩阵是对角矩阵, 对角线



上的项是  $\lambda_1, \lambda_2, \dots, \lambda_n$ , 其余位置都是 0, 省略不写:

$$(2.1) \quad [A]_{\mathcal{B}\mathcal{B}} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

另一方面, 如果一个算子  $A$  在基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  下的矩阵由 (2.1) 给出, 那么显然  $A\mathbf{b}_k = \lambda_k \mathbf{b}_k$ , 即  $\lambda_k$  是特征值,  $\mathbf{b}_k$  是相应的特征向量。

再次注意, 上述推理对于复数和实数向量空间 (甚至对于任意域  $\mathbb{F}$  上的向量空间) 都成立。

将上述推理应用于  $\mathbb{F}^n$  中的算子 (矩阵), 我们立即得到以下定理。注意, 虽然本书中  $\mathbb{F}$  是  $\mathbb{C}$  或  $\mathbb{R}$ , 但本定理对任意域  $\mathbb{F}$  都成立。

**定理 2.1.** 一个矩阵  $A$  (每项的值都在  $\mathbb{F}$  内) 可以表示为  $A = SDS^{-1}$ , 其中  $D$  是一个对角矩阵,  $S$  是一个可逆矩阵 (两者项中的值都在  $\mathbb{F}$  内), 当且仅当  $\mathbb{F}^n$  存在一个由  $A$  的特征向量组成的基。

而且, 在这种情况下,  $D$  的对角线上的项都是特征值, 而  $S$  的列是相应的特征向量 (第  $k$  列对应于  $D$  的第  $k$  个对角线元素)。

**证明:** 设  $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ , 并设  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  是  $S$  的列 (注意, 由于  $S$  可逆, 它的列构成了  $\mathbb{F}^n$  的一组基)。

那么恒等式  $A = SDS^{-1}$  意味着  $D = S^{-1}AS = [I]_{S,\mathcal{B}}A[I]_{\mathcal{B},S}$ , 其中  $S = [I]_{S,\mathcal{B}}$  是从  $\mathcal{B}$  到标准基  $S$  的坐标变换矩阵。这正好意味着  $D = [A]_{\mathcal{B},\mathcal{B}}$ 。

正如我们上面讨论的, 当且仅当  $\lambda_k$  是  $A$  的特征值,  $\mathbf{b}_k$  是  $A$  的相应特征向量时,  $[A]_{\mathcal{B},\mathcal{B}} = D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ .  $\square$

**注记** 注意, 如果一个矩阵允许表示为  $A = SDS^{-1}$  并且  $D$  是一个对角矩阵, 那么通过直接计算就可以表明,  $S$  的列是  $A$  的特征向量, 而  $D$  的对角线元素是相应的特征值。这为定理 2.1 中相应陈述提供了另一种证明。

正如我们上面讨论的, 一个可对角化的算子  $A: V \rightarrow V$  恰好有  $n = \dim V$  个特征值 (计入重数); 一个复向量空间中的算子也恰好有  $n$  个特征值 (计入重数); 另一方面, 一个实数空间中的算子可能没有实数特征值。

我们将遵循谱理论中的惯例, 将实数矩阵视为  $\mathbb{C}^n$  中的算子, 从而允许复数特征值和特征向量。除非另有说明, 我们将把对矩阵的复数对角化简单说成对角化, 即  $A = SDS^{-1}$  中, 矩阵  $S$  和  $D$  的项可以是复数。

一个实数矩阵何时可以实数对角化 ( $A = SDS^{-1}$ , 其中  $S$  和  $D$  都是实数矩阵) 的问题, 实际上是一个非常简单的问题, 见下面的定理 2.9.

## 2.2. 一些动机: 算子函数

设一个算子  $A$  在基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  下的矩阵是 (2.1) 中给出的对角矩阵。那么很容易找到算子  $A$  的  $N$  次幂。即, 在基  $\mathcal{B}$  下  $A^N$  的矩阵是

$$[A^N]_{\mathcal{B}\mathcal{B}} = \text{diag}\{\lambda_1^N, \lambda_2^N, \dots, \lambda_n^N\} = \begin{pmatrix} \lambda_1^N & & & \\ & \lambda_2^N & & \\ & & \ddots & \\ & & & \lambda_n^N \end{pmatrix}.$$

而且, 算子的函数也相对容易计算: 例如, 算子 (矩阵) 指数  $e^{tA}$  定义为  $e^{tA} = I + tA + \frac{t^2 A^2}{2!} + \frac{t^3 A^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$ , 并且它在基  $\mathcal{B}$  下的矩阵是

$$[e^{tA}]_{\mathcal{B}\mathcal{B}} = \text{diag}\{e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}\} = \begin{pmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_n t} \end{pmatrix}.$$

设  $A$  是  $\mathbb{F}^n$  中的一个算子。为了在标准基  $\mathcal{S}$  下找到算子  $A^N$  和  $e^{tA}$  的矩阵, 我们需要回忆一下坐标变换矩阵  $[I]_{\mathcal{S}\mathcal{B}}$  是一个以  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  为列的矩阵。设这个矩阵为  $S$ , 那么根据坐标变换公式我们得到

$$A = [A]_{\mathcal{S}\mathcal{S}} = S \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} S^{-1} = SDS^{-1},$$

其中我们使用  $D$  来表示中间的对角矩阵。

类似地,

$$A^N = SD^N S^{-1} = S \begin{pmatrix} \lambda_1^N & & & \\ & \lambda_2^N & & \\ & & \ddots & \\ & & & \lambda_n^N \end{pmatrix} S^{-1},$$

并且对于  $e^{tA}$  也类似。

另一种求可对角化算子的幂 (或其他函数) 的方式是, 注意到, 如果算子  $A$  可以表示为  $A = SDS^{-1}$ , 那么

$$A^N = \underbrace{(SDS^{-1})(SDS^{-1}) \dots (SDS^{-1})}_{N \text{ times}} = SD^N S^{-1}.$$

于是, 计算对角矩阵的  $N$  次幂就很容易了。

### 2.3. $n$ 个不同特征值的情况

我们现在给出一个算子可对角化的非常简单的**充分条件**, 见下面的推论 2.3。

**定理 2.2.** 设  $\lambda_1, \lambda_2, \dots, \lambda_r$  是  $A$  的不同特征值, 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是相应的特征向量。那么向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是线性无关的。

**证明:** 我们将使用数学归纳法处理  $r$ .  $r = 1$  的情况是平凡的, 因为根据定义, 特征向量是非零的, 并且由一个非零向量组成的系统是线性无关的。

假设定理的陈述对  $r - 1$  是正确的。如果存在一个非平凡线性组合

$$(2.2) \quad c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_r \mathbf{v}_r = \sum_{k=1}^r c_k \mathbf{v}_k = \mathbf{0}.$$

将  $(A - \lambda_r I)$  作用于 (2.2) 并利用  $(A - \lambda_r I)\mathbf{v}_r = \mathbf{0}$  的事实, 我们得到

$$\sum_{k=1}^{r-1} c_k (\lambda_k - \lambda_r) \mathbf{v}_k = \mathbf{0}.$$

根据归纳假设, 向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r-1}$  是线性无关的, 所以  $c_k (\lambda_k - \lambda_r) = 0$  对于  $k = 1, 2, \dots, r - 1$  成立。由于  $\lambda_k \neq \lambda_r$ , 我们可以得出  $c_k = 0$  对于  $k < r$  成立。然后从 (2.2) 可知  $c_r = 0$ , 也就是说我们得到了平凡线性组合。□

**推论 2.3** 如果一个算子  $A: V \rightarrow V$  恰好有  $n = \dim V$  个不同的特征值, 那么它是可对角化的。

**证明:** 对于每个特征值  $\lambda_k$ , 设  $\mathbf{v}_k$  是一个相应的特征向量 (对每个特征值只选取一个特征向量)。根据定理 2.2, 系统  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是线性无关的, 并且由于它恰好包含  $n = \dim V$  个向量, 所以它是一组基。□

### 2.4. 子空间的基 (又名子空间的直和)

为了描述可对角化算子, 我们需要引入一些新定义。

设  $V_1, V_2, \dots, V_p$  是向量空间  $V$  的子空间。我们说子空间的系统是  $V$  的一组基, 如果任何向量  $\mathbf{v} \in V$  都存在唯一的表示:

$$(2.3) \quad \mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_p = \sum_{k=1}^p \mathbf{v}_k, \quad \mathbf{v}_k \in V_k.$$

我们也说, 子空间的系统  $\{V_1, V_2, \dots, V_p\}$  是线性无关的, 如果方程

$$\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_p = \mathbf{0}, \quad \mathbf{v}_k \in V_k$$

只有平凡解 ( $\mathbf{v}_k = \mathbf{0}, \forall k = 1, 2, \dots, p$ )。

另一种表述方式是, 子空间的系统  $\{V_1, V_2, \dots, V_p\}$  是线性无关的, 当且仅当任何由非零向量  $\mathbf{v}_k$  ( $\mathbf{v}_k \in V_k$ ) 组成的系统是线性无关的。

我们说子空间的系统  $\{V_1, V_2, \dots, V_p\}$  是生成 (或完备, 或张成) 的, 如果任何向量  $\mathbf{v} \in V$  可以表示为 (2.3) (不一定是唯一的)。

**注记 2.4** 从上述定义可以立即看出, 定理 2.2 实际上表明, 算子  $A$  的特征子空间

$$E_k := \text{Ker}(A - \lambda_k I), \quad \lambda_k \in \sigma(A)$$

的系统是线性无关的。

**注记 2.5** 很容易看出, 与向量基类似, 子空间的系统  $\{V_1, V_2, \dots, V_p\}$  是一组基当且仅当它既是生成集又是线性无关的。我们将此事实的证明留给读者作为练习。

有一个子空间基的简单例子。设  $V$  是一个向量空间, 有一组基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ . 将下标集  $\{1, 2, \dots, n\}$  分为  $p$  个子集  $\Lambda_1, \Lambda_2, \dots, \Lambda_p$ , 并定义子空间  $V_k := \text{span}\{\mathbf{v}_j : j \in \Lambda_k\}$ . 显然, 子空间  $V_k$  构成  $V$  的一组基。

下面的定理表明, 在有限维情况下, 这是子空间基唯一可能的例子。

**定理 2.6.** 设  $\{V_1, V_2, \dots, V_p\}$  是一些子空间的基, 并且在每个子空间  $V_k$  中都有一组基 (向量基)  $\mathcal{B}_k$ .<sup>33</sup> 那么这些基的并集  $\cup_k \mathcal{B}_k$  是  $V$  的一组基。

为了证明定理, 我们需要以下引理:

**引理 2.7.** 设  $\{V_1, V_2, \dots, V_p\}$  是一个线性无关的子空间族, 并且在每个子空间  $V_k$  中都有一个线性无关的向量系统  $\mathcal{B}_k$ .<sup>34</sup> 那么这些基的并集  $\mathcal{B} := \cup_k \mathcal{B}_k$  是一个线性无关的系统。

**证明:** 如果稍微思考一下, 引理的证明几乎是平凡的。书写证明的主要困难在于选取合适的记号。为了避免使用两个下标 (一个表示  $k$ , 另一个表示  $\mathcal{B}_k$  中向量的编号), 让我们使用 “扁平化” 的记号。

即, 设  $n$  是  $\mathcal{B} := \cup_k \mathcal{B}_k$  中向量的数量。让我们对  $\mathcal{B}$  中的向量集进行排序, 例如如下: 首先列出  $\mathcal{B}_1$  中的所有向量, 然后是  $\mathcal{B}_2$  中的所有向量, 依此类推, 最后列出  $\mathcal{B}_p$  中的所有向量。

这样, 我们将  $\mathcal{B}$  中的所有向量用整数  $1, 2, \dots, n$  下标, 并且下标集  $\{1, 2, \dots, n\}$  被分成集合  $\Lambda_1, \Lambda_2, \dots, \Lambda_p$ , 使得集合  $\mathcal{B}_k$  由向量  $\{\mathbf{b}_j : j \in \Lambda_k\}$  组成。

<sup>33</sup> 我们不具体列出  $\mathcal{B}_k$  中的向量, 只需记住每个  $\mathcal{B}_k$  都包含有限数量的向量。

<sup>34</sup> 同样, 这里我们不单独命名  $\mathcal{B}_k$  中的每个向量, 我们只是记住每个集合  $\mathcal{B}_k$  都包含有限数量的向量。

假设我们有一个非平凡的线性组合

$$(2.4) \quad c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n = \sum_{j=1}^n c_j \mathbf{b}_j = \mathbf{0}.$$

设

$$\mathbf{v}_k := \sum_{j \in \Lambda_k} c_j \mathbf{b}_j.$$

那么 (2.4) 可以重写为

$$\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_p = \mathbf{0}.$$

由于  $\mathbf{v}_k \in V_k$  并且子空间  $\{V_1, V_2, \dots, V_p\}$  是线性无关的, 所以  $\mathbf{v}_k = \mathbf{0} \forall k$ . 这意味着对于每个  $k$ ,

$$\sum_{j \in \Lambda_k} c_j \mathbf{b}_j = \mathbf{0},$$

并且由于向量系统  $\{\mathbf{b}_j : j \in \Lambda_k\}$  (即系统  $B_k$ ) 是线性无关的, 我们得到  $c_j = 0$  对于所有  $j \in \Lambda_k$ . 由于这对所有  $\Lambda_k$  都成立, 我们可以得出  $c_j = 0$  对于所有  $j$ .  $\square$

**定理 2.6 的证明** 为了证明定理, 我们将使用与引理 2.7 证明相同的记号, 即系统  $B_k$  由向量  $\{\mathbf{b}_j : j \in \Lambda_k\}$  组成。

引理 2.7 断言向量系统  $\{\mathbf{b}_j : j = 1, 2, \dots, n\}$  是线性无关的, 所以剩下的事情就是证明这个系统是完备的。

由于子空间系统  $\{V_1, V_2, \dots, V_p\}$  是一组基, 任何向量  $\mathbf{v} \in V$  可以表示为

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_p = \sum_{k=1}^p \mathbf{v}_k, \quad \mathbf{v}_k \in V_k.$$

由于向量  $\{\mathbf{b}_j : j \in \Lambda_k\}$  构成了  $V_k$  的基, 向量  $\mathbf{v}_k$  可以表示为

$$\mathbf{v}_k = \sum_{j \in \Lambda_k} c_j \mathbf{b}_j.$$

因此,  $\mathbf{v} = \sum_{j=1}^n c_j \mathbf{b}_j$ .  $\square$

## 2.5. 可对角化判据

首先, 让我们回顾一个要用到的条件。由于对角矩阵  $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  的特征值 (计入重数) 恰好是  $\lambda_1, \lambda_2, \dots, \lambda_n$ , 我们发现, 如果一个算子  $A : V \rightarrow V$  是可对角化的, 它恰好有  $n = \dim V$  个特征值 (计入重数)。

下面的定理对实数和复数向量空间都成立 (甚至对任意域上的空间也成立)。

**定理 2.8.** 设一个算子  $A : V \rightarrow V$  恰好有  $n = \dim V$  个特征值 (计入重数)。<sup>35</sup>

<sup>35</sup> 由于任何复向量空间中的算子都恰好有  $n$  个特征值 (计入重数), 因此在复数情况下, 此假设是多余的。

那么  $A$  是可对角化的当且仅当对于每个特征值  $\lambda$ , 特征子空间  $\text{Ker}(A - \lambda I)$  的维数 (即几何重数) 等于  $\lambda$  的代数重数。

**证明:** 首先, 我们注意到, 对于一个对角矩阵, 特征值的代数重数和几何重数是相等的, 因此对于可对角化算子也是如此。

现在我们来证明另一个蕴含关系。设  $\lambda_1, \lambda_2, \dots, \lambda_p$  是  $A$  的特征值, 设  $E_k := \text{Ker}(A - \lambda_k I)$  是相应的特征子空间。根据注记 2.4, 子空间  $E_k, k = 1, 2, \dots, p$  是线性无关的。

设  $\mathcal{B}_k$  是  $E_k$  的一组基。根据引理 2.7, 向量系统  $\mathcal{B} := \cup_k \mathcal{B}_k$  是一个线性无关系统。

我们知道每个  $\mathcal{B}_k$  由  $\dim E_k$  (即  $\lambda_k$  的重数) 个向量组成。所以  $\mathcal{B}$  中的向量数量等于特征值  $\lambda_k$  的重数之和。但是特征值重数之和就是计入重数的特征值数量, 这恰好是  $n = \dim V$ 。因此, 我们得到了一个含  $n = \dim V$  个线性无关的特征向量组成的系统, 这意味着它是一组基。  $\square$

## 2.6. 实数分解

下面的定理实际上已经证明过了 (它本质上是定理 2.8 在实数空间上的情况)。我们在此陈述只是为了总结实数矩阵可实数对角化的情形。

**定理 2.9.** 一个  $n \times n$  实数矩阵  $A$  允许实数分解 (即表示为  $A = SDS^{-1}$ , 其中  $S$  和  $D$  是实数矩阵,  $D$  是对角矩阵且  $S$  可逆) 当且仅当它允许复数分解并且  $A$  的所有特征值都是实数。

## 2.7. 一些例子

### 2.7.1. 实数特征值

考虑矩阵

$$A = \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}.$$

它的特征多项式等于

$$\begin{vmatrix} 1 - \lambda & 2 \\ 8 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - 16,$$

其根 (特征值) 是  $\lambda = 5$  和  $\lambda = -3$ 。对于特征值  $\lambda = 5$ ,

$$A - 5I = \begin{pmatrix} 1 - 5 & 2 \\ 8 & 1 - 5 \end{pmatrix} = \begin{pmatrix} -4 & 2 \\ 8 & -4 \end{pmatrix}.$$

其零空间的基由向量  $(1, 2)^T$  构成, 所以这是对应的特征向量。

类似地, 对于  $\lambda = -3$ ,

$$A - \lambda I = A + 3I = \begin{pmatrix} 1+3 & 2 \\ 8 & 1+3 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 8 & 4 \end{pmatrix}.$$

$\text{Ker}(A + 3I)$  的零空间由向量  $(1, -2)^T$  张成, 所以这是对应的特征向量。矩阵  $A$  可以被对角化为

$$A = \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}^{-1}.$$

### 2.7.2. 复数特征值

考虑矩阵

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

其特征多项式是

$$\begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 + 4,$$

特征值 (特征多项式的根) 是  $\lambda = 1 \pm i$ . 对于  $\lambda = 1 + i$ ,

$$A - \lambda I = \begin{pmatrix} 1 - (1+i) & 2 \\ -2 & 1 - (1+i) \end{pmatrix} = \begin{pmatrix} -i & 2 \\ -2 & -i \end{pmatrix}.$$

这个矩阵的秩是 1, 所以特征子空间  $\text{Ker}(A - \lambda I)$  由一个向量, 例如  $(1, i)^T$  张成。

由于矩阵  $A$  是实数的, 我们不需要计算  $\lambda = 1 - i$  的特征向量: 通过取上述特征向量的复共轭, 我们可以自动获得它, 见下面的练习 2.2。所以, 对于  $\lambda = 1 - 2i$ , 一个相应的特征向量是  $(1, -i)^T$ , 因此矩阵  $A$  可以被对角化为

$$A = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} 1+2i & 0 \\ 0 & 1-2i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}^{-1}.$$

### 2.7.3. 一个不可对角化的矩阵

考虑矩阵

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

其特征多项式是

$$\begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2,$$

所以  $A$  有一个重数为 2 的特征值 1。然而, 很容易看出  $\dim \text{Ker}(A - I) = 1$  (有 1 个主元, 所以有  $2 - 1 = 1$  个自由变量)。因此, 特征值 1 的几何重数与其代数重数不同, 所以  $A$  是不可对角化的。

还有一个不使用定理 2.8 的解释。即, 我们得到的特征子空间  $\text{Ker}(A - I)$  是一维的 (由向量  $(1, 0)^T$  张成)。如果  $A$  是可对角化的, 那么它将在某组基下具有对角形式  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ , <sup>36</sup> 因此特征子空间的维数将是 2, 矛盾。所以  $A$  不能被对角化。

### 练习

2.1. 设  $A$  是  $n \times n$  矩阵。判断正误:

- $A^T$  与  $A$  具有相同的特征值。
  - $A^T$  与  $A$  具有相同的特征向量。
  - 如果  $A$  是可对角化的, 那么  $A^T$  也是可对角化的。
- 证明你的结论。

2.2. 设  $A$  是一个实数方阵,  $\lambda$  是它的一个复数特征值。假设  $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$  是一个相应的特征向量,  $A\mathbf{v} = \lambda\mathbf{v}$ . 证明  $\bar{\lambda}$  是  $A$  的一个特征值, 并且  $\bar{\mathbf{v}}$  是  $A$  的相应特征向量。这里  $\bar{\mathbf{v}}$  是向量  $\mathbf{v}$  的复共轭,  $\bar{\mathbf{v}} := (\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n)^T$ 。

2.3. 设

$$A = \begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}.$$

通过对  $A$  进行对角化, 求出  $A^{2004}$ 。

2.4. 构建一个特征值为 1 和 3, 相应特征向量为  $(1, 2)^T$  和  $(1, 1)^T$  的矩阵  $A$ . 这样的矩阵是唯一的吗?

2.5. 如果可以, 对以下矩阵进行对角化:

- $\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}$ .
- $\begin{pmatrix} -1 & -1 \\ 6 & 4 \end{pmatrix}$ .
- $\begin{pmatrix} -2 & 2 & 6 \\ 5 & 1 & -6 \\ -5 & 2 & 9 \end{pmatrix}$  (提示:  $\lambda = 2$  是其中一个特征值)

2.6. 考虑矩阵

$$A = \begin{pmatrix} 2 & 6 & -6 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{pmatrix}.$$

- 求它的特征值。在不计算的情况下能否直接观察出特征值?
- 这个矩阵可对角化吗? 在不进行计算的情况下找出答案。
- 如果矩阵可对角化, 请对其进行对角化。

2.7. 对下面的矩阵进行对角化:

$$\begin{pmatrix} 2 & 0 & 6 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{pmatrix}$$

2.8. 求矩阵

$$A = \begin{pmatrix} 5 & 2 \\ -3 & 0 \end{pmatrix}$$

的所有平方根, 即求所有满足  $B^2 = A$  的矩阵  $B$ . **提示:** 对角矩阵的平方根很容易求。你可以将答案留作乘积形式。

---

<sup>36</sup>注意, 唯一具有某种基下矩阵为  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  的线性变换是恒等变换  $I$ . 由于  $A$  肯定不是恒等变换, 我们可以立即得出  $A$  不能被对角化, 所以计算特征子空间的维数是没有必要的。



- 2.9. 回顾一下著名的斐波那契数列:  $0, 1, 1, 2, 3, 5, 8, 13, 21, \dots$ , 它由以下方式定义: 令  $\phi_0 = 0$ ,  $\phi_1 = 1$ , 并定义

$$\phi_{n+2} = \phi_{n+1} + \phi_n.$$

我们想找到  $\phi_n$  的一个公式。可以按照以下步骤来:

- a) 找到一个  $2 \times 2$  矩阵  $A$ , 使得

$$\begin{pmatrix} \phi_{n+2} \\ \phi_{n+1} \end{pmatrix} = A \begin{pmatrix} \phi_{n+1} \\ \phi_n \end{pmatrix}.$$

**提示:** 结合平凡方程  $\phi_{n+1} = \phi_{n+1}$  和斐波那契关系  $\phi_{n+2} = \phi_{n+1} + \phi_n$ .

- b) 对  $A$  进行对角化, 并找到关于  $A^n$  的一个式子。

- c) 注意到

$$\begin{pmatrix} \phi_{n+1} \\ \phi_n \end{pmatrix} = A^n \begin{pmatrix} \phi_1 \\ \phi_0 \end{pmatrix} = A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

找到关于  $\phi_n$  的一个式子。(你需要计算一个逆矩阵并进行乘法运算。)

- d) 证明向量  $(\phi_{n+1}/\phi_n, 1)^T$  收敛到一个  $A$  的特征向量。你认为这是一个巧合吗?

- 2.10. 设  $A$  是一个  $5 \times 5$  矩阵, 且仅有 3 个特征值 (不计重数)。假设我们知道其中一个特征子空间是三维的。你能说  $A$  是否可对角化吗?
- 2.11. 给出一个  $3 \times 3$  矩阵的例子, 它不能被对角化。在构造了矩阵之后, 你能使它“一般”一些, 使得矩阵的特殊结构不明显吗?
- 2.12. 设一个非零矩阵  $A$  满足  $A^5 = 0$ 。证明  $A$  不能被对角化。更一般地说, 任何非零幂零矩阵, 即满足  $\exists N \in \mathbb{N}_+$ , 使得  $A^N = 0$  成立的矩阵, 都不能被对角化。
- 2.13. 转置的特征值:
- a) 考虑  $2 \times 2$  矩阵空间  $M_{2 \times 2}$  上的变换  $T(A) = A^T$ 。找出它所有的特征值和特征向量。这个变换可能被对角化吗?
- 提示:** 虽然可以写出这个线性变换在某组基下的矩阵, 计算特征多项式等等, 但直接从定义中找出特征值和特征向量会更容易。
- b) 在  $n \times n$  矩阵空间中, 能否做同样的事情?
- 2.14. 证明两个子空间  $V_1$  和  $V_2$  是线性无关的当且仅当  $V_1 \cap V_2 = \{0\}$ 。



## 第五章 内积空间

内积空间的理论仅针对实空间和复空间进行发展，因此本章中的  $\mathbb{F}$  始终代表  $\mathbb{R}$  或  $\mathbb{C}$ ；其结果通常不能推广到任意域上的向量空间。

本章中的大部分内容在实数和复数情况下都成立（且表述相同）。在实数和复数情况存在差异的罕见情况下，我将明确说明所考虑的情况：否则，所有内容对两种情况都适用。

最后，当公式对复数和实数情况都适用时，我们将使用复数情况的公式；在实数情况下，这些公式也能给出正确的结果，尽管有时会显得稍微复杂一些。

### 1. $\mathbb{R}^n$ 与 $\mathbb{C}^n$ 中的内积 内积空间

#### 1.1. $\mathbb{R}^n$ 中的内积和范数

在二维和三维空间中，我们通过勾股定理定义了向量  $\mathbf{x}$  的长度（即其终点到原点的距离），例如在  $\mathbb{R}^3$  中，向量的长度定义为

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

自然地，我们可以将这个公式推广到所有正整数  $n$ ，定义向量  $\mathbf{x} \in \mathbb{R}^n$  的**范数**(norm)为

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

这里的“范数”一词只是“长度”一词的一种更专业的说法。

在  $\mathbb{R}^3$  中，我们定义了**点积** (dot product) 为  $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3$ ，其中  $\mathbf{x} = (x_1, x_2, x_3)^T$ ,  $\mathbf{y} = (y_1, y_2, y_3)^T$ 。

类似地，在  $\mathbb{R}^n$  中，可以定义两个向量  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  的**内积** (inner product)<sup>37</sup>（记作  $(\mathbf{x}, \mathbf{y})$ ）为：

$$(\mathbf{x}, \mathbf{y}) := x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T \mathbf{x}$$

因此， $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ 。

注意， $\mathbf{y}^T \mathbf{x} = \mathbf{x}^T \mathbf{y}$ ，我们使用  $\mathbf{y}^T \mathbf{x}$  的记法只是为了保持一致性。

---

<sup>37</sup>虽然  $\mathbf{x} \cdot \mathbf{y}$  的记法和“点积”的术语都经常用于表示内积，但在下文中我们将会看到，为什么我更喜欢  $(\mathbf{x}, \mathbf{y})$  的记法。

## 1.2. $\mathbb{C}^n$ 中的内积和范数

现在我们来定义  $\mathbb{C}^n$  的范数和内积。正如我们之前所见，从谱理论的角度来看，复数空间  $\mathbb{C}^n$  是最自然的。即使我们从具有实系数的矩阵（或实向量空间上的算子）开始，特征值也可能是复数，这时我们就需要在一个复数空间中进行运算。

对于复数  $z = x + iy$ ，我们有  $|z|^2 = x^2 + y^2 = z\bar{z}$ 。如果复数  $\mathbf{z} \in \mathbb{C}^n$  表示为：

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} x_1 + iy_1 \\ x_2 + iy_2 \\ \vdots \\ x_n + iy_n \end{pmatrix}$$

那么可以自然地定义其**范数**  $\|\mathbf{z}\|$  为：

$$\|\mathbf{z}\|^2 = \sum_{k=1}^n (x_k^2 + y_k^2) = \sum_{k=1}^n |z_k|^2.$$

我们尝试定义一个  $\mathbb{C}^n$  上的内积，使得  $\|\mathbf{z}\|^2 = (\mathbf{z}, \mathbf{z})$ 。一种选择是定义  $(\mathbf{z}, \mathbf{w})$  为：

$$(\mathbf{z}, \mathbf{w}) = z_1 \bar{w}_1 + z_2 \bar{w}_2 + \cdots + z_n \bar{w}_n = \sum_{k=1}^n z_k \bar{w}_k,$$

我们将此定义为  $\mathbb{C}^n$  中的**标准内积**。

为了简化记法，让我们引入一个新概念。对于矩阵  $A$ ，我们定义其**埃尔米特伴随** (Hermitian adjoint)（或简称**伴随**） $A^*$  为  $A^* = \overline{A^T}$ ，这意味着我们对矩阵进行转置，然后取每个元素的复共轭（注：如果让这两个步骤颠倒也可以得到一样的结果）。注意，对于实数矩阵  $A$ ，有  $A^* = A^T$ 。

使用  $A^*$  的概念，我们可以将  $\mathbb{C}^n$  中的标准内积写成

$$(\mathbf{z}, \mathbf{w}) = \mathbf{w}^* \mathbf{z}.$$

**注记** 很容易看出，我们也可以定义一个形式上与之不同的  $\mathbb{C}^n$  内积，使得  $\|\mathbf{z}\|^2 = (\mathbf{z}, \mathbf{z})$ ，即内积定义为

$$(\mathbf{z}, \mathbf{w})_1 = \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n = \mathbf{z}^T \mathbf{w}.$$

我们还没有明确说明希望内积满足什么性质，但  $\mathbf{w}^* \mathbf{z}$  和  $\mathbf{z}^T \mathbf{w}$  是给出  $\|\mathbf{z}\|^2 = (\mathbf{z}, \mathbf{z})$  的唯一合理的选择。

注意，上述两个内积的选择本质上是等价的：它们之间的唯一区别在于记法，因为  $(\mathbf{z}, \mathbf{w})_1 = (\mathbf{w}, \mathbf{z})$ 。

虽然第二个内积选择看起来更自然，但第一个内积  $(\mathbf{z}, \mathbf{w}) = \mathbf{w}^* \mathbf{z}$  使用更广泛，因此我们将采用它。

### 1.3. 内积空间

我们在  $\mathbb{R}^n$  和  $\mathbb{C}^n$  中定义的内积满足以下性质：

1. **(共轭) 对称性**： $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$ ；注意，对于实数空间，此性质就是对称性， $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$ ；
2. **线性性质**： $(\alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z}) = \alpha(\mathbf{x}, \mathbf{z}) + \beta(\mathbf{y}, \mathbf{z})$  对所有向量  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  和所有标量  $\alpha, \beta$  成立；
3. **非负性**： $(\mathbf{x}, \mathbf{x}) \geq 0 \quad \forall \mathbf{x}$  成立；
4. **非退化性**： $(\mathbf{x}, \mathbf{x}) = 0$  当且仅当  $\mathbf{x} = \mathbf{0}$ 。

设  $V$  是一个（复数或实数）向量空间。**内积**是这样一个函数，它为每对向量  $\mathbf{x}, \mathbf{y}$  分配一个标量，记作  $(\mathbf{x}, \mathbf{y})$ ，并满足上述性质 1-4。

注意，对于实数空间  $V$ ，我们假定  $(\mathbf{x}, \mathbf{y})$  始终是实数；对于复数空间，内积  $(\mathbf{x}, \mathbf{y})$  可以是复数。

一个带有内积的向量空间  $V$  被称为**内积空间(inner product space)**。给定一个内积空间，我们可以定义其上的范数

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}.$$

#### 1.3.1. 例子

**例子 1.1.** 设  $V$  为  $\mathbb{R}^n$  或  $\mathbb{C}^n$ 。我们已经定义了内积  $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^* \mathbf{x} = \sum_{k=1}^n x_k \overline{y_k}$ 。

这个内积称为  $\mathbb{R}^n$  或  $\mathbb{C}^n$  中的**标准内积** (standard inner product)。

我们将符号  $\mathbb{F}$  用来同时表示  $\mathbb{C}$  和  $\mathbb{R}$ 。当我们在  $\mathbb{F}^n$  上有一些陈述时，这意味着该陈述对  $\mathbb{R}^n$  和  $\mathbb{C}^n$  都成立。

**例子 1.2.** 设  $V$  为次数最多为  $n$  的多项式空间  $\mathbb{P}_n$ 。定义内积为

$$(f, g) = \int_{-1}^1 f(t) \overline{g(t)} dt.$$

很容易验证上述性质 1-4 得到满足。

这个定义对复数和实数情况都有效。在实数情况下，我们只允许有实系数多项式，并且不需要取复共轭。

我们回顾一下，对于方阵  $A$ ，其**迹**定义为对角线元素之和，

$$\text{trace } A := \sum_{k=1}^n a_{k,k}.$$

**例子 1.3.** 对于  $m \times n$  矩阵空间  $M_{m \times n}$ ，我们定义所谓的**弗罗贝尼乌斯内积**

(Frobenius inner product) 为

$$(A, B) = \text{trace}(B^* A).$$

再次, 很容易验证性质 1-4 得到满足, 即我们确实定义了一个内积。

注意,

$$\text{trace}(B^* A) = \sum_{j,k} A_{j,k} \overline{B_{j,k}}.$$

这意味着此内积与  $\mathbb{C}^{mn}$  中的标准内积一致。

## 1.4. 内积的性质

我们在本节中得到的陈述对任何抽象域上的内积空间都是成立的, 而不仅仅是  $\mathbb{F}^n$ . 为了证明它们, 我们仅使用内积的性质 1-4.

首先, 我们注意到性质 1 和 2 暗示了:

$$2'. (\mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z}) = \overline{\alpha}(\mathbf{x}, \mathbf{y}) + \overline{\beta}(\mathbf{x}, \mathbf{z})$$

的确, 对于复数空间:

$$(\mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z}) = \overline{(\alpha\mathbf{y} + \beta\mathbf{z}, \mathbf{x})} = \overline{\alpha(\mathbf{y}, \mathbf{x}) + \beta(\mathbf{z}, \mathbf{x})} = \overline{\alpha}(\overline{(\mathbf{y}, \mathbf{x})}) + \overline{\beta}(\overline{(\mathbf{z}, \mathbf{x})}) = \overline{\alpha}(\mathbf{x}, \mathbf{y}) + \overline{\beta}(\mathbf{x}, \mathbf{z})$$

同时还注意到性质 2 暗示了对于所有向量  $\mathbf{x}$ :

$$(\mathbf{0}, \mathbf{x}) = (\mathbf{x}, \mathbf{0}) = 0.$$

**引理 1.4.** 设  $\mathbf{x}$  是内积空间  $V$  中的一个向量。则  $\mathbf{x} = \mathbf{0}$  当且仅当

$$(1.1) \quad (\mathbf{x}, \mathbf{y}) = 0 \quad \forall \mathbf{y} \in V.$$

**证明:** 由于  $(\mathbf{0}, \mathbf{y}) = 0$ , 我们只需要证明  $(1.1) \quad \forall \mathbf{y} \in V$  蕴含  $\mathbf{x} = \mathbf{0}$ . 将  $\mathbf{y} = \mathbf{x}$  代入 (1.1), 我们得到  $(\mathbf{x}, \mathbf{x}) = 0$ , 因此  $\mathbf{x} = \mathbf{0}$ .  $\square$

将上述引理应用于差值  $\mathbf{x} - \mathbf{y}$ , 我们得到以下:

**推论 1.5.** 设  $\mathbf{x}, \mathbf{y}$  是内积空间  $V$  中的向量。则  $\mathbf{x} = \mathbf{y}$  当且仅当

$$(\mathbf{x}, \mathbf{z}) = (\mathbf{y}, \mathbf{z}) \quad \forall \mathbf{z} \in V$$

成立。

以下推论非常简单, 但将被大量使用:

**推论 1.6.** 假设两个算子  $A, B: X \rightarrow Y$  满足

$$(A\mathbf{x}, \mathbf{y}) = (B\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x} \in X, \forall \mathbf{y} \in Y.$$

则  $A = B$ .

**证明：**根据上一个推论（固定  $\mathbf{x}$  并考虑所有可能的  $\mathbf{y}$ ），我们得到  $A\mathbf{x} = B\mathbf{x}$ 。由于这对所有  $\mathbf{x} \in X$  都成立，因此算子  $A$  和  $B$  是相同的。  $\square$

以下性质将范数和内积联系起来。

**定理 1.7.** (柯西-施瓦茨不等式)

$$|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

**证明：**我们将给出的证明不是最短的，但它显示了证明主要思想的来源。

我们先考虑实数情况。如果  $\mathbf{y} = \mathbf{0}$ ，则该陈述是平凡的，因此我们可以假设  $\mathbf{y} \neq \mathbf{0}$ 。根据内积的性质，对于所有标量  $t$ ，有：

$$0 \leq \|\mathbf{x} - t\mathbf{y}\|^2 = (\mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y}) = \|\mathbf{x}\|^2 - 2t(\mathbf{x}, \mathbf{y}) + t^2\|\mathbf{y}\|^2.$$

特别地，该不等式对于  $t = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2}$  应该成立，<sup>38</sup> 并且在该点该不等式变为：

$$0 \leq \|\mathbf{x}\|^2 - 2 \frac{(\mathbf{x}, \mathbf{y})^2}{\|\mathbf{y}\|^2} + \frac{(\mathbf{x}, \mathbf{y})^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{(\mathbf{x}, \mathbf{y})^2}{\|\mathbf{y}\|^2},$$

这正是我们需要证明的不等式。

处理复数情况有几种可能的方法。一种方法是将  $\mathbf{x}$  替换为  $\alpha\mathbf{x}$ ，其中  $\alpha$  是一个模为 1 的复常数，使得  $(\alpha\mathbf{x}, \mathbf{y})$  为实数，然后对实数情况重复证明。

另一种可能性是再次考虑

$$\begin{aligned} 0 \leq \|\mathbf{x} - t\mathbf{y}\|^2 &= (\mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y}) = (\mathbf{x}, \mathbf{x} - t\mathbf{y}) - t(\mathbf{y}, \mathbf{x} - t\mathbf{y}) \\ &= \|\mathbf{x}\|^2 - t(\mathbf{y}, \mathbf{x}) - \bar{t}(\mathbf{x}, \mathbf{y}) + |t|^2\|\mathbf{y}\|^2. \end{aligned}$$

将  $t = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2} = \frac{\overline{(\mathbf{y}, \mathbf{x})}}{\|\mathbf{y}\|^2}$  代入该不等式，我们得到：

$$0 \leq \|\mathbf{x}\|^2 - \frac{|(\mathbf{x}, \mathbf{y})|^2}{\|\mathbf{y}\|^2}$$

这就是我们想要的 $\square$ 不等式。

注意，上述段落实际上是定理的完整形式证明。在此之前的推理只是为了解释为什么我们需要选择这个特定的  $t$  值。

柯西-施瓦茨不等式的直接推论是以下引理。

**引理 1.8.** (三角不等式)

对于内积空间  $V$  中的任意向量  $\mathbf{x}, \mathbf{y}$ ，有

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

<sup>38</sup> 这是上述二次多项式取得最小值的点：例如，可以通过对  $t$  取导数并令其等于 0 来计算。

$$\begin{aligned}
\text{证明: } \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) \\
&\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + |(\mathbf{x}, \mathbf{y})| + |(\mathbf{y}, \mathbf{x})| = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2|(\mathbf{x}, \mathbf{y})| \\
&\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.
\end{aligned}$$

□

以下**极化恒等式** (polarization identities) 允许我们从范数重构内积:

**引理 1.9.** (极化恒等式) 对于  $\mathbf{x}, \mathbf{y} \in V$ :

$$(\mathbf{x}, \mathbf{y}) = \frac{1}{4}(\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2)$$

如果  $V$  是一个实内积空间,

$$(\mathbf{x}, \mathbf{y}) = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|\mathbf{x} + \alpha \mathbf{y}\|^2$$

如果  $V$  是一个复内积空间。

该引理可以通过直接计算来证明。我们把它留给读者作为练习。

范数在内积空间中的另一个重要性质也可以通过直接计算来验证。

**引理 1.10.** (平行四边形恒等式) 对于任意向量  $\mathbf{u}, \mathbf{v}$ :

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2).$$

在二维空间中, 这个引理将平行四边形的边与其对角线的长度联系起来, 这就是其名称的由来。它是平面几何中一个众所周知的结论。

## 1.5. 范数 赋范空间

我们之前已经证明, 范数  $\|\mathbf{v}\|$  满足以下性质:

1. **齐次性**:  $\|\alpha \mathbf{v}\| = |\alpha| \cdot \|\mathbf{v}\|$  对所有向量  $\mathbf{v}$  和所有标量  $\alpha$  成立。
2. **三角不等式**:  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ 。
3. **非负性**:  $\|\mathbf{v}\| \geq 0$  对所有向量  $\mathbf{v}$  成立。
4. **非退化性**:  $\|\mathbf{v}\| = 0$  当且仅当  $\mathbf{v} = \mathbf{0}$ 。

假设在一个向量空间  $V$  中, 我们为每个向量  $\mathbf{v}$  分配一个数字  $\|\mathbf{v}\|$ , 并且它满足上述性质 1-4。那么我们称函数  $\mathbf{v} \mapsto \|\mathbf{v}\|$  是一个**范数**。一个配备了范数的向量空间  $V$  被称为**赋范空间**(normed space)。

任何内积空间都是赋范空间, 因为范数  $\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}$  满足上述性质 1-4。然而, 也存在许多其他的赋范空间。例如, 给定  $p$ ,  $1 \leq p < \infty$ , 我们可以在  $\mathbb{R}^n$  或  $\mathbb{C}^n$  上定义范数  $\|\cdot\|_p$  为:

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p} = \left( \sum_{k=1}^n |x_k|^p \right)^{1/p}.$$



我们还可以定义范数  $\|\cdot\|_\infty$  ( $p = \infty$ ) 为:

$$\|\mathbf{x}\|_\infty = \max\{|x_k| : k = 1, 2, \dots, n\}.$$

当  $p = 2$  时, 范数  $\|\cdot\|_2$  与由内积得到的常规范数一致。

为了验证  $\|\cdot\|_p$  确实是一个范数, 我们需要验证它满足上述所有性质 1-4。性质 1, 3, 4 非常容易验证, 我们将其留作读者练习。当  $p = 1$  和  $p = \infty$  时, 三角不等式很容易验证 (并且我们在前面已经对  $p = 2$  证明了)。

对于所有其他的  $p$ , 可以验证三角不等式都是成立的, 但证明并不简单, 我们在此不予呈现。 $\|\cdot\|_p$  的三角不等式甚至有一个特殊的名字: 它被称为**闵可夫斯基不等式**, 以德国数学家 H. Minkowski 的名字命名。

注意, 当  $p \neq 2$  时, 范数  $\|\cdot\|_p$  不能由内积得到。很容易看出, 该范数不是由  $\mathbb{R}^n$  ( $\mathbb{C}^n$ ) 中的**标准**内积得到的。但我们将声称一个更强的结论! 我们声称**不可能**引入一个产生范数  $\|\cdot\|_p$ ,  $p \neq 2$  的内积。

这个陈述实际上很容易证明。根据引理 1.10, 任何由内积产生的范数都必须满足平行四边形恒等式。很容易看出, 当  $p \neq 2$  时, 平行四边形恒等式对范数  $\|\cdot\|_p$  不成立, 我们可以在  $\mathbb{R}^2$  中很容易地找到一个反例, 这进而会在所有其他空间中产生反例。

事实上, 如下面的定理所述, 平行四边形恒等式完全刻画了由内积产生的范数。

**定理 1.11.** 赋范空间中的一个范数当且仅当它满足平行四边形恒等式

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) \quad \forall \mathbf{u}, \mathbf{v} \in V$$

时, 才能由某个内积产生。

引理 1.10 断言, 由内积产生的范数满足平行四边形恒等式。

反向蕴含更复杂。如果我们给出一个范数, 并且该范数来自一个内积, 那么没有余地; 这个内积必须由极化恒等式给出 (见引理 1.9)。但是, 我们需要证明 (我们从极化恒等式得到的)  $(\mathbf{x}, \mathbf{y})$  确实是一个内积, 即它满足内积的所有性质。

事实上, 可以验证, 如果范数满足平行四边形恒等式, 那么从极化恒等式得到的  $(\mathbf{x}, \mathbf{y})$  确实是一个内积 (即满足内积的所有性质)。然而, 证明有点过于复杂, 我们在此不予呈现。

## 练习

### 1.1. 计算

$$(3 + 2i)(5 - 3i), \quad \frac{2-3i}{1-2i}, \quad \operatorname{Re}\left(\frac{2-3i}{1-2i}\right), \quad (1 + 2i)^3, \quad \operatorname{Im}((1 + 2i)^3).$$

### 1.2. 对于向量 $\mathbf{x} = (1, 2i, 1 + i)^T$ 和 $\mathbf{y} = (i, 2 - i, 3)^T$ , 计算:

$$\text{a) } (\mathbf{x}, \mathbf{y}), \quad \|\mathbf{x}\|^2, \quad \|\mathbf{y}\|^2, \quad \|\mathbf{y}\|;$$

b)  $(3\mathbf{x}, 2i\mathbf{y}), (2\mathbf{x}, i\mathbf{x} + 2\mathbf{y})$ ;

c)  $\|\mathbf{x} + 2\mathbf{y}\|$ .

注记: 完成部分 a) 后, 你可以不实际计算所有向量, 而只通过使用内积的性质来完成部分 b) 和 c)。

1.3. 设  $\|\mathbf{u}\| = 2, \|\mathbf{v}\| = 3, (\mathbf{u}, \mathbf{v}) = 2 + i$ . 计算

$$\|\mathbf{u} + \mathbf{v}\|^2, \|\mathbf{u} - \mathbf{v}\|^2, (\mathbf{u} + \mathbf{v}, \mathbf{u} - i\mathbf{v}), (\mathbf{u} + 3i\mathbf{v}, 4i\mathbf{u}).$$

1.4. 证明在内积空间中, 对于向量  $\mathbf{x}, \mathbf{y}$  有

$$\|\mathbf{x} \pm \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \pm 2\operatorname{Re}(\mathbf{x}, \mathbf{y}).$$

回忆  $\operatorname{Re} z = \frac{1}{2}(z + \bar{z})$ .

1.5. 解释为什么下面每个定义都不是给定向量空间上的内积:

a)  $(\mathbf{x}, \mathbf{y}) = x_1y_1 - x_2y_2$  在  $\mathbb{R}^2$  上;

b)  $(A, B) = \operatorname{trace}(A + B)$  在  $2 \times 2$  实矩阵空间上;

c)  $(f, g) = \int_0^1 f'(t)g(t)dt$  在多项式空间上;  $f'(t)$  表示导数。

1.6. (柯西-施瓦茨不等式中的等号) 证明

$$|(\mathbf{x}, \mathbf{y})| = \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

当且仅当其中一个向量是另一个向量的倍数。

提示: 分析柯西-施瓦茨不等式的证明。

1.7. 证明内积空间  $V$  中的平行四边形恒等式:

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2).$$

1.8. 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是内积空间  $V$  中的一个生成集 (特别地, 一组基)。证明:

a) 如果  $(\mathbf{x}, \mathbf{v}) = 0 \quad \forall \mathbf{v} \in V$  成立, 则  $\mathbf{x} = \mathbf{0}$ ;

b) 如果  $(\mathbf{x}, \mathbf{v}_k) = 0 \quad \forall k$ , 则  $\mathbf{x} = \mathbf{0}$ ;

c) 如果  $(\mathbf{x}, \mathbf{v}_k) = (\mathbf{y}, \mathbf{v}_k) \quad \forall k$ , 则  $\mathbf{x} = \mathbf{y}$ .

1.9. 考虑范数  $\|\cdot\|_p$  (在 1.5 节中引入) 的  $\mathbb{R}^2$  空间。对于  $p = 1, 2, \infty$ , 在范数  $\|\cdot\|_p$  下绘制 “单位球”  $B_p$ :

$$B_p := \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_p \leq 1\}.$$

你能猜测其他  $p$  的球  $B_p$  是什么样的吗?

## 2. 正交性 正交与标准正交基

**定义 2.1.** 两个向量  $\mathbf{u}$  和  $\mathbf{v}$  被称为**正交** (orthogonal) (也称为**垂直** (perpendicular)) 如果  $(\mathbf{u}, \mathbf{v}) = 0$ . 我们将用  $\mathbf{u} \perp \mathbf{v}$  来表示向量是正交的。

注意, 对于正交向量  $\mathbf{u}$  和  $\mathbf{v}$ , 我们有以下所谓的**勾股定理** (Pythagorean identity):

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad \text{如果 } \mathbf{u} \perp \mathbf{v}.$$

证明是简单直接的计算:

$$\|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) = (\mathbf{u}, \mathbf{u}) + (\mathbf{v}, \mathbf{v}) + (\mathbf{u}, \mathbf{v}) + (\mathbf{v}, \mathbf{u}) = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

(( $\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u}) = 0$  是因为正交性)。

**定义 2.2.** 我们说向量  $\mathbf{v}$  正交于子空间  $E$ , 如果  $\mathbf{v}$  正交于  $E$  中的所有向量  $\mathbf{w}$ .

我们说子空间  $E$  和  $F$  是正交的, 如果  $E$  中的所有向量都正交于  $F$ , 即  $E$  中的所有向量都正交于  $F$  中的所有向量。

以下引理显示了如何检查一个向量是否正交于一个子空间。

**引理 2.3.** 设  $E$  由向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  生成。则  $\mathbf{v} \perp E$  当且仅当

$$\mathbf{v} \perp \mathbf{v}_k \quad \forall k = 1, 2, \dots, r$$

成立。

**证明:** 根据定义, 如果  $\mathbf{v} \perp E$ , 则  $\mathbf{v}$  正交于  $E$  中的所有向量。特别地,  $\mathbf{v} \perp \mathbf{v}_k$  对  $k = 1, 2, \dots, r$  成立。

另一方面, 设  $\mathbf{v} \perp \mathbf{v}_k$  对  $k = 1, 2, \dots, r$  成立。由于向量  $\mathbf{v}_k$  生成  $E$ , 则  $E$  中的任何向量  $\mathbf{w}$  都可以表示为线性组合  $\mathbf{w} = \sum_{k=1}^r \alpha_k \mathbf{v}_k$ 。那么

$$(\mathbf{v}, \mathbf{w}) = \sum_{k=1}^r \alpha_k (\mathbf{v}, \mathbf{v}_k) = 0,$$

因此,  $\mathbf{v} \perp \mathbf{w}$ . □

**定义 2.4.** 向量组  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  被称为**正交的**, 如果任意两个向量相互正交 (即  $(\mathbf{v}_j, \mathbf{v}_k) = 0$  当  $j \neq k$  时)。

如果此外  $\|\mathbf{v}_k\| = 1 \quad \forall k$  成立, 则我们称该向量组为**标准正交** (orthonormal) 的<sup>39</sup>

**引理 2.5.** (广义勾股定理) 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是一个正交向量组。则:

$$\left\| \sum_{k=1}^n \alpha_k \mathbf{v}_k \right\|^2 = \sum_{k=1}^n |\alpha_k|^2 \|\mathbf{v}_k\|^2$$

当  $\|\mathbf{v}_k\| = 1$  时, 该公式显得特别简单。

**引理 2.5 的证明**

$$\left\| \sum_{k=1}^n \alpha_k \mathbf{v}_k \right\|^2 = \left( \sum_{k=1}^n \alpha_k \mathbf{v}_k, \sum_{j=1}^n \alpha_j \mathbf{v}_j \right) = \sum_{k=1}^n \sum_{j=1}^n \alpha_k \overline{\alpha_j} (\mathbf{v}_k, \mathbf{v}_j)$$

由于正交性, 当  $j \neq k$  时,  $(\mathbf{v}_k, \mathbf{v}_j) = 0$ 。因此, 我们只需要对  $j = k$  的项求和, 这就得到了:

$$\sum_{k=1}^n |\alpha_k|^2 (\mathbf{v}_k, \mathbf{v}_k) = \sum_{k=1}^n |\alpha_k|^2 \|\mathbf{v}_k\|^2.$$

□

**推论 2.6.** 任何一个不含零向量的正交向量组  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是线性无关的。

<sup>39</sup>译者注: 有些书上也称作“么正”的。

**证明：**假设存在  $\alpha_1, \alpha_2, \dots, \alpha_n$ , 使得  $\sum_{k=1}^n \alpha_k \mathbf{v}_k = \mathbf{0}$ . 那么根据广义勾股定理 (引理 2.5), 有

$$0 = \|\mathbf{0}\|^2 = \sum_{k=1}^n |\alpha_k|^2 \|\mathbf{v}_k\|^2.$$

由于  $\|\mathbf{v}_k\| \neq 0$  (因为  $\mathbf{v}_k \neq \mathbf{0}$ ), 我们得出

$$\alpha_k = 0 \quad \forall k$$

成立, 因此只有平凡线性组合才得到  $\mathbf{0}$ . □

**注记** 在后续讨论中, 我们通常将“正交向量组”的含义理解为“不含零向量的正交向量组”。由于零向量  $\mathbf{0}$  正交于一切向量, 它可以随时添加到任何正交向量组中, 但考虑含有零向量的正交向量组并没有太大意义。

## 2.1. 正交和标准正交基

**定义 2.7.** 一个既正交 (或标准正交), 同时又是基的向量组, 称为正交 (或标准正交) 基。

在  $\dim V = n$  的情况下, 任何包含  $n$  个非零向量的正交向量组显然是一组正交基。

正如我们之前研究过的, 要找到向量在某组基下的坐标, 需要解一个线性方程组。然而, 对于一组正交基, 找到向量的坐标要容易得多。具体来说, 假设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是一组正交基, 且

$$\mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \sum_{j=1}^n \alpha_j \mathbf{v}_j.$$

将方程两边分别与  $\mathbf{v}_1$  作内积, 我们得到

$$(\mathbf{x}, \mathbf{v}_1) = \left( \sum_{j=1}^n \alpha_j \mathbf{v}_j, \mathbf{v}_1 \right) = \sum_{j=1}^n \alpha_j (\mathbf{v}_j, \mathbf{v}_1) = \alpha_1 (\mathbf{v}_1, \mathbf{v}_1) = \alpha_1 \|\mathbf{v}_1\|^2$$

(所有内积  $(\mathbf{v}_j, \mathbf{v}_1) = 0$  当  $j \neq 1$  时)。因此,

$$\alpha_1 = \frac{(\mathbf{x}, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2}.$$

类似地, 将两边分别与  $\mathbf{v}_k$  作内积, 我们得到

$$(\mathbf{x}, \mathbf{v}_k) = \left( \sum_{j=1}^n \alpha_j \mathbf{v}_j, \mathbf{v}_k \right) = \sum_{j=1}^n \alpha_j (\mathbf{v}_j, \mathbf{v}_k) = \alpha_k (\mathbf{v}_k, \mathbf{v}_k) = \alpha_k \|\mathbf{v}_k\|^2,$$

因此:

$$(2.1) \quad \alpha_k = \frac{(\mathbf{x}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}.$$

因此,

要找到向量在一组正交基下的坐标, 无需求解线性方程组, 坐标可由公式 (2.1) 确定。

当  $\|\mathbf{v}_k\| = 1$  时, 这个公式对于标准正交基尤为简单。

具体来说, 如果  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是一组标准正交基, 则任何向量  $\mathbf{v}$  可以表示为:

$$(2.2) \quad \mathbf{v} = \sum_{k=1}^n (\mathbf{v}, \mathbf{v}_k) \mathbf{v}_k.$$

这个公式有时被称为(一个简化的版本(a baby version))**抽象正交傅里叶分解**(abstract orthogonal Fourier decomposition)。经典的(非抽象)傅里叶分解处理的是具体的标准正交系统(正弦和余弦, 或复指数)。我们将这个公式称为“简化版本”, 因为真实的傅里叶分解处理的是无限的标准正交系统。

**注记 2.8.** 标准正交基的重要性在于, 如果我们固定了内积空间  $V$  中的一组标准正交基, 我们就可以像处理  $\mathbb{F}^n$  中的向量一样处理该基下的坐标。具体来说, 正如本书开头所讨论的(参见第一章的注记 2.4), 如果我们有一个具有基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的向量空间  $V$  (在域  $\mathbb{F}$  上), 那么我们可以通过处理基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  下坐标向量列的标准向量运算(向量加法和标量乘法), 以完全相同的方式进行。<sup>40</sup>

如练习 2.3 所示, 如果我们有一内积空间  $V$  中的**标准正交基**, 我们可以通过取这些坐标向量列并计算这些向量列在  $\mathbb{C}^n$  或  $\mathbb{R}^n$  中的标准内积来计算  $V$  中两个向量的内积。

如下文第 3 节所示, 任何有限维内积空间都有一组标准正交基。因此, 在有限维情况下, 标准内积空间  $\mathbb{C}^n$  (或实数情况下的  $\mathbb{R}^n$ ) 实质上是唯一的有限维内积空间。

### 练习

2.1. 找出  $\mathbb{R}^4$  中所有正交于向量  $(1, 1, 1, 1)^T$  和  $(1, 2, 3, 4)^T$  的向量。

2.2. 设  $A$  是一个  $m \times n$  实数矩阵。描述  $(\text{Ran } A^T)^\perp$  和  $(\text{Ran } A)^\perp$ 。

2.3. 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组标准正交基。

a) 证明对于任意  $\mathbf{x} = \sum_{k=1}^n \alpha_k \mathbf{v}_k, \mathbf{y} = \sum_{k=1}^n \beta_k \mathbf{v}_k$ , 有

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n \alpha_k \overline{\beta_k}.$$

b) 从 a) 推导出**帕塞瓦尔恒等式**(Parseval's identity):

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n (\mathbf{x}, \mathbf{v}_k) \overline{(\mathbf{y}, \mathbf{v}_k)}.$$

c) 现在假设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  仅仅是一组正交基, 而不是标准正交基。你能写出在这种情况下帕塞瓦尔恒等式吗?

<sup>40</sup>这是一个非常重要的注记, 它允许我们将关于标准内积空间  $\mathbb{F}^n$  的任何陈述转换到具有标准正交基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  的内积空间。

这个问题表明, 如果我们有一组标准正交基, 我们可以像使用  $\mathbb{C}^n$  (或  $\mathbb{R}^n$ ) 中的标准坐标一样使用该基下的坐标。

下面的问题表明, 我们可以通过声明一组基为标准正交基来定义一个内积。

- 2.4. 设  $V$  是一个向量空间,  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $V$  中的一组基。对于  $\mathbf{x} = \sum_{k=1}^n \alpha_k \mathbf{v}_k$ ,  $\mathbf{y} = \sum_{k=1}^n \beta_k \mathbf{v}_k$ , 定义  $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{k=1}^n \alpha_k \beta_k$ .

证明  $\langle \mathbf{x}, \mathbf{y} \rangle$  定义了  $V$  上的一个内积。

- 2.5. 设  $A$  是一个  $m \times n$  实矩阵。请描述  $\mathbb{F}^m$  中所有正交于  $\text{Ran } A$  的向量的集合。

### 3. 正交投影和格拉姆-施密特正交化

回想一下二维平面几何中正交投影的定义, 人们可以引入以下定义。设  $E$  是内积空间  $V$  的一个子空间。

**定义 3.1** 对于向量  $\mathbf{v}$ , 它到子空间  $E$  的**正交投影** (orthogonal projection)  $P_E \mathbf{v}$  是一个向量  $\mathbf{w}$ , 满足:

1.  $\mathbf{w} \in E$ ;
2.  $\mathbf{v} - \mathbf{w} \perp E$ .

我们将使用记号  $\mathbf{w} = P_E \mathbf{v}$  表示正交投影。

在引入一个对象之后, 自然要问:

1. 该对象是否存在?
2. 该对象是否唯一?
3. 如何找到它?

我们将首先证明投影是唯一的。然后我们将给出一个找到投影的方法, 证明它的存在性。

下面的定理表明了为什么正交投影很重要, 同时这也证明了它的唯一性。

**定理 3.2.** 正交投影  $\mathbf{w} = P_E \mathbf{v}$  使  $\mathbf{v}$  到  $E$  的距离最小化, 即对于所有  $\mathbf{x} \in E$ ,

$$\|\mathbf{v} - \mathbf{w}\| \leq \|\mathbf{v} - \mathbf{x}\|.$$

而且, 如果对于某个  $\mathbf{x} \in E$ , 有

$$\|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - \mathbf{x}\|,$$

则  $\mathbf{x} = \mathbf{w}$ .

**证明:** 设  $\mathbf{y} = \mathbf{w} - \mathbf{x}$ . 那么

$$\mathbf{v} - \mathbf{x} = \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{x} = \mathbf{v} - \mathbf{w} + \mathbf{y}.$$

由于  $\mathbf{v} - \mathbf{w} \perp E$ , 所以  $\mathbf{y} \perp \mathbf{v} - \mathbf{w}$ , 因此根据勾股定理

$$\|\mathbf{v} - \mathbf{x}\|^2 = \|\mathbf{v} - \mathbf{w}\|^2 + \|\mathbf{y}\|^2 \geq \|\mathbf{v} - \mathbf{w}\|^2.$$

注意, 当且仅当  $\mathbf{y} = \mathbf{0}$ , 即  $\mathbf{x} = \mathbf{w}$  时, 等号成立。  $\square$

下面的命题表明, 如果我们知道  $E$  中的一组正交基, 我们就能找到  $E$  上的正交投影。

**命题 3.3** 设  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  是  $E$  的一组正交基。那么向量  $\mathbf{v}$  到  $E$  的正交投影  $P_E \mathbf{v}$  由以下公式给出:

$$P_E \mathbf{v} = \sum_{k=1}^r \alpha_k \mathbf{v}_k, \quad \text{其中} \quad \alpha_k = \frac{(\mathbf{v}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}.$$

换句话说,

$$(3.1) \quad P_E \mathbf{v} = \sum_{k=1}^r \frac{(\mathbf{v}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \mathbf{v}_k.$$

注意,  $\alpha_k$  的公式与 (2.1) 一致, 即如果将这个公式应用于一个正交系统 (而不是基) 就可以得到其张成的子空间上的投影。

**注记 3.4.** 由公式 (3.1) 可知, 正交投影  $P_E$  是一个线性变换。

也可以直接从正交投影的定义和唯一性看出其线性。事实上, 容易验证, 对于任意向量  $\mathbf{x}$  和  $\mathbf{y}$  以及常数  $\alpha$  和  $\beta$ , 向量  $\alpha \mathbf{x} + \beta \mathbf{y} - (\alpha P_E \mathbf{x} + \beta P_E \mathbf{y})$  与  $E$  中的任何向量都正交, 因此根据定义,  $P_E(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha P_E \mathbf{x} + \beta P_E \mathbf{y}$ 。

**注记 3.5.** 回忆在  $\mathbb{C}^n$  和  $\mathbb{R}^n$  中内积的定义, 可以从上述公式 (3.1) 得到将  $\mathbb{C}^n$  (或  $\mathbb{R}^n$ ) 中的向量投影到  $E$  上的正交投影矩阵  $P_E$ , 它由下式给出:

$$(3.2) \quad P_E = \sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \mathbf{v}_k^*$$

其中列向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  构成  $E$  中的一组正交基。

**命题 3.3 的证明** 设

$$\mathbf{w} := \sum_{k=1}^r \alpha_k \mathbf{v}_k, \quad \text{其中} \quad \alpha_k = \frac{(\mathbf{v}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}.$$

我们想证明  $\mathbf{v} - \mathbf{w} \perp E$ . 根据引理 2.3, 只要证明  $\mathbf{v} - \mathbf{w} \perp \mathbf{v}_k$ ,  $\forall k = 1, 2, \dots, r$  即可。计算内积, 我们得到对于  $k = 1, 2, \dots, r$ ,

$$\begin{aligned} (\mathbf{v} - \mathbf{w}, \mathbf{v}_k) &= (\mathbf{v}, \mathbf{v}_k) - (\mathbf{w}, \mathbf{v}_k) = (\mathbf{v}, \mathbf{v}_k) - \left( \sum_{j=1}^r \alpha_j \mathbf{v}_j, \mathbf{v}_k \right) \\ &= (\mathbf{v}, \mathbf{v}_k) - \alpha_k (\mathbf{v}_k, \mathbf{v}_k) = (\mathbf{v}, \mathbf{v}_k) - \frac{(\mathbf{v}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \|\mathbf{v}_k\|^2 = 0. \end{aligned}$$

$\square$

因此, 如果我们知道  $E$  中的一组正交基, 我们就可以找到到  $E$  的正交投影。特别地, 由于任何只包含一个向量的系统都是正交系统, 我们也知道了如何进行到一维空间上的正交投影。

但是如果我们只知道  $E$  的一组基, 我们该如何找到正交投影呢? 幸运的是, 存在一个简单的算法, 可以把一组基化为正交基。

### 3.1. 格拉姆-施密特正交化方法

假设我们有一个线性无关系统  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ . 格拉姆-施密特方法从这个系统出发构造一个正交系统  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ , 使得

$$\text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}.$$

而且, 对于所有  $r \leq n$ , 都有

$$\text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}.$$

现在我们来描述这个算法。

**步骤 1** 令  $\mathbf{v}_1 := \mathbf{x}_1$ . 记  $E_1 := \text{span}\{\mathbf{x}_1\} = \text{span}\{\mathbf{v}_1\}$ .

**步骤 2** 定义  $\mathbf{v}_2$  为

$$\mathbf{v}_2 := \mathbf{x}_2 - P_{E_1} \mathbf{x}_2 = \mathbf{x}_2 - \frac{(\mathbf{x}_2, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2} \mathbf{v}_1.$$

定义  $E_2 := \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ . 注意  $\text{span}\{\mathbf{x}_1, \mathbf{x}_2\} = E_2$ .

**步骤 3** 定义  $\mathbf{v}_3$  为

$$\mathbf{v}_3 := \mathbf{x}_3 - P_{E_2} \mathbf{x}_3 = \mathbf{x}_3 - \frac{(\mathbf{x}_3, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \frac{(\mathbf{x}_3, \mathbf{v}_2)}{\|\mathbf{v}_2\|^2} \mathbf{v}_2.$$

令  $E_3 := \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ . 注意  $\text{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = E_3$ . 也注意  $\mathbf{x}_3 \notin E_2$  所以  $\mathbf{v}_3 \neq 0$ .

...

**步骤  $r+1$**  假设我们已经完成了过程的  $r$  步, 构造了一个正交系统 (包含非零向量)  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ , 使得  $E_r := \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r\}$ . 定义

$$\mathbf{v}_{r+1} := \mathbf{x}_{r+1} - P_{E_r} \mathbf{x}_{r+1} = \mathbf{x}_{r+1} - \sum_{k=1}^r \frac{(\mathbf{x}_{r+1}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \mathbf{v}_k.$$

注意  $\mathbf{x}_{r+1} \notin E_r$  所以  $\mathbf{v}_{r+1} \neq 0$ .

...

继续进行这个算法有限次, 我们将得到一个正交系统  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ .



### 3.2. 一个例子

假设我们有向量

$$\mathbf{x}_1 = (1, 1, 1)^T, \quad \mathbf{x}_2 = (0, 1, 2)^T, \quad \mathbf{x}_3 = (1, 0, 2)^T,$$

我们想通过格拉姆-施密特方法来正交化它们。第一步定义

$$\mathbf{v}_1 = \mathbf{x}_1 = (1, 1, 1)^T.$$

第二步我们得到

$$\mathbf{v}_2 = \mathbf{x}_2 - P_{E_1} \mathbf{x}_2 = \mathbf{x}_2 - \frac{(\mathbf{x}_2, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2} \mathbf{v}_1.$$

计算

$$(\mathbf{x}_2, \mathbf{v}_1) = \left( \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = 3, \quad \|\mathbf{v}_1\|^2 = 3,$$

我们得到

$$\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{3}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

最后, 定义

$$\mathbf{v}_3 = \mathbf{x}_3 - P_{E_2} \mathbf{x}_3 = \mathbf{x}_3 - \frac{(\mathbf{x}_3, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \frac{(\mathbf{x}_3, \mathbf{v}_2)}{\|\mathbf{v}_2\|^2} \mathbf{v}_2.$$

计算

$$\left( \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = 3, \quad \left( \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right) = 1, \quad \|\mathbf{v}_1\|^2 = 3, \quad \|\mathbf{v}_2\|^2 = 2.$$

( $\|\mathbf{v}_1\|^2$  已经计算过了) 我们得到

$$\mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} - \frac{3}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix}.$$

**注记** 由于对某个向量乘以一个标量不改变向量组正交性, 因此可以将由格拉姆-施密特正交化得到的向量  $\mathbf{v}_k$  乘以任意非零数。

特别地, 在许多理论构造中, 人们通过将向量  $\mathbf{v}_k$  除以它们各自的范数  $\|\mathbf{v}_k\|$  来归一化它们。然后得到的结果系统将是标准正交的, 并且公式会更简单。

另一方面, 在进行计算时, 人们可能希望避免出现分数项。办法是将向量乘以各项分母数字的最小公倍数。因此, 可以将上面例子中的向量  $\mathbf{v}_3$  替换为  $(1, -2, 1)^T$ 。

### 3.3. 正交补 正交分解 $V = E \oplus E^\perp$

**定义** 对于子空间  $E$ , 其**正交补**(orthogonal complement)  $E^\perp$  是所有与  $E$  正交的向量的集合,

$$E^\perp := \{\mathbf{x} : \mathbf{x} \perp E\}.$$

如果  $\mathbf{x}, \mathbf{y} \perp E$ , 则对于任意线性组合都有  $\alpha\mathbf{x} + \beta\mathbf{y} \perp E$  (你能看出为什么吗?). 因此  $E^\perp$  是一个子空间。

根据正交投影的定义, 任何内积空间  $V$  中的向量都可以唯一地表示为

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2, \quad \mathbf{v}_1 \in E, \quad \mathbf{v}_2 \perp E \text{ (等价地, } \mathbf{v}_2 \in E^\perp \text{)}$$

(其中显然  $\mathbf{v}_1 = P_E \mathbf{v}$ )。

这个陈述通常被象征性地写成  $V = E \oplus E^\perp$ , 这意味着任何向量都可以进行上述**正交分解**, 并且分解是唯一的。

以下命题给出了正交补的一个重要性质。

**命题 3.6** 对于子空间  $E$ ,

$$(E^\perp)^\perp = E.$$

证明留给读者作为练习, 见下面的练习 3.12。

#### 练习

- 3.1. 将向量  $(1, 2, -2)^T$ ,  $(1, -1, 4)^T$ ,  $(2, 1, 1)^T$  格拉姆-施密特正交化。
- 3.2. 将向量  $(1, 2, 3)^T$ ,  $(1, 3, 1)^T$  格拉姆-施密特正交化。写出由这两个向量张成的二维子空间的**正交投影**矩阵。
- 3.3. 将上一个问题中得到的正交系统补全为  $\mathbb{R}^3$  中的一组正交基, 即向系统中添加一些向量 (有多少个?) 以得到一组正交基。  
你能描述如何将一个正交系统补全为一般情况 ( $\mathbb{R}^n$  或  $\mathbb{C}^n$ ) 中的一组正交基吗?
- 3.4. 求向量  $(2, 3, 1)^T$  到由向量  $(1, 2, 3)^T$ ,  $(1, 3, 1)^T$  张成的子空间的距离。注意, 只计算到子空间的距离, 而不是正交投影。
- 3.5. 找到向量  $(1, 1, 1, 1)^T$  到由向量  $\mathbf{v}_1 = (1, 3, 1, 1)^T$  和  $\mathbf{v}_2 = (2, -1, 1, 0)^T$  张成的子空间的**正交投影** (注意  $\mathbf{v}_1 \perp \mathbf{v}_2$ )。
- 3.6. 求向量  $(1, 2, 3, 4)^T$  到由向量  $\mathbf{v}_1 = (1, -1, 1, 0)^T$  和  $\mathbf{v}_2 = (1, 2, 1, 1)^T$  张成的子空间的距离 (注意  $\mathbf{v}_1 \perp \mathbf{v}_2$ )。能否在不实际计算投影的情况下找到距离? 这将简化计算。
- 3.7. 判断正误: 如果  $E$  是  $V$  的子空间, 则有  $\dim E + \dim(E^\perp) = \dim V$ . 证明你的结论。
- 3.8. 设  $P$  是到内积空间  $V$  的子空间  $E$  的正交投影,  $\dim V = n$ ,  $\dim E = r$ . 找出它的特征值和特征向量 (特征子空间)。找出每个特征值的代数重数和几何重数。
- 3.9. (使用特征值计算行列式)。
  - a) 求到由向量  $(1, 1, \dots, 1)^T$  张成的一维子空间的**正交投影**矩阵;
  - b) 设  $A$  是一个所有项均为 1 的  $n \times n$  矩阵。计算它的特征值和重数 (使用上一个问题);
  - c) 计算矩阵  $A - I$  (即主对角线全为零, 其他所有元素都为 1 的矩阵) 的特征值 (和重数);
  - d) 计算  $\det(A - I)$ .

- 3.10. (勒让德多项式 (Legendre's polynomials)): 设内积在多项式空间上由  $(f, g) = \int_{-1}^1 f(t)g(t) dt$  定义。将格拉姆-施密特正交化应用于系统  $\{1, t, t^2, t^3\}$ 。  
勒让德多项式是所谓的正交多项式的特例，它们在数学的许多分支中起着重要作用。
- 3.11. 设  $P$  是到子空间  $E$  的正交投影。证明：  
a) 矩阵  $P$  是**自伴随**的，即  $P^* = P$ 。  
b)  $P^2 = P$ 。  
**注**：以上 2 个性质完全刻画了正交投影，即满足这些性质的任何矩阵都是某个正交投影的矩阵。我们稍后将讨论这一点。
- 3.12. 证明对于子空间  $E$ ，有  $(E^\perp)^\perp = E$ 。**提示**：很容易看出  $E$  正交于  $E^\perp$  (为什么?)。为了证明任何正交于  $E^\perp$  的向量  $\mathbf{x}$  属于  $E$ ，使用上面第 3.3 节中的分解  $V = E \oplus E^\perp$ 。
- 3.13. 假设  $P$  是到子空间  $E$  的正交投影，而  $Q$  是到其正交补  $E^\perp$  的正交投影。  
a)  $P + Q$  和  $PQ$  是什么？  
b) 证明  $P - Q$  是它自己的逆。

## 4. 最小二乘解 正交投影的公式

正如第二章第 2 节所讨论的，方程

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

有解当且仅当  $\mathbf{b} \in \text{Ran } A$ 。但对于无解的方程该怎么办？

这似乎是一个愚蠢的问题，因为如果没有解，那么就没有解。但是，在一些实际情况中，常会出现我们要解一个没有解的方程。例如，从实验中得到一个线性方程，如果没有任何误差，那么右侧  $\mathbf{b}$  属于列空间  $\text{Ran } A$ ，方程是相容的。但是在现实生活中，无法避免测量误差，所以一个理论上应该相容的方程可能没有解。那么，在这种情况下我们能做什么？

### 4.1. 最小二乘解

最简单的想法是写出误差

$$\|A\mathbf{x} - \mathbf{b}\|$$

并尝试找到能最小化它的  $\mathbf{x}$ 。如果我们能找到一个  $\mathbf{x}$  使得误差为 0，那么系统就是相容的，我们就得到了精确解。否则，我们就得到所谓的**最小二乘解**(least square solution)。

**最小二乘**这个术语源于最小化  $\|A\mathbf{x} - \mathbf{b}\|$ ，这等价于最小化

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \sum_{k=1}^m |(A\mathbf{x})_k - \mathbf{b}_k|^2 = \sum_{k=1}^m \left| \sum_{j=1}^n A_{k,j}x_j - \mathbf{b}_k \right|^2,$$

即最小化线性函数平方和。

有几种方法可以找到最小二乘解。如果我们处于  $\mathbb{R}^n$  中，并且所有项都是实数，我们可以忽略绝对值。然后我们可以对每个变量  $x_j$  取偏导数，并找到所有偏导数

都为 0 的地方，这将给我们最小值。

#### 4.1.1. 几何方法

然而，有一个更简单的寻找最小值的方法。即，如果我们考虑所有可能的向量  $\mathbf{x}$ ，那么  $A\mathbf{x}$  会给出  $\text{Ran } A$  中的所有可能向量，所以最小化  $\|A\mathbf{x} - \mathbf{b}\|$  就是找到从  $\mathbf{b}$  到  $\text{Ran } A$  的距离。因此， $\|A\mathbf{x} - \mathbf{b}\|$  的值最小当且仅当  $A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}$ ，其中  $P_{\text{Ran } A}$  表示到列空间  $\text{Ran } A$  的正交投影。

所以，为了找到最小二乘解，我们只需要解方程

$$A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}.$$

如果我们知道  $\text{Ran } A$  中的一组正交基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ ，我们可以通过公式

$$P_{\text{Ran } A}\mathbf{b} = \sum_{k=1}^n \frac{(\mathbf{b}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \mathbf{v}_k$$

来找到向量  $P_{\text{Ran } A}\mathbf{b}$ 。

如果我们只知道  $\text{Ran } A$  中的一组基，我们需要使用格拉姆-施密特正交化来从它得到一组正交基。

因此，理论上，问题已经解决了，但解决方案并不那么简单：它涉及格拉姆-施密特正交化，这在计算上可能很繁杂。幸运的是，存在一个更简单的解决方案。

#### 4.1.2. 正规方程

即， $A\mathbf{x}$  是正交投影  $P_{\text{Ran } A}\mathbf{b}$  当且仅当  $\mathbf{b} - A\mathbf{x} \perp \text{Ran } A$  (对所有  $\mathbf{x}$ ,  $A\mathbf{x} \in \text{Ran } A$ )。

如果  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  是  $A$  的列，那么条件  $A\mathbf{x} \perp \text{Ran } A$  可以重写为

$$\mathbf{b} - A\mathbf{x} \perp \mathbf{a}_k, \quad \forall k = 1, 2, \dots, n.$$

这意味着

$$0 = (\mathbf{b} - A\mathbf{x}, \mathbf{a}_k) = \mathbf{a}_k^*(\mathbf{b} - A\mathbf{x}), \quad \forall k = 1, 2, \dots, n.$$

将行  $\mathbf{a}_k^*$  拼接起来，我们得到这些方程等价于

$$A^*(\mathbf{b} - A\mathbf{x}) = \mathbf{0},$$

这反过来等价于所谓的**正规方程** (normal equation)

$$A^*A\mathbf{x} = A^*\mathbf{b}.$$

该方程的解给出了  $A\mathbf{x} = \mathbf{b}$  的最小二乘解。

注意，当且仅当  $A^*A$  可逆时，最小二乘解是唯一的。

## 4.2. 正交投影公式

如上所述, 如果  $\mathbf{x}$  是正规方程  $A^*A\mathbf{x} = A^*\mathbf{b}$  的解 (即  $A\mathbf{x} = \mathbf{b}$  的最小二乘解), 那么  $A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}$ . 所以, 为了找到  $\mathbf{b}$  到列空间  $\text{Ran } A$  的正交投影, 我们需要解正规方程  $A^*A\mathbf{x} = A^*\mathbf{b}$ , 然后将解乘以  $A$ .

如果算子  $A^*A$  可逆, 则正规方程  $A^*A\mathbf{x} = A^*\mathbf{b}$  的解由  $\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}$  给出, 因此正交投影  $P_{\text{Ran } A}\mathbf{b}$  可以计算为

$$P_{\text{Ran } A}\mathbf{b} = A(A^*A)^{-1}A^*\mathbf{b}.$$

由于这对所有  $\mathbf{b}$  都成立,

$$P_{\text{Ran } A} = A(A^*A)^{-1}A^*$$

就是到  $\text{Ran } A$  的正交投影矩阵的公式。

下面的定理意味着, 对于一个  $m \times n$  矩阵  $A$ , 矩阵  $A^*A$  是可逆的当且仅当  $\text{rank } A = n$ .

**定理 4.1.** 对于一个  $m \times n$  矩阵  $A$ ,

$$\text{Ker } A = \text{Ker}(A^*A).$$

确实, 根据秩定理, 当且仅当  $\text{rank } A = n$  时,  $\text{Ker } A = \{\mathbf{0}\}$ . 所以, 当且仅当  $\text{rank } A = n$  时,  $\text{Ker}(A^*A) = \{\mathbf{0}\}$ . 因为  $A^*A$  矩阵是方阵, 故而, 当且仅当  $\text{rank } A = n$  时, 矩阵  $A^*A$  是可逆的。

我们把定理的证明留给读者。要证明  $\text{Ker } A = \text{Ker}(A^*A)$ , 需要证明两个包含关系  $\text{Ker}(A^*A) \subseteq \text{Ker } A$  和  $\text{Ker } A \subseteq \text{Ker}(A^*A)$ . 其中一个包含关系是平凡的, 对于另一个, 使用事实

$$\|A\mathbf{x}\|^2 = (A\mathbf{x}, A\mathbf{x}) = (A^*A\mathbf{x}, \mathbf{x}).$$

## 4.3. 一个例子: 直线拟合

让我们引入几个最小二乘解自然出现的例子。假设我们知道两个量  $x$  和  $y$  之间的关系由线性规律  $y = a + bx$  给出。系数  $a$  和  $b$  是未知的, 我们希望通过实验数据找到它们。

假设我们进行了  $n$  次实验, 得到了  $n$  对  $(x_k, y_k)$ ,  $k = 1, 2, \dots, n$ . 理想情况下, 所有点  $(x_k, y_k)$  都应该在一条直线上, 但由于测量误差, 通常不会这样: 点通常接近某条直线, 但并不完全在其上面。这时最小二乘解就能用上了!

理想情况下, 系数  $a$  和  $b$  应该满足方程

$$a + bx_k = y_k, \quad k = 1, 2, \dots, n$$

(注意这里,  $x_k$  和  $y_k$  是一些固定的数字, 而未知数是  $a$  和  $b$ )。如果可以直接求出这样的  $a$  和  $b$ , 那就是非常幸运的。如果不行, 标准做法是最小化总的二次误差

$$\sum_{k=1}^n |a + bx_k - y_k|^2.$$

但是, 最小化这个误差恰好是求解系统

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

的最小二乘解 (回忆  $x_k$  和  $y_k$  是一些给定的数字, 未知数是  $a$  和  $b$ )。

#### 4.3.1. 一个例子

假设我们的数据  $(x_k, y_k)$  由数对

$$(-2, 4), \quad (-1, 2), \quad (0, 1), \quad (2, 1), \quad (3, 1)$$

组成。那么我们需要求解方程

$$\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

的最小二乘解。那么

$$A^*A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 2 & 18 \end{pmatrix},$$

并且

$$A^*\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \end{pmatrix}.$$

所以正规方程  $A^*Ax = A^*b$  被重写为

$$\begin{pmatrix} 5 & 2 \\ 2 & 18 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \end{pmatrix}.$$

该方程的解是

$$a = 2, \quad b = -1/2,$$

因此最佳拟合直线是

$$y = 2 - \frac{1}{2}x.$$

#### 4.4. 其他例子：曲线和平面拟合

最小二乘法不限于直线拟合。它也可以应用于更一般的曲线，以及更高维度中的曲面。这里唯一的限制是我们要寻找的参数必须是线性的。一般算法如下：

1. 找到理想情况下数据精确拟合时应该满足的方程；
2. 将这些方程写成一个线性系统，其中未知数是我们想要寻找的参数。注意，系统不一定是一致的（通常不是）；
3. 找到该系统的最小二乘解。

##### 4.4.1. 曲线拟合例子

例如，假设我们知道  $x$  和  $y$  之间的关系由二次函数  $y = a + bx + cx^2$  给出，所以我们想为数据拟合出一个抛物线  $y = a + bx + cx^2$ ，那么我们的未知数  $a, b, c$  应该满足方程

$$a + bx_k + cx_k^2 = y_k, \quad k = 1, 2, \dots, n$$

或者，以矩阵形式

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

例如，对于上例中的数据，我们需要求解方程

$$\begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

的最小二乘解, 那么

$$A^*A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 & 3 \\ 4 & 1 & 0 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} = \begin{pmatrix} 5 & 2 & 18 \\ 2 & 18 & 26 \\ 18 & 26 & 114 \end{pmatrix},$$

并且

$$A^*\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 & 3 \\ 4 & 1 & 0 & 4 & 9 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \\ 31 \end{pmatrix}.$$

因此, 正规方程  $A^*A\mathbf{x} = A^*\mathbf{b}$  是

$$\begin{pmatrix} 5 & 2 & 18 \\ 2 & 18 & 26 \\ 18 & 26 & 114 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \\ 31 \end{pmatrix}$$

它有且仅有一解

$$a = 86/77, \quad b = -62/77, \quad c = 43/154.$$

因此,

$$y = \frac{86}{77} - \frac{62}{77}x + \frac{43}{154}x^2$$

是最佳拟合抛物线。

#### 4.4.2. 平面拟合

再举一个例子, 我们拟合一个平面  $z = a + bx + cy$  到数据

$$(x_k, y_k, z_k) \in \mathbb{R}^3, \quad k = 1, 2, \dots, n.$$

在精确拟合的情况下, 我们应该有的方程是

$$a + bx_k + cy_k = z_k, \quad k = 1, 2, \dots, n,$$

或者, 矩阵形式

$$\begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$



所以，为了找到最佳拟合平面，我们需要找到这个系统（未知数是  $a, b, c$ ）的最小二乘解。

### 练习

#### 4.1. 求解方程组

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

的最小二乘解。

#### 4.2. 找到矩阵

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

的列空间的正交投影矩阵  $P$ 。

使用两种方法：格拉姆-施密特正交化和投影公式。

比较所得结果。

#### 4.3. 找到点 $(-2, 4), (-1, 3), (0, 1), (2, 0)$ 的最佳直线拟合（最小二乘解）。

#### 4.4. 将平面 $z = a + bx + cy$ 拟合到四个点 $(1, 1, 3), (0, 3, 6), (2, 1, 5), (0, 0, 0)$ 。

为此：

a) 找出 4 个关于 3 个未知数  $a, b, c$  的方程，使得平面通过所有 4 个点（这个系统不一定有解）；

b) 找到该系统的最小二乘解。

#### 4.5. 最小范数解 设方程 $A\mathbf{x} = \mathbf{b}$ 有解，并且设 $A$ 有非平凡的核（因此解不唯一）。证明：

a) 存在唯一一个  $A\mathbf{x} = \mathbf{b}$  的解  $\mathbf{x}_0$ ，它最小化范数  $\|\mathbf{x}\|$ ，即存在唯一的  $\mathbf{x}_0$  使得  $A\mathbf{x}_0 = \mathbf{b}$  且  $\|\mathbf{x}_0\| \leq \|\mathbf{x}\|$  对于任何满足  $A\mathbf{x} = \mathbf{b}$  的  $\mathbf{x}$ 。

b)  $\mathbf{x}_0 = P_{(\text{Ker } A)^\perp} \mathbf{x}$  对于任何满足  $A\mathbf{x} = \mathbf{b}$  的  $\mathbf{x}$ 。

#### 4.6. 最小范数的最小二乘解 将上一问题应用于方程 $A\mathbf{x} = P_{\text{Ran } A} \mathbf{b}$ ，证明 $A\mathbf{x} = \mathbf{b}$ 的一个最小范数的最小二乘解 $\mathbf{x}_0$ 存在且唯一。

a) 存在唯一的最小二乘解  $\mathbf{x}_0$  最小化范数  $\|\mathbf{x}\|$ 。

b)  $\mathbf{x}_0 = P_{(\text{Ker } A)^\perp} \mathbf{x}$  对于任何  $A\mathbf{x} = \mathbf{b}$  的最小二乘解  $\mathbf{x}$ 。

## 5. 线性变换的伴随，基本子空间的再次回顾

### 5.1. 伴随矩阵与伴随算子

让我们回忆一下，对于一个  $m \times n$  矩阵  $A$ ，其共轭转置（或简单地说是伴随） $A^*$  定义为  $A^* := \overline{A^T}$ 。换句话说，矩阵  $A^*$  是通过转置矩阵  $A^T$  然后取每个元素的复共轭得到的。

以下恒等式是伴随矩阵的主要性质：

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y}) \quad \forall \mathbf{x} \in \mathbb{C}^n, \forall \mathbf{y} \in \mathbb{C}^m.$$

在证明这个恒等式之前，让我们引入一些有用的公式。回忆一下，对于转置矩阵我们有恒等式  $(AB)^T = B^T A^T$ 。由于对于复数  $z$  和  $w$  我们有  $\overline{zw} = \overline{z}\overline{w}$ ，所以对于伴随有恒等式

$$(AB)^* = B^* A^*.$$

同样, 由于  $(A^T)^T = A$  且  $\bar{\bar{z}} = z$ ,

$$(A^*)^* = A.$$

现在, 我们准备证明正主恒等式:

$$(A\mathbf{x}, \mathbf{y}) = \mathbf{y}^* A\mathbf{x} = (A^*\mathbf{y})^* \mathbf{x} = (\mathbf{x}, A^*\mathbf{y});$$

这里第一个和最后一个等式遵循内积在  $\mathbb{F}^n$  中的定义, 而中间的等式遵循事实

$$(A^*\mathbf{x})^* = \mathbf{x}^* (A^*)^* = \mathbf{x}^* A.$$

### 5.1.1. 伴随的唯一性

上述恒等式  $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y})$  在其他地方通常用作对伴随算子的定义。让我们首先注意到伴随算子是唯一的: 如果一个矩阵  $B$  满足

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, B\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y},$$

则  $B = A^*$ . 确实, 根据  $A^*$  的定义, 对于给定的  $\mathbf{y}$ , 我们有

$$(\mathbf{x}, A^*\mathbf{y}) = (\mathbf{x}, B\mathbf{y}) \quad \forall \mathbf{x},$$

因此根据推论 1.5,  $A^*\mathbf{y} = B\mathbf{y}$ . 由于这对所有  $\mathbf{y}$  都成立, 即这是线性变换, 因此矩阵  $A^*$  和  $B$  是相等的。

### 5.1.2. 抽象环境下的伴随变换

前面提到过, 恒等式  $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y})$  可用于在抽象环境中定义伴随算子, 其中  $A: V \rightarrow W$  是作用在一个内积空间到另一个内积空间上的算子。即, 我们定义  $A^*: W \rightarrow V$  为满足

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y}) \quad \forall \mathbf{x} \in V, \forall \mathbf{y} \in W$$

的算子。为什么这样的算子存在? 我们可以简单地构造它: 考虑  $V$  中的一组标准正交基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  和  $W$  中的一组标准正交基  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$ . 如果  $[A]_{BA}$  是这两组基下  $A$  的矩阵, 我们以定义其矩阵  $[A^*]_{AB}$  来定义算子  $A^*$ :

$$[A^*]_{AB} = ([A]_{BA})^*.$$

我们将该算子是伴随算子的证明留给读者作为练习。

注意, 上述第 5.1.1 节中的推理意味着伴随算子是唯一的。

### 5.1.3. 有用的公式

下面我们给出将广泛使用的伴随算子 (矩阵) 的性质。我们将证明留给读者作为练习。

1.  $(A + B)^* = A^* + B^*$ ;
2.  $(\alpha A)^* = \overline{\alpha} A^*$ ;
3.  $(AB)^* = B^* A^*$ ;
4.  $(A^*)^* = A$ ;
5.  $(y, Ax) = (A^* y, x)$ .

## 5.2. 基本子空间之间的关系

**定理 5.1.** 设  $A : V \rightarrow W$  是从一个内积空间作用到另一个内积空间上的算子。那么

1.  $\text{Ker } A^* = (\text{Ran } A)^\perp$ ;
2.  $\text{Ker } A = (\text{Ran } A^*)^\perp$ ;
3.  $\text{Ran } A = (\text{Ker } A^*)^\perp$ ;
4.  $\text{Ran } A^* = (\text{Ker } A)^\perp$ .

**注记** 在第二章第 7 节, 基本子空间的定义 (如文献中常见的那样) 使用  $A^T$  而不是  $A^*$ . 当然, 对于实数矩阵, 两者并无区别, 所以在实数情况下, 本定理给出了那里定义的基本子空间的几何描述。

第八章第 3 节 (定理 3.7) 给出了使用  $A^T$  定义的基本子空间的几何解释。本定理中的公式与定理 5.1 中的公式基本相同, 只是解释略有不同。

**定理 5.1 的证明** 首先, 我们注意到, 对于子空间  $E$ , 我们有  $(E^\perp)^\perp = E$ , 所以陈述 1 和 3, 陈述 2 和 4 都是等价的。最后, 陈述 2 恰好是将算子  $A$  替换为  $A^*$  的陈述 1 (这里我们使用了  $(A^*)^* = A$  的事实)。

因此, 我们只需要证明陈述 1。

我们将为此陈述提供两种证明: “矩阵”证明和 “不变” (或 “坐标无关”) 证明。

在 “矩阵” 证明中, 我们假设  $A$  是一个  $m \times n$  矩阵, 即  $A : \mathbb{F}^n \rightarrow \mathbb{F}^m$ . 一般情况总可以通过选取  $V$  和  $W$  中的标准正交基来简化为这种情况。

设  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  是  $A$  的列。注意,  $\mathbf{x} \in (\text{Ran } A)^\perp$  当且仅当  $\mathbf{x} \perp \mathbf{a}_k$  (即  $(\mathbf{x}, \mathbf{a}_k) = 0$ )  $\forall k = 1, 2, \dots, n$ .

根据  $\mathbb{F}^n$  中内积的定义, 这意味着

$$0 = (\mathbf{x}, \mathbf{a}_k) = \mathbf{a}_k^* \mathbf{x} \quad \forall k = 1, 2, \dots, n.$$

由于  $\mathbf{a}_k^*$  是  $A^*$  的第  $k$  行, 上述  $n$  个等式等价于方程

$$A^*\mathbf{x} = \mathbf{0}.$$

所以, 我们证明了  $\mathbf{x} \in (\text{Ran } A)^\perp$  当且仅当  $A^*\mathbf{x} = \mathbf{0}$ , 而这恰好是定理 1 的陈述。

现在, 让我们给出“坐标无关”的证明。 $\mathbf{x} \in (\text{Ran } A)^\perp$  的含义是  $\mathbf{x}$  正交于所有形式为  $A\mathbf{y}$  的向量, 即

$$(\mathbf{x}, A\mathbf{y}) = 0 \quad \forall \mathbf{y}.$$

由于  $(\mathbf{x}, A\mathbf{y}) = (A^*\mathbf{x}, \mathbf{y})$ , 这个恒等式等价于

$$(A^*\mathbf{x}, \mathbf{y}) = 0 \quad \forall \mathbf{y},$$

并且根据引理 1.4, 这当且仅当  $A^*\mathbf{x} = \mathbf{0}$ . 所以我们证明了  $\mathbf{x} \in (\text{Ran } A)^\perp$  当且仅当  $A^*\mathbf{x} = \mathbf{0}$ , 而这恰好是定理 1 的陈述。□

### 5.3. 线性变换的“本质”部分

上述定理使得算子  $A$  的结构以及基本子空间的几何学更加清晰。从该定理可以得出, 算子  $A$  可以表示为到  $\text{Ran } A^*$  的正交投影与从  $\text{Ran } A^*$  到  $\text{Ran } A$  的同构的组合。

确实, 设  $\tilde{A}: \text{Ran } A^* \rightarrow \text{Ran } A$  是  $A$  对定义域  $\text{Ran } A^*$  和目标空间  $\text{Ran } A$  的限制,

$$\tilde{A}\mathbf{x} = A\mathbf{x}, \quad \forall \mathbf{x} \in \text{Ran } A^*.$$

由于  $\text{Ker } A = (\text{Ran } A^*)^\perp$ , 我们有

$$A\mathbf{x} = AP_{\text{Ran } A^*}\mathbf{x} = \tilde{A}P_{\text{Ran } A^*}\mathbf{x} \quad \forall \mathbf{x} \in X;$$

这里使用了  $\mathbf{x} - P_{\text{Ran } A^*}\mathbf{x} \in (\text{Ran } A^*)^\perp = \text{Ker } A$  的事实。因此我们可以写成

$$(5.1) \quad A = \tilde{A}P_{\text{Ran } A^*} \quad \forall \mathbf{x} \in X,$$

或者等价地说,  $A = \tilde{A}P_{\text{Ran } A^*}$ .

还需注意,  $\tilde{A}: \text{Ran } A^* \rightarrow \text{Ran } A$  是一个可逆变换。首先我们注意到  $\text{Ker } \tilde{A} = \{\mathbf{0}\}$ : 如果  $\mathbf{x} \in \text{Ran } A^*$  且  $\tilde{A}\mathbf{x} = A\mathbf{x} = \mathbf{0}$ , 那么  $\mathbf{x} \in \text{Ker } A = (\text{Ran } A^*)^\perp$ , 所以  $\mathbf{x} \in \text{Ran } A^* \cap (\text{Ran } A^*)^\perp$ , 因此  $\mathbf{x} = \mathbf{0}$ . 然后为了证明  $\tilde{A}$  是满射的 (surjective), 必须确保  $\tilde{A}$  是满射的。但这直接从 (5.1) 得出:

$$\text{Ran } \tilde{A} = \tilde{A}(\text{Ran } A^*) = AP_{\text{Ran } A^*}X = AX = \text{Ran } A.$$

同构  $\tilde{A}$  有时被称为算子  $A$  的“本质部分” (essential part) (非标准术语)。

“本质部分”  $\tilde{A}: \text{Ran } A^* \rightarrow \text{Ran } A$  是一个同构, 这隐含了以下“复数”秩定理:  $\text{rank } A = \text{rank } A^*$ . 但是, 当然, 这个定理也来自一组基本事实: 取复共轭不改变矩阵的秩,  $\text{rank } A = \text{rank } \bar{A}$ .

### 练习

5.1. 证明对于方阵  $A$ ,  $\det(A^*) = \overline{\det(A)}$  成立。

5.2. 找出矩阵

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 2 & 4 & 3 \end{pmatrix}.$$

的所有四组基本子空间的正交投影矩阵。注意, 实际上只需要计算其中两个投影。如果你选择合适的两个, 其他的 2 个可以很容易地从它们得到 (回想一下, 投影到  $E$  和  $E^\perp$  的关系)。

5.3. 设  $A$  是一个  $m \times n$  矩阵。证明  $\text{Ker } A = \text{Ker}(A^*A)$ . 为此你需要证明两个包含关系  $\text{Ker}(A^*A) \subseteq \text{Ker } A$  和  $\text{Ker } A \subseteq \text{Ker}(A^*A)$ . 其中一个包含关系是平凡的, 对于另一个, 使用事实

$$\|Ax\|^2 = (Ax, Ax) = (A^*Ax, x).$$

5.4. 使用等式  $\text{Ker } A = \text{Ker}(A^*A)$  来证明:

a)  $\text{rank } A = \text{rank}(A^*A)$ ;

b) 如果  $Ax = 0$  只有平凡解, 则  $A$  是左可逆的。(你只需要写出一个左逆的公式)。

5.5. 假设矩阵  $A$  的  $A^*A$  是可逆的, 因此到  $\text{Ran } A$  的正交投影由公式  $A(A^*A)^{-1}A^*$  给出。你能写出到其他 3 组基本子空间 ( $\text{Ker } A$ ,  $\text{Ker } A^*$ ,  $\text{Ran } A^*$ ) 的正交投影的公式吗?

5.6. 设矩阵  $P$  是自伴随的 ( $P^* = P$ ) 并且  $P^2 = P$ . 证明  $P$  是一个正交投影的矩阵。提示: 考虑分解  $x = x_1 + x_2$ ,  $x_1 \in \text{Ran } P$ ,  $x_2 \perp \text{Ran } P$ , 并证明  $Px_1 = x_1$ ,  $Px_2 = 0$ . 对于其中一个等式, 你将需要用到自伴随性, 对于另一个等式, 你需要用到  $P^2 = P$  的性质。

## 6. 等距同构和酉算子 酉矩阵和正交矩阵

### 6.1. 基本定义

**定义** 算子  $U: X \rightarrow Y$  被称为**等距同构(isometry)**, 如果它保持范数, 即

$$\|Ux\| = \|x\| \quad \forall x \in X.$$

以下定理表明等距同构保持内积:

**定理 6.1.** 算子  $U: X \rightarrow Y$  是等距同构当且仅当它保持内积, 即当且仅当

$$(x, y) = (Ux, Uy) \quad \forall x, y \in X.$$

**证明：** 证明使用了极化恒等式 (第 5 章引理 1.9)。例如, 如果  $X$  是复数空间,

$$\begin{aligned}(U\mathbf{x}, U\mathbf{y}) &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|U\mathbf{x} + \alpha U\mathbf{y}\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|U(\mathbf{x} + \alpha\mathbf{y})\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|\mathbf{x} + \alpha\mathbf{y}\|^2 = (\mathbf{x}, \mathbf{y}).\end{aligned}$$

类似地, 对于实数空间  $X$ ,

$$\begin{aligned}(U\mathbf{x}, U\mathbf{y}) &= \frac{1}{4} (\|U\mathbf{x} + U\mathbf{y}\|^2 - \|U\mathbf{x} - U\mathbf{y}\|^2) \\ &= \frac{1}{4} (\|U(\mathbf{x} + \mathbf{y})\|^2 - \|U(\mathbf{x} - \mathbf{y})\|^2) \\ &= \frac{1}{4} (\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2) = (\mathbf{x}, \mathbf{y}).\end{aligned}$$

□

**引理 6.2.** 算子  $U : X \rightarrow Y$  是等距同构当且仅当  $U^*U = I$ .

**证明：** 如果  $U^*U = I$ , 那么根据伴随算子的定义,

$$(\mathbf{x}, \mathbf{x}) = (U^*U\mathbf{x}, \mathbf{x}) = (U\mathbf{x}, U\mathbf{x}) \quad \forall \mathbf{x} \in X.$$

因此  $\|\mathbf{x}\| = \|U\mathbf{x}\|$ , 所以  $U$  是等距同构。

另一方面, 如果  $U$  是等距同构, 那么根据伴随算子的定义和定理 6.1, 对于所有  $\mathbf{x} \in X$ ,

$$(U^*U\mathbf{x}, \mathbf{y}) = (U\mathbf{x}, U\mathbf{y}) = (\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{y} \in X,$$

因此根据推论 1.5,  $U^*U\mathbf{x} = \mathbf{x}$ . 由于这对所有  $\mathbf{x} \in X$  都成立, 所以  $U^*U = I$ . □

上面的引理意味着等距同构总是左可逆的 ( $U^*$  是左逆)。

**定义** 等距同构  $U : X \rightarrow Y$  被称为酉算子(unitary operator), 如果它是可逆的。

**命题 6.3** 等距同构  $U : X \rightarrow Y$  是酉算子当且仅当  $\dim X = \dim Y$ .

**证明：** 由于  $U$  是等距同构, 它是左可逆的, 并且由于  $\dim X = \dim Y$ , 它是可逆的 (左可逆的方阵是可逆的)。

另一方面, 如果  $U : X \rightarrow Y$  是可逆的, 那么  $\dim X = \dim Y$  (只有方阵才是可逆的, 同构的空间具有相等的维度)。□

一个方阵  $U$  被称为酉矩阵, 如果  $U^*U = I$ , 即酉矩阵是作用在  $\mathbb{F}^n$  上的酉算子的矩阵。

实数项的酉矩阵被称为**正交矩阵**。一个正交矩阵可以解释为作用在实数空间  $\mathbb{R}^n$  上的酉算子的矩阵。

酉算子的几个性质：

1. 对于酉变换  $U$ ,  $U^{-1} = U^*$ ;
2. 如果  $U$  是酉的, 那么  $U^* = U^{-1}$  也必须是酉的;
3. 如果  $U$  是等距同构, 并且  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是一组标准正交基, 那么  $\{U\mathbf{v}_1, U\mathbf{v}_2, \dots, U\mathbf{v}_n\}$  是一个标准正交系统。而且, 如果  $U$  是酉的,  $\{U\mathbf{v}_1, U\mathbf{v}_2, \dots, U\mathbf{v}_n\}$  是一组标准正交基。
4. 酉算子的乘积也是酉算子。

## 6.2. 例子

首先, 让我们注意到,

一个矩阵  $U$  是等距同构当且仅当它的列构成一个标准正交系统。

通过计算乘积  $U^*U$  可以很容易地检验这一点。可以很容易地检验出旋转矩阵

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

的列彼此正交, 并且每列的范数都为 1。因此, 旋转矩阵是等距同构, 并且由于它是方阵, 所以它是酉的。由于旋转矩阵的所有元素都是实数, 它是一个正交矩阵。

下一个例子更抽象。设  $X$  和  $Y$  是内积空间,  $\dim X = \dim Y = n$ , 并且设  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  和  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  分别是  $X$  和  $Y$  中的标准正交基。定义一个算子  $U: X \rightarrow Y$  为

$$U\mathbf{x}_k = \mathbf{y}_k, \quad k = 1, 2, \dots, n.$$

由于对于向量  $\mathbf{x} = \sum_{k=1}^n c_k \mathbf{x}_k$ ,

$$\|\mathbf{x}\|^2 = \sum_{k=1}^n |c_k|^2$$

且

$$\|U\mathbf{x}\|^2 = \left\| \sum_{k=1}^n c_k \mathbf{y}_k \right\|^2 = \sum_{k=1}^n |c_k|^2,$$

我们可以得出  $\|U\mathbf{x}\| = \|\mathbf{x}\| \quad \forall \mathbf{x} \in X$ , 所以  $U$  是一个酉算子。

## 6.3. 酉算子的性质

**命题 6.4** 设  $U$  是一个酉矩阵。那么

1.  $|\det U| = 1$ . 特别地, 对于正交矩阵  $\det U = \pm 1$ ;

2. 如果  $\lambda$  是  $U$  的一个特征值, 那么  $|\lambda| = 1$ .

**注记** 注意, 对于正交矩阵, 特征值 (不像行列式) 不一定必须是实数。我们老朋友, 旋转矩阵就是一个例子。

**命题 6.4 的证明** 设  $\det U = z$ . 由于  $\det(U^*) = \det(U)$ , 见问题 5.1, 我们有

$$|z|^2 = \bar{z} = \det(U^*U) = \det I = 1,$$

所以  $|\det U| = |z| = 1$ . 陈述 1 证毕。

为了证明陈述 2, 我们注意到如果  $U\mathbf{x} = \lambda\mathbf{x}$ , 那么

$$\|U\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|,$$

所以  $|\lambda| = 1$ . □

## 6.4. 酉等价算子

**定义** 算子 (矩阵)  $A$  和  $B$  被称为**酉等价的** (unitarily equivalent), 如果存在一个酉算子  $U$  使得  $A = UBU^*$ .

由于对于酉  $U$  我们有  $U^{-1} = U^*$ , 任何两个酉等价的矩阵也是相似的。

反之则不然, 我们很容易构造出一对相似但不酉等价的矩阵。

下面的命题提供了一种构造反例的方法。

**命题 6.5** 矩阵  $A$  酉等价于一个对角矩阵当且仅当它具有一组**正交** (或标准正交) 的特征向量基。

**证明:** 设  $A$  与对角矩阵  $D$  酉等价, 即  $A = UDU^*$ . 设  $B\mathbf{x} = \lambda\mathbf{x}$ . 那么  $AU\mathbf{x} = UBU^*U\mathbf{x} = UB\mathbf{x} = U(\lambda\mathbf{x}) = \lambda U\mathbf{x}$ , 即  $U\mathbf{x}$  是  $A$  的特征向量。

所以, 设  $A$  具有由特征向量  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  组成的**正交基**。通过将每个向量  $\mathbf{u}_k$  除以其范数 (如果需要), 我们可以假设系统  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  是一组**标准正交基**。设  $D$  是  $A$  在基  $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  下的矩阵。显然,  $D$  是一个对角矩阵。

设  $U$  为以  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  为列的矩阵。由于列向量构成了标准正交基,  $U$  是酉的。标准坐标变换公式意味着

$$A = [A]_{SS} = [I]_{SB}[A]_{BB}[I]_{BS} = UDU^{-1},$$

并且由于  $U$  是酉的,  $A = UDU^*$ . □

### 练习

6.1. 对以下矩阵进行**正交对角化**, 即对每个矩阵  $A$ , 找出酉矩阵  $U$  和对角矩阵  $D$ , 使得  $A = UDU^*$ :

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}.$$



6.2. 判断正误：一个矩阵酉等价于一个对角矩阵当且仅当它具有一组**正交**的特征向量基。

6.3. 证明极化恒等式

$$(A\mathbf{x}, \mathbf{y}) = \frac{1}{4}[(A(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) - (A(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y})] \quad (\text{实数情况, } A = A^*),$$

以及

$$(A\mathbf{x}, \mathbf{y}) = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha (A(\mathbf{x} + \alpha \mathbf{y}), \mathbf{x} + \alpha \mathbf{y}) \quad (\text{复数情况, } A \text{ 任意}).$$

6.4. 证明酉（正交）矩阵的乘积也是酉（正交）的。

6.5. 设  $U: X \rightarrow X$  是一个有限维内积空间上的线性变换。证明或证伪：

a) 如果  $\|U\mathbf{x}\| = \|\mathbf{x}\| \quad \forall \mathbf{x} \in X$ ，那么  $U$  是酉的。

b) 如果  $\|U\mathbf{e}_k\| = \|\mathbf{e}_k\|, k = 1, 2, \dots, n$  对于某组标准正交基  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ ，那么  $U$  是酉的。

6.6. 设  $A$  和  $B$  是酉等价的  $n \times n$  矩阵。

a) 证明  $\text{trace}(A^*A) = \text{trace}(B^*B)$ 。

b) 用 a) 证明

$$\sum_{j,k=1}^n |A_{j,k}|^2 = \sum_{j,k=1}^n |B_{j,k}|^2.$$

c) 用 b) 证明矩阵

$$\begin{pmatrix} 1 & 2 \\ 2 & i \end{pmatrix} \text{ 和 } \begin{pmatrix} i & 4 \\ 1 & 1 \end{pmatrix}$$

不是酉等价的。

6.7. 以下哪些矩阵对是酉等价的：

a)  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  和  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 。

b)  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  和  $\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$ 。

c)  $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  和  $\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ 。

d)  $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  和  $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -i & 0 \\ 0 & 0 & i \end{pmatrix}$ 。

e)  $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$  和  $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ 。

**提示：**很容易排除不酉等价的矩阵：记住酉等价矩阵是相似的，而相似矩阵的迹、行列式和特征值是相同的。

同样，前面的问题有助于排除非酉等价矩阵。

一个矩阵酉等价于一个对角矩阵当且仅当它具有一组特征向量的正交基。

6.8. 设  $U$  是一个行列式为 1 的  $2 \times 2$  正交矩阵。证明  $U$  是一个旋转矩阵。

6.9. 设  $U$  是一个行列式为 1 的  $3 \times 3$  正交矩阵。证明：

a) 1 是  $U$  的一个特征值。

b) 如果  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  是一组标准正交基，使得  $U\mathbf{v}_1 = \mathbf{v}_1$ （记住 1 是一个特征值），那么在基  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  下  $U$  的矩阵是

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix},$$

其中  $\alpha$  是某个角度。

**提示：**证明，由于  $\mathbf{v}_1$  是  $U$  的特征向量，1 下方的所有元素必须为零，并且由于  $\mathbf{v}_1$  也是  $U^*$  的特征向量（为什么？），1 右侧的所有元素也必须为零。然后证明下方的  $2 \times 2$  矩阵是一个行列式为 1 的正交矩阵，并用上一问题的结论。

## 7. $\mathbb{R}^n$ 中的刚性运动

$\mathbb{R}^n$  中的一个**刚性运动** (rigid motion) 是一个变换  $f: V \rightarrow V$ , 它保持点之间的距离, 即

$$\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in V.$$

注意, 定义中我们没有假设变换  $f$  是线性的。显然, 任何酉变换都是刚性运动。另一个刚性运动的例子是平移 (移位)  $\mathbf{a} \in V$ ,  $f(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ .

下面的定理是主要结论, 它表明任何实内积空间中的刚性运动都是正交变换和平移的组合。

**定理 7.1.** 设  $f$  是实内积空间  $X$  中的一个刚性运动, 设  $T(\mathbf{x}) := f(\mathbf{x}) - f(\mathbf{0})$ . 那么  $T$  是一个正交变换。

为了证明这个定理, 我们需要一个简单的引理。

**引理 7.2.** 设  $T$  如定理 7.1 中所定义。那么对于所有  $\mathbf{x}, \mathbf{y} \in X$ :

1.  $\|T\mathbf{x}\| = \|\mathbf{x}\|$ ;
2.  $\|T(\mathbf{x}) - T(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$ ;
3.  $(T\mathbf{x}, T\mathbf{y}) = (\mathbf{x}, \mathbf{y})$ .

**证明:** 为了证明陈述 1, 注意到

$$\|T(\mathbf{x})\| = \|f(\mathbf{x}) - f(\mathbf{0})\| = \|\mathbf{x} - \mathbf{0}\| = \|\mathbf{x}\|.$$

陈述 2 源于以下恒等式链:

$$\begin{aligned} \|T(\mathbf{x}) - T(\mathbf{y})\| &= \|(f(\mathbf{x}) - f(\mathbf{0})) - (f(\mathbf{y}) - f(\mathbf{0}))\| \\ &= \|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|. \end{aligned}$$

另一种解释是  $T$  是两个刚性运动的组合 (先是  $f$ , 然后是平移  $-f(\mathbf{0})$ ), 并且可以很容易地看出刚性运动的组合也还是刚性运动。由于  $T(\mathbf{0}) = \mathbf{0}$ , 并且  $\|T(\mathbf{x})\| = \|T(\mathbf{x}) - T(\mathbf{0})\|$ , 所以陈述 1 可以视为陈述 2 的一个特例。

为了证明陈述 3, 让我们注意到在实内积空间中

$$\|T(\mathbf{x}) - T(\mathbf{y})\|^2 = \|T(\mathbf{x})\|^2 + \|T(\mathbf{y})\|^2 - 2(T(\mathbf{x}), T(\mathbf{y})),$$

并且

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2(\mathbf{x}, \mathbf{y}).$$

回想一下  $\|T(\mathbf{x}) - T(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$  并且  $\|T(\mathbf{x})\| = \|\mathbf{x}\|$ ,  $\|T(\mathbf{y})\| = \|\mathbf{y}\|$ , 我们立即得到了期望的结论。□

**定理 7.1 的证明** 首先, 注意到对于所有  $\mathbf{x} \in X$ ,

$$\|T\mathbf{x}\| = \|f(\mathbf{x}) - f(\mathbf{0})\| = \|\mathbf{x} - \mathbf{0}\| = \|\mathbf{x}\|,$$

所以  $T$  保持范数,  $\|T\mathbf{x}\| = \|\mathbf{x}\|$ .

我们想说由于  $\|T\mathbf{x}\| = \|\mathbf{x}\|$ , 所以  $T$  是一个等距同构, 但要能够得到这一点, 我们还需证明  $T$  是一个线性变换。

为此, 让我们在  $X$  中固定一组标准正交基  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ , 并设  $\mathbf{b}_k := T(\mathbf{e}_k)$ ,  $k = 1, 2, \dots, n$ . 由于  $T$  保持内积 (引理 7.2 的陈述 3), 我们可以得出  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是一个标准正交系统。事实上, 由于  $\dim X = n$  (因为基  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  包含  $n$  个向量), 我们可以得出  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是一组标准正交基。

设  $\mathbf{x} = \sum_{k=1}^n \alpha_k \mathbf{e}_k$ . 回忆一下, 根据抽象正交傅里叶分解 (2.2), 我们有  $\alpha_k = (\mathbf{x}, \mathbf{e}_k)$ . 将抽象正交傅里叶分解 (2.2) 应用于  $T(\mathbf{x})$  和标准正交基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ , 我们得到

$$T(\mathbf{x}) = \sum_{k=1}^n (T(\mathbf{x}), \mathbf{b}_k) \mathbf{b}_k.$$

由于

$$(T(\mathbf{x}), \mathbf{b}_k) = (T(\mathbf{x}), T(\mathbf{e}_k)) = (\mathbf{x}, \mathbf{e}_k) = \alpha_k,$$

我们得到

$$T\left(\sum_{k=1}^n \alpha_k \mathbf{e}_k\right) = \sum_{k=1}^n \alpha_k \mathbf{b}_k.$$

这意味着  $T$  是一个线性变换, 其在基  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  和  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  下的矩阵是单位矩阵,  $[T]_{B,S} = I$ .

另一种证明  $T$  是线性变换的方法是进行以下直接计算:

$$\begin{aligned} & \|T(\mathbf{x} + \alpha\mathbf{y}) - (T(\mathbf{x}) + \alpha T(\mathbf{y}))\|^2 = \|(T(\mathbf{x} + \alpha\mathbf{y}) - T(\mathbf{x})) - \alpha T(\mathbf{y})\|^2 \\ &= \|T(\mathbf{x} + \alpha\mathbf{y}) - T(\mathbf{x})\|^2 + \alpha^2 \|T(\mathbf{y})\|^2 - 2\alpha (T(\mathbf{x} + \alpha\mathbf{y}) - T(\mathbf{x}), T(\mathbf{y})) \\ &= \|\mathbf{x} + \alpha\mathbf{y} - \mathbf{x}\|^2 + \alpha^2 \|\mathbf{y}\|^2 - 2\alpha (T(\mathbf{x} + \alpha\mathbf{y}), T(\mathbf{y})) + 2\alpha (T(\mathbf{x}), T(\mathbf{y})) \\ &= \|\alpha\mathbf{y}\|^2 + \alpha^2 \|\mathbf{y}\|^2 - 2\alpha (\mathbf{x} + \alpha\mathbf{y}, \mathbf{y}) + 2\alpha (\mathbf{x}, \mathbf{y}) \\ &= \alpha^2 \|\mathbf{y}\|^2 + \alpha^2 \|\mathbf{y}\|^2 - 2\alpha (\mathbf{x}, \mathbf{y}) - 2\alpha^2 (\mathbf{y}, \mathbf{y}) + 2\alpha (\mathbf{x}, \mathbf{y}) = 0. \end{aligned}$$

因此

$$T(\mathbf{x} + \alpha\mathbf{y}) = T(\mathbf{x}) + \alpha T(\mathbf{y}),$$

这意味着  $T$  是线性的 (取  $\mathbf{x} = \mathbf{0}$  或  $\alpha = 1$  即可得到线性变换定义中的两个性质)。

所以,  $T$  是一个满足  $\|T\mathbf{x}\| = \|\mathbf{x}\|$  的线性变换, 即  $T$  是一个等距同构。由于  $T: X \rightarrow X$ ,  $T$  是一个酉变换 (见命题 6.3)。这完成了证明, 因为一个正交变换仅仅是一个实内积空间中的酉变换。  $\square$

### 练习

7.1. 在  $\mathbb{C}^n$  中给出一个刚性运动  $T, T(\mathbf{0}) = \mathbf{0}$ , 但  $T$  不是线性变换。

## 8. 复化与反复化

本节可能比本章的其余部分更抽象一些, 首次阅读时可以跳过。

### 8.1. 反复化

#### 8.1.1. 向量空间的复化

任何复数向量空间都可以解释为一个实向量空间: 我们只需要忘记我们可以乘以复数, 并且只允许乘以实数。

例如, 空间  $\mathbb{C}^n$  **典型地** 被对应为实向量空间  $\mathbb{R}^{2n}$ : 每个复数坐标  $z_k = x_k + iy_k$  给出两个实坐标  $x_k$  和  $y_k$ .

“典型地” 在这里意味着这是一种标准、最自然地对应  $\mathbb{C}^n$  和  $\mathbb{R}^{2n}$  的方式。注意, 虽然上述定义给了我们一种从复数坐标得到实数坐标的典型方法, 但它并没有说明坐标的排序。

事实上, 有两种标准的方法来排序坐标  $x_k, y_k$ . 一种方法是先取所有实部, 然后取所有虚部, 所以排序是  $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ . 另一种标准选择是排序  $x_1, y_1, x_2, y_2, \dots, x_n, y_n$ . 本节的内容不依赖于坐标的排序选择, 所以读者不必担心选择排序。

#### 8.1.2. 内积的复化

结果表明, 如果我们有一个复内积 (在一个复数空间中), 我们可以以典型的方式从中得到一个实内积: 实际上, 你可能已经在不知不觉中做过了。即, 考虑  $\mathbb{C}^n$  的上述例子, 它典型地被对应为  $\mathbb{R}^{2n}$ . 设  $(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$  表示  $\mathbb{C}^n$  中的标准内积,  $(\mathbf{x}, \mathbf{y})_{\mathbb{R}}$  表示  $\mathbb{R}^{2n}$  中的标准内积 (注意  $\mathbb{R}^n$  中的标准内积不依赖于坐标的排序)。那么 (见下面的练习 8.1)

$$(8.1) \quad (\mathbf{x}, \mathbf{y})_{\mathbb{R}} = \operatorname{Re}((\mathbf{x}, \mathbf{y})_{\mathbb{C}})$$

这个公式可以用于典型地从复内积中定义一个实内积, 在一般情况下也是如此。即, 很容易核验出, 如果  $(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$  是一个复内积空间中的内积, 那么  $(\mathbf{x}, \mathbf{y})_{\mathbb{R}}$  (定义为 (8.1)) 是一个实内积 (在其相应的实空间上)。

总结一下, 我们可以说,

要对一个复内积空间进行反复化, 我们只需“忘记”我们可以乘以复数, 即我们只允许乘以实数。被反复化空间中的典型实内积由公式 (8.1) 给出。

**注记** 任何作用在  $\mathbb{C}^n$  上的 (复) 线性变换 (或更广泛地说, 在复向量空间上) 都

会产生一个实线性变换：这仅仅是因为如果  $T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y}$  对  $\alpha, \beta \in \mathbb{C}$  成立，那么它当然对  $\alpha, \beta \in \mathbb{R}$  也成立。

反之则不成立，即作用在  $\mathbb{C}^n$  的反复化  $\mathbb{R}^{2n}$  上的（实数）线性变换并不总是产生  $\mathbb{C}^n$  的（复数）线性变换（在抽象情况下也是一样）。

例如，如果考虑  $n = 1$  的情况，那么乘以一个复数  $z$ （复数空间  $\mathbb{C}^1$  中的线性变换的一般形式）被视为  $\mathbb{R}^2$  中的线性变换时，具有一个非常特殊的结构（你能描述它吗？）。

## 8.2. 复化

我们也可以做相反的事情，即从一个实空间得到一个复空间：实际上，你可能已经做过了，而没有太在意。

即，给定一个实内积空间  $\mathbb{R}^n$ ，我们可以从它得到一个复空间  $\mathbb{C}^n$ ，方法是允许复数坐标（在两种情况下都使用标准内积）。在这种情况下，空间  $\mathbb{R}^n$  将是  $\mathbb{C}^n$  的一个**实子空间**<sup>41</sup>，由具有实数坐标的向量组成。

抽象地说，这个构造可以描述如下：给定一个实向量空间  $X$ ，我们可以将其复化  $X_{\mathbb{C}}$  定义为所有对  $[\mathbf{x}_1, \mathbf{x}_2]$ ， $\mathbf{x}_1, \mathbf{x}_2 \in X$  的集合，其中加法和实数  $\alpha$  的乘法是逐坐标定义的：

$$[\mathbf{x}_1, \mathbf{x}_2] + [\mathbf{y}_1, \mathbf{y}_2] = [\mathbf{x}_1 + \mathbf{y}_1, \mathbf{x}_2 + \mathbf{y}_2], \quad \alpha[\mathbf{x}_1, \mathbf{x}_2] = [\alpha\mathbf{x}_1, \alpha\mathbf{x}_2].$$

如果  $X = \mathbb{R}^n$ ，那么向量  $\mathbf{x}_1$  由  $\mathbb{C}^n$  中复数坐标的实部组成，向量  $\mathbf{x}_2$  由虚部组成。因此，非正式地说，我们可以将对  $[\mathbf{x}_1, \mathbf{x}_2]$  写成  $\mathbf{x}_1 + i\mathbf{x}_2$ 。

为了定义复数乘法，我们定义  $i$  的乘法为

$$i[\mathbf{x}_1, \mathbf{x}_2] = [-\mathbf{x}_2, \mathbf{x}_1]$$

（将  $[\mathbf{x}_1, \mathbf{x}_2]$  写成  $\mathbf{x}_1 + i\mathbf{x}_2$  时，我们可以看到它必须这样定义），并使用第二分配律  $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$  来定义任意复数的乘法。

如果  $X$  是一个内积空间，我们还可以将内积扩展到  $X_{\mathbb{C}}$ ：

$$([\mathbf{x}_1, \mathbf{x}_2], [\mathbf{y}_1, \mathbf{y}_2])_{X_{\mathbb{C}}} = (\mathbf{x}_1, \mathbf{y}_1)_X + (\mathbf{x}_2, \mathbf{y}_2)_X.$$

要看到一切都定义良好，最简单的方法是固定一组基（在实内积空间的情况下是标准正交基），然后检验坐标表示下会发生什么。我们可以看到，如果我们把向量  $\mathbf{x}_1$  看作由复数坐标的实部组成的向量，把向量  $\mathbf{x}_2$  看作由坐标的虚部组成的向量，那么这个构造恰好是  $\mathbb{R}^n$  的标准复化（通过允许复数坐标），正如上面描述的那样。

<sup>41</sup>实子空间是指在实数加法和实数乘法下封闭的集合。

我们可以用坐标无关的方式来解释这个构造，而不必选取基并处理坐标，这意味着结果不依赖于基的选择。

所以，思考复化的最简单方法可能是这样的：

要构造一个实向量空间  $X$  的复化，我们可以选取一组基（如果  $X$  是实内积空间，则选取一组标准正交基），然后处理坐标，允许复数坐标。结果空间不依赖于基的选择；我们可以通过标准的坐标变换公式从一个坐标集得到另一个。

注意，任何实空间  $X$  中的线性变换  $T$  都会产生其复化  $X_{\mathbb{C}}$  中的一个线性变换  $T_{\mathbb{C}}$ 。

看到这一点最简单的方法是固定  $X$  中的一组基（如果  $X$  是实内积空间，则选取一组标准正交基），并以坐标表示进行处理：在这种情况下， $T_{\mathbb{C}}$  与  $T$  具有相同的矩阵。在抽象表示中，我们可以写成

$$T_{\mathbb{C}}[\mathbf{x}_1, \mathbf{x}_2] = [T\mathbf{x}_1, T\mathbf{x}_2].$$

另一方面，并非所有  $X_{\mathbb{C}}$  中的线性变换都可以从  $X$  中的变换得到；如果我们进行坐标复化，只有具有实数矩阵的变换才有效。

请注意，这与第 8.1 节中描述的反复化的情景完全相反。

细心的读者可能已经注意到，复化和反复化的操作不是彼此的逆。首先，空间及其复化具有相同的维度，而一个  $n$  维空间反复化的维度为  $2n$ 。此外，正如我们刚才讨论的，实数和复数线性变换之间的关系在这些情况下是完全相反的。

在下一节中，我们将讨论一个操作，在某种意义上是反复化的逆。

### 8.3. 向实数空间引入复数结构

本节介绍的构造仅适用于偶数维的实数空间。

#### 8.3.1. 引入复数结构的初等方法

设  $X$  是一个  $2n$  维的实内积空间。我们想通过引入  $X$  上的复数结构来逆转反复化过程，即对应这个空间与一个复数空间，使其反复化（见第 8.1 节）得到原始空间  $X$ 。最简单的想法是固定  $X$  中的一组标准正交基，然后将该基下的坐标分成两半。

我们将一半坐标（例如，坐标  $x_1, x_2, \dots, x_n$ ）视为复数坐标的实部，并将其余部分视为虚部。然后我们需要将实部和虚部组合起来：例如，如果我们让  $x_1, x_2, \dots, x_n$  作为实部， $x_{n+1}, x_{n+2}, \dots, x_{2n}$  作为虚部，那么可以定义复数坐标  $z_k = x_k + ix_{n+k}$ 。

当然，结果通常取决于标准正交基的选择，以及我们如何分割实数坐标为实部和虚部，以及如何将它们组合起来。

从 8.1 节描述的反复化构造也可以看出, 所有实内积空间  $X$  上的复数结构都可以通过这种方式获得。

### 8.3.2. 从初等到抽象构造复数结构

上述构造可以用抽象的、坐标无关的方式来描述。即, 设我们将空间  $X$  分解为  $X = E \oplus E^\perp$ , 其中  $E$  是一个子空间,  $\dim E = n$  (因此  $\dim E^\perp = n$ ), 并且设  $U_0: E \rightarrow E^\perp$  是一个酉 (更确切地说, 是正交的, 因为我们的空间是实数的) 变换。

注意, 如果  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $E$  中的一组标准正交基, 那么系统  $\{U_0\mathbf{v}_1, U_0\mathbf{v}_2, \dots, U_0\mathbf{v}_n\}$  是  $E^\perp$  中的一组标准正交基, 因此

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n, U_0\mathbf{v}_1, U_0\mathbf{v}_2, \dots, U_0\mathbf{v}_n\}$$

是整个空间  $X$  中的一组标准正交基。

如果  $x_1, x_2, \dots, x_{2n}$  是该基下向量  $\mathbf{x}$  的坐标, 并且我们将  $x_k + ix_{n+k}$ ,  $k = 1, 2, \dots, n$  作为  $\mathbf{x}$  的复数坐标, 那么  $i$  的乘法由正交变换  $U$  表示, 该变换在子空间  $E, E^\perp$  的正交基下由块对角矩阵

$$U = \begin{pmatrix} \mathbf{0} & -U_0^* \\ U_0 & \mathbf{0} \end{pmatrix}$$

给出。这意味着

$$i \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = U \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & -U_0^* \\ U_0 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix},$$

$\mathbf{x}_1 \in E, \mathbf{x}_2 \in E^\perp$ .

显然,  $U$  是一个正交变换, 且  $U^2 = -I$ . 因此, 任何实内积空间  $X$  上的复数结构都由满足  $U^2 = -I$  的正交变换  $U$  给出; 变换  $U$  赋予了我们虚数单位  $i$  的乘法。

反之亦然, 即任何满足  $U^2 = -I$  的正交变换  $U$  都可以给出一个实内积空间  $X$  上的复结构。让我们来解释一下。

### 8.3.3. 复结构的抽象构造

我们先考虑一个抽象的解释。要定义一个复结构, 我们需要定义向量与复数的乘法 (最初我们只能与实数相乘)。实际上, 我们只需要定义与  $i$  的乘法, 其余的运算将从原实空间中的线性性质推导出来。而与  $i$  的乘法由满足  $U^2 = -I$  的正交变换  $U$  给出。也就是说, 如果与  $i$  的乘法由  $U$  给出, 即  $i\mathbf{x} = U\mathbf{x}$ , 那么复数乘法必须由下式定义:

$$(8.2) \quad (\alpha + \beta i)\mathbf{x} := \alpha\mathbf{x} + \beta U\mathbf{x} = (\alpha I + \beta U)\mathbf{x}, \quad \alpha, \beta \in \mathbb{R}, \quad \mathbf{x} \in X.$$

我们将使用这个公式来定义复数乘法。

不难核验出, 对于由 (8.2) 定义的复数乘法, 所有复向量空间的公理都得到了满足。例如, 可以利用实空间  $X$  中的线性性质, 并注意到, 关于代数运算 (加法和乘法), 形式为

$$\alpha I + \beta U, \quad \alpha, \beta \in \mathbb{R},$$

的线性变换的行为方式与复数  $\alpha + \beta i$  完全相同, 即这样的变换给了我们复数域的一个表示。

这意味着首先, 形式为  $\alpha I + \beta U$  的变换的和与积是相同形式的变换, 并且为了得到结果的系数  $\alpha, \beta$ , 我们可以对相应的复数执行运算并取结果的实部和虚部。注意, 这里我们需要  $U^2 = -I$  的恒等式, 但我们不需要  $U$  是正交变换的事实。

因此, 我们得到了一个复向量空间的结构。为了得到一个复内积空间, 我们需要引入一个复内积, 使得原始实内积是它的实部。

我们在这里确实没有其他选择: 注意到对于复内积

$$\operatorname{Im}(\mathbf{x}, \mathbf{y}) = \operatorname{Re}[-i(\mathbf{x}, \mathbf{y})_{\mathbb{R}}] = \operatorname{Re}((\mathbf{x}, i\mathbf{y})_{\mathbb{R}}),$$

我们发现定义复内积的唯一方法是

$$(8.3) \quad (\mathbf{x}, \mathbf{y})_{\mathbb{C}} := (\mathbf{x}, \mathbf{y})_{\mathbb{R}} + i(\mathbf{x}, U\mathbf{y})_{\mathbb{R}}.$$

我们来证明这是一个内积。我们需要  $U^* = -U$  的事实, 见下面的练习 8.4 (这里的  $U^*$  是指相对于原始实内积的伴随)。

为了证明  $(\mathbf{y}, \mathbf{x})_{\mathbb{C}} = \overline{(\mathbf{x}, \mathbf{y})_{\mathbb{C}}}$ , 我们使用恒等式  $U^* = -U$  和实内积的对称性:

$$\begin{aligned} (\mathbf{y}, \mathbf{x})_{\mathbb{C}} &= (\mathbf{y}, \mathbf{x})_{\mathbb{R}} + i(\mathbf{y}, U\mathbf{x})_{\mathbb{C}} \\ &= (\mathbf{x}, \mathbf{y})_{\mathbb{R}} + i(U\mathbf{x}, \mathbf{y})_{\mathbb{R}} \\ &= (\mathbf{x}, \mathbf{y})_{\mathbb{R}} - i(\mathbf{x}, U\mathbf{y})_{\mathbb{R}} \\ &= \overline{(\mathbf{x}, \mathbf{y})_{\mathbb{R}} + i(\mathbf{x}, U\mathbf{y})_{\mathbb{R}}} \\ &= \overline{(\mathbf{x}, \mathbf{y})_{\mathbb{C}}}. \end{aligned}$$

为了证明复内积的线性性质, 让我们首先注意到  $(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$  在第一个 (实际上是每个) 参数上是**实线性的**, 即对于  $\alpha, \beta \in \mathbb{R}$ ,  $(\alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z})_{\mathbb{C}} = \alpha(\mathbf{x}, \mathbf{z})_{\mathbb{C}} + \beta(\mathbf{y}, \mathbf{z})_{\mathbb{C}}$ ; 这是正确的, 因为右侧的每个加数在参数上都是实线性的。

使用  $(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$  的实线性以及  $U^* = -U$  (这意味着  $(U\mathbf{x}, \mathbf{y})_{\mathbb{R}} = -(\mathbf{x}, U\mathbf{y})_{\mathbb{R}}$ ) 以及



$U$  的正交性, 我们得到以下等式链:

$$\begin{aligned}
 (\alpha I + \beta U)\mathbf{x}, \mathbf{y})_{\mathbb{C}} &= (\alpha \mathbf{x}, \mathbf{y})_{\mathbb{C}} + \beta (U\mathbf{x}, \mathbf{y})_{\mathbb{C}} \\
 &= \alpha (\mathbf{x}, \mathbf{y})_{\mathbb{C}} + \beta [(U\mathbf{x}, \mathbf{y})_{\mathbb{R}} + i(U\mathbf{x}, U\mathbf{y})_{\mathbb{R}}] \\
 &= \alpha (\mathbf{x}, \mathbf{y})_{\mathbb{C}} + \beta [-(\mathbf{x}, U\mathbf{y})_{\mathbb{R}} + i(\mathbf{x}, \mathbf{y})_{\mathbb{R}}] \\
 &= \alpha (\mathbf{x}, \mathbf{y})_{\mathbb{C}} + \beta i[(\mathbf{x}, \mathbf{y})_{\mathbb{R}} + i(\mathbf{x}, U\mathbf{y})_{\mathbb{R}}] \\
 &= \alpha (\mathbf{x}, \mathbf{y})_{\mathbb{C}} + \beta i(\mathbf{x}, \mathbf{y})_{\mathbb{C}} \\
 &= (\alpha + \beta i)(\mathbf{x}, \mathbf{y})_{\mathbb{C}}.
 \end{aligned}$$

这证明了复线性。

最后, 为了证明  $(\mathbf{x}, \mathbf{x})_{\mathbb{C}}$  的非负性, 让我们注意到 (见练习 8.3)  $(\mathbf{x}, U\mathbf{x})_{\mathbb{R}} = 0$ , 所以

$$(\mathbf{x}, \mathbf{x})_{\mathbb{C}} = (\mathbf{x}, \mathbf{x})_{\mathbb{R}} = \|\mathbf{x}\|^2 \geq 0.$$

#### 8.3.4. 通过初等方法进行的抽象构造

对于不习惯如此“高明”和抽象的证明的读者, 还有另一种更实际的解释。

即, 可以证明 (见练习 8.5), 存在一个子空间  $E$ ,  $\dim E = n$  (回想一下  $\dim X = 2n$ ), 使得在分解  $X = E \oplus E^{\perp}$  下  $U$  的矩阵由块对角矩阵

$$U = \begin{pmatrix} \mathbf{0} & -U_0^* \\ U_0 & \mathbf{0} \end{pmatrix}$$

给出, 其中  $U_0 : E \rightarrow E^{\perp}$  是某个正交变换。

设  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $E$  中的一组标准正交基。那么  $\{U_0\mathbf{v}_1, U_0\mathbf{v}_2, \dots, U_0\mathbf{v}_n\}$  是  $E^{\perp}$  中的一组标准正交基, 所以

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n, U_0\mathbf{v}_1, U_0\mathbf{v}_2, \dots, U_0\mathbf{v}_n\}$$

是整个空间  $X$  中的一组标准正交基。考虑该基下的坐标  $x_1, x_2, \dots, x_{2n}$ , 并将  $x_k + ix_{n+k}$ ,  $k = 1, 2, \dots, n$  作为  $\mathbf{x}$  的复数坐标, 那么  $i$  的乘法由变换  $U$  来表示: 它对于  $\mathbf{x} \in E$  是平凡的, 对于  $\mathbf{y} \in E^{\perp}$  也是平凡的, 因此它对于所有实数线性组合  $\alpha\mathbf{x} + \beta\mathbf{y}$  都成立, 即对于  $X$  中的所有向量都成立。

但这意味着抽象复数结构的引入以及相应的初等方法给出了相同的结果! 而且, 由于初等方法清楚地给出复数结构, 抽象方法也给出了相同的复数结构。

#### 练习

8.1. 证明公式 (8.1)。即, 证明如果

$$\mathbf{x} = (z_1, z_2, \dots, z_n)^T, \quad \mathbf{y} = (w_1, w_2, \dots, w_n)^T,$$

$z_k = x_k + iy_k, w_k = u_k + iv_k, x_k, y_k, u_k, v_k \in \mathbb{R}$ , 那么

$$\operatorname{Re}\left(\sum_{k=1}^n z_k \overline{w_k}\right) = \sum_{k=1}^n x_k u_k + \sum_{k=1}^n y_k v_k.$$

8.2. 证明如果  $(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$  是复内积空间中的内积, 那么由 (8.1) 定义的  $(\mathbf{x}, \mathbf{y})_{\mathbb{R}}$  是一个实内积空间。

8.3. 设  $U$  是一个满足  $U^2 = -I$  的正交变换 (在实内积空间  $X$  中)。证明对于所有  $\mathbf{x} \in X$ ,

$$U\mathbf{x} \perp \mathbf{x}.$$

8.4. 证明, 如果  $U$  是一个满足  $U^2 = -I$  的正交变换, 那么  $U^* = -U$ 。

8.5. 设  $U$  是一个满足  $U^2 = -I$  的实内积空间中的正交变换。证明在这种情况下  $\dim X = 2n$ , 并且存在一个子空间  $E \subset X$ ,  $\dim E = n$ , 以及一个正交变换  $U_0 : E \rightarrow E^\perp$ , 使得在  $X = E \oplus E^\perp$  的分解下,  $U$  由块对角矩阵

$$U = \begin{pmatrix} \mathbf{0} & -U_0^* \\ U_0 & \mathbf{0} \end{pmatrix}$$

给出。这个陈述可以很容易地从第六章定理 5.1 得到, 如果我们注意到  $\mathbb{R}^2$  中的唯一满足  $R_\alpha^2 = -I$  的旋转  $R_\alpha$  是角度为  $\pm\pi/2$  的旋转。

但是, 可以找到一个初等的证明, 而无需使用该定理。例如, 该陈述在  $\dim X = 2$  时是平凡的: 在这种情况下, 我们可以选任何一维子空间作为  $E$ , 见练习 8.3。

然后, 不难证明, 这样的变换于  $\mathbb{R}^2$  中  $U$  不存在, 并且我们可以通过归纳  $\dim X$  来完成证明。

## 第六章 内积空间中算子的结构

在本章中，我们再次假设所有空间都是有限维的。同样，我们只处理复数或实数空间，内积空间的理论不适用于任意域上的空间。当没有提及我们所处的空间时，所有结果都适用于复数和实数空间。

为了避免重复书写基本相同的公式，我们将使用复数情况的记号：在实数情况下，它给出正确但有时稍显复杂的公式。

### 1. 算子的上三角（舒尔）表示

**定理 1.1.** 设  $A: X \rightarrow X$  是作用在复内积空间中的算子。则在  $X$  中存在一组标准正交基  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ ，使得  $A$  在该基下的矩阵是上三角矩阵。

换句话说，任何  $n \times n$  矩阵  $A$  都可以表示为  $A = UTU^*$ ，其中  $U$  是酉矩阵，而  $T$  是上三角矩阵。

**证明：** 我们使用  $\dim X$  的数学归纳法来证明定理。如果  $\dim X = 1$ ，则定理是平凡的，因为任何  $1 \times 1$  矩阵都是上三角矩阵。

假设当  $\dim X = n - 1$  时定理成立，并且我们想证明对于  $\dim X = n$  时定理成立。

设  $\lambda_1$  是  $A$  的一个特征值，设  $\mathbf{u}_1, \|\mathbf{u}_1\| = 1$  是相应的特征向量， $A\mathbf{u}_1 = \lambda_1\mathbf{u}_1$ 。记  $E = \mathbf{u}_1^\perp$ ，并设  $\{\mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $E$  中的某组标准正交基（显然  $\dim E = \dim X - 1 = n - 1$ ），那么  $\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $X$  中的一组标准正交基。在这一基下， $A$  的矩阵具有形式

$$(1.1) \quad \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix};$$

这里  $\lambda_1$  下方的所有元素都是零，而  $*$  表示我们不关心  $\lambda_1$  右侧的元素。

我们足够关心右下角的  $(n - 1) \times (n - 1)$  块，以至于给它命名：我们将其记为  $A_1$ 。

注意， $A_1$  定义了  $E$  中的一个线性变换，并且由于  $\dim E = n - 1$ ，归纳假设表明存在一组标准正交基（我们将其记为  $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ ），使得  $A_1$  在该基下的矩阵

是上三角矩阵。

所以,  $A$  的矩阵在正交基  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  具有形式 (1.1), 其中矩阵  $A_1$  是上三角矩阵。因此,  $A$  在该标准正交基  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  下的矩阵也是上三角矩阵。□

**注记** 注意, 在证明中引入的子空间  $E = \mathbf{u}_1^\perp$  对于  $A$  不是不变的, 即  $AE \subseteq E$  不一定成立。这意味着  $A_1$  不是  $A$  的一部分, 它是从  $A$  构建的某个算子。

还需注意,  $AE \subseteq E$  当且仅当所有标记为  $*$  的元素 (即除了  $\lambda_1$  之外的第一行的所有元素) 都为零。

**注记** 注意, 即使我们从一个实数矩阵  $A$  开始, 矩阵  $U$  和  $T$  也可以是复数的。旋转矩阵

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad \alpha \neq k\pi, k \in \mathbb{Z}$$

与实数上三角矩阵不是酉等价的 (甚至不是相似的)。因为该矩阵的特征值是复数, 而上三角矩阵的特征值是对角线元素。

**注记** 定理 1.1 的一个类似版本可以陈述于任意向量空间, 而不要求它具有内积。在这种情况下, 定理声称在某组基下任何算子都有上三角形式。我们可以通过模仿定理 1.1 的证明来完成此证明。另一种方法是为  $V$  装备内积, 方法是固定一组基并声明它是标准正交基, 见第 5 章第 2.4 节。

注意, 内积空间版本 (定理 1.1) 比向量空间版本更强大, 因为它说明我们总能找到一组标准正交基, 而不仅仅是一组基。

下面的定理是定理 1.1 的实数版本:

**定理 1.2.** 设  $A: X \rightarrow X$  是作用在实数内积空间中的算子。假设  $A$  的所有特征值都是实数 (意味着  $A$  恰好有  $n = \dim X$  个实数特征值, 计入重数)。那么  $X$  中存在标准正交基  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ , 使得  $A$  在该基下的矩阵是上三角矩阵。

换句话说, 任何特征值均为实数的  $n \times n$  实矩阵  $A$  都可以表示为  $A = UTU^* = UTU^T$ , 其中  $U$  是正交矩阵, 而  $T$  是实数上三角矩阵。

**证明:** 为了证明定理, 我们只需要分析定理 1.1 的证明。不妨假设 (我们可以无损于一般性地这样做) 算子 (矩阵)  $A$  作用在  $\mathbb{R}^n$  上。

假设定理对  $(n-1) \times (n-1)$  矩阵成立。在定理 1.1 的证明中, 设  $\lambda_1$  是  $A$  的一个实特征值,  $\mathbf{u}_1 \in \mathbb{R}^n, \|\mathbf{u}_1\| = 1$  是相应的特征向量, 设  $\{\mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $\mathbb{R}^n$  中的一个标准正交系统, 使得  $\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $\mathbb{R}^n$  中的一组标准正交基。

在该基下  $A$  的矩阵具有形式 (1.1), 其中  $A_1$  是某个实数矩阵。

如果我们能证明矩阵  $A_1$  只有实数特征值，那么我们就完成了证明。确实，根据归纳假设，存在  $E = \mathbf{u}_1^\perp$  中的一组标准正交基  $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ ，使得  $A_1$  在该基下的矩阵是上三角矩阵，因此  $A$  在  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  基下的矩阵也是上三角矩阵。

为了证明  $A_1$  只有实数特征值，让我们注意到

$$\det(A - \lambda I) = (\lambda_1 - \lambda) \det(A_1 - \lambda)$$

(例如，通过第一行的代数余子式展开)，所以  $A_1$  的任何特征值也是  $A$  的特征值。但是  $A$  只有实数特征值！□

### 练习

1.1. 利用算子的上三角表示，给出行列式是乘积，迹是计重数的特征值之和这一事实的另一种证明。

## 2. 自伴随和正规算子的谱定理

在本章中，我们处理的酉等价于对角矩阵的矩阵（算子）。

让我们回忆一下，如果一个算子满足  $A = A^*$ ，则称其为**自伴随**的。在某组标准正交基下的自伴（随）算子（即满足  $A^* = A$  的矩阵）称为**埃尔米特矩阵**。术语“自伴随”和“埃尔米特”其实是同义的。但通常人们在谈论算子（变换）时会说自伴随，而在谈论矩阵时说埃尔米特。我们将尝试遵循这个约定，但由于我们经常不区分算子和它们的矩阵，所以有时会混合使用这两个术语。

**定理 2.1.** 设  $A = A^*$  是内积空间  $X$ （空间可以是复数或实数）中的一个自伴随算子。那么  $A$  的所有特征值都是实数，并且  $X$  中存在由  $A$  特征向量组成的标准正交基。

这个定理可以用矩阵形式重述如下：

**定理 2.2.** 设  $A = A^*$  是一个自伴随（因此是方阵）矩阵。那么  $A$  可以表示为

$$A = UDU^*,$$

其中  $U$  是酉矩阵， $D$  是实对角矩阵。

而且，如果矩阵  $A$  是实数的，矩阵  $U$  可以选择为实数的（即正交的）。

**证明：** 为了证明定理 2.1 和定理 2.2，我们首先对内积空间  $X$ （或实数空间  $X$ ）应用定理 1.1（或定理 1.2）来找到一组标准正交基，使得  $A$  在该基下的矩阵是上三角矩阵。现在让我们问自己一个问题：什么样的上三角矩阵是自伴随的？

答案是显而易见的：上三角矩阵是自伴随的当且仅当它是具有实数项的对角矩阵。定理 2.1（以及因此定理 2.2）得证。□

**注记** 注意, 在许多教科书中只考虑实数矩阵, 并且定理 2.2 通常被称为“**对称矩阵的谱定理**”。然而, 应该强调, 定理 2.2 的结论对于**复**对称矩阵是不成立的: 该定理适用于埃尔米特矩阵, 特别地, **实**对称矩阵。

让我们给出一个  $A = A^*$  的算子的特征值是实数的独立证明。设  $A = A^*$  且  $A\mathbf{x} = \lambda\mathbf{x}$ ,  $\mathbf{x} \neq 0$ . 那么

$$(A\mathbf{x}, \mathbf{x}) = (\lambda\mathbf{x}, \mathbf{x}) = \lambda(\mathbf{x}, \mathbf{x}) = \lambda\|\mathbf{x}\|^2.$$

另一方面,

$$(A\mathbf{x}, \mathbf{x}) = (\mathbf{x}, A^*\mathbf{x}) = (\mathbf{x}, A\mathbf{x}) = (\mathbf{x}, \lambda\mathbf{x}) = \bar{\lambda}(\mathbf{x}, \mathbf{x}) = \bar{\lambda}\|\mathbf{x}\|^2,$$

(这里我们用到了结论  $(\mathbf{x}, \lambda\mathbf{y}) = \bar{\lambda}(\mathbf{x}, \mathbf{y})$ )。所以  $\lambda\|\mathbf{x}\|^2 = \bar{\lambda}\|\mathbf{x}\|^2$ . 由于  $\mathbf{x} \neq 0$ ,  $\|\mathbf{x}\|^2 \neq 0$ , 我们可以得出  $\lambda = \bar{\lambda}$ , 所以  $\lambda$  是实数。

从定理 2.1 也可以得出, 自伴随算子的特征子空间是相互正交的。但让我们给出一个该结果的独立证明。

**命题 2.3** 设  $A = A^*$  是一个自伴随算子, 设  $\mathbf{u}, \mathbf{v}$  是它的特征向量,  $A\mathbf{u} = \lambda\mathbf{u}$ ,  $A\mathbf{v} = \mu\mathbf{v}$ . 那么, 如果  $\lambda \neq \mu$ , 则特征向量  $\mathbf{u}$  和  $\mathbf{v}$  是正交的。

**证明:** 这个命题虽然可以从谱定理 (定理 1.1) 得出, 但我们在这里给出一个直接的证明。即,

$$(A\mathbf{u}, \mathbf{v}) = (\lambda\mathbf{u}, \mathbf{v}) = \lambda(\mathbf{u}, \mathbf{v}).$$

另一方面,

$$(A\mathbf{u}, \mathbf{v}) = (\mathbf{u}, A^*\mathbf{v}) = (\mathbf{u}, A\mathbf{v}) = (\mathbf{u}, \mu\mathbf{v}) = \mu(\mathbf{u}, \mathbf{v})$$

(最后一个等式成立是因为自伴随算子的特征值都是实数), 所以  $\lambda(\mathbf{u}, \mathbf{v}) = \mu(\mathbf{u}, \mathbf{v})$ . 如果  $\lambda \neq \mu$ , 则这只有在  $(\mathbf{u}, \mathbf{v}) = 0$  时才可能。□

现在让我们尝试找到哪些矩阵酉等价于一个对角矩阵。可以很容易地检验出, 对于对角矩阵  $D$ ,

$$D^*D = DD^*.$$

因此, 如果  $A$  在某组标准正交基下的矩阵是对角矩阵, 那么  $A^*A = AA^*$ .

**定义** 称算子 (矩阵)  $N$  是**正规**(normal) 的, 如果  $N^*N = NN^*$ .

显然, 任何自伴随算子 ( $A^*A = AA^*$ ) 都是正规的。同样, 任何酉算子  $U: X \rightarrow X$  也是正规的, 因为  $U^*U = UU^* = I$ .

注意, 正规算子是作用在同一个空间上的算子, 而不是从一个空间到另一个空间。所以, 如果  $U$  是作用在一个空间到另一个空间上的酉算子, 我们就不能说  $U$  是正规的。

**定理 2.4.** 任何复数向量空间中的正规算子  $N$  都有一组由标准正交的特征向量组成的基。

换句话说, 任何满足  $N^*N = NN^*$  的矩阵  $N$  都可以表示为

$$N = UDU^*,$$

其中  $U$  是酉矩阵,  $D$  是对角矩阵。

**注记** 注意, 在上述定理中, 即使  $N$  是实数矩阵, 我们也没有声称矩阵  $U$  和  $D$  是实数。而且, 可以很容易地证明, 如果  $D$  是实数, 那么  $N$  必须是自伴随的。

**定理 2.4 的证明** 为了证明定理 2.4, 我们应用定理 1.1 来得到一组标准正交基, 使得  $N$  在该基下的矩阵是上三角矩阵。为了完成定理的证明, 我们只需要证明一个上三角正规矩阵必须是对角矩阵。

我们将使用对矩阵维数的数学归纳法来证明这一点。  $1 \times 1$  矩阵的情况是平凡的, 因为任何  $1 \times 1$  矩阵都是对角矩阵。

假设我们已经证明了任何  $(n-1) \times (n-1)$  的上三角正规矩阵都是对角矩阵, 并且我们想证明对于  $n \times n$  矩阵也成立。设  $N$  是一个  $n \times n$  上三角正规矩阵。我们可以将其写成

$$N = \left( \begin{array}{c|ccc} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \hline 0 & & & \\ \vdots & & N_1 & \\ 0 & & & \end{array} \right)$$

其中  $N_1$  是一个  $(n-1) \times (n-1)$  的上三角矩阵。

让我们比较  $N^*N$  和  $NN^*$  的左上角元素 (第一行第一列)。直接计算表明

$$(N^*N)_{1,1} = \bar{a}_{1,1}a_{1,1} = |a_{1,1}|^2,$$

而

$$(NN^*)_{1,1} = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2.$$

所以,  $(N^*N)_{1,1} = (NN^*)_{1,1}$  当且仅当  $a_{1,2} = \cdots = a_{1,n} = 0$ . 因此, 矩阵  $N$  具有形式

$$N = \left( \begin{array}{c|ccc} a_{1,1} & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & N_1 & \\ 0 & & & \end{array} \right).$$

从上述表示可以得出

$$N^*N = \left( \begin{array}{c|ccc} |a_{1,1}|^2 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & N_1^*N_1 & \\ 0 & & & \end{array} \right), \quad NN^* = \left( \begin{array}{c|ccc} |a_{1,1}|^2 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & N_1N_1^* & \\ 0 & & & \end{array} \right).$$

所以  $N_1^*N_1 = N_1N_1^*$ . 这意味着矩阵  $N_1$  也是正规的, 并且根据归纳假设它是对角矩阵. 所以矩阵  $N$  也是对角矩阵.  $\square$

以下命题给出了正规算子的一个非常有用的刻画。

**命题 2.5** 算子  $N: X \rightarrow X$  是正规的当且仅当

$$\|N\mathbf{x}\| = \|N^*\mathbf{x}\| \quad \forall \mathbf{x} \in X.$$

**证明:** 设  $N$  是正规的,  $N^*N = NN^*$ . 那么

$$\|N\mathbf{x}\|^2 = (N\mathbf{x}, N\mathbf{x}) = (N^*N\mathbf{x}, \mathbf{x}) = (NN^*\mathbf{x}, \mathbf{x}) = (N^*\mathbf{x}, N^*\mathbf{x}) = \|N^*\mathbf{x}\|^2,$$

所以  $\|N\mathbf{x}\| = \|N^*\mathbf{x}\|$ . 现在设

$$\|N\mathbf{x}\| = \|N^*\mathbf{x}\| \quad \forall \mathbf{x} \in X.$$

极化恒等式 (第 5 章引理 1.9) 暗示对于所有  $\mathbf{x}, \mathbf{y} \in X$ ,

$$\begin{aligned} (N^*N\mathbf{x}, \mathbf{y}) &= (N\mathbf{x}, N\mathbf{y}) = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|N\mathbf{x} + \alpha N\mathbf{y}\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|N(\mathbf{x} + \alpha\mathbf{y})\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|N^*(\mathbf{x} + \alpha\mathbf{y})\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|N^*\mathbf{x} + \alpha N^*\mathbf{y}\|^2 \\ &= (N^*\mathbf{x}, N^*\mathbf{y}) = (NN^*\mathbf{x}, \mathbf{y}). \end{aligned}$$

因此 (见推论 1.6),  $N^*N = NN^*$ .  $\square$

## 练习

### 2.1. 判断正误:

- 任何酉算子  $U: X \rightarrow X$  都是正规的。
- 矩阵是酉的当且仅当它是可逆的。
- 如果两个矩阵酉等价, 那么它们也相似。
- 两个自伴随算子之和是自伴随的。
- 酉算子的伴随是酉的。
- 正规算子的伴随是正规的。
- 如果一个线性算子的所有特征值都是 1, 那么该算子必须是酉的或正交的。



- h) 如果一个正规算子的所有特征值都是 1, 那么该算子是恒等 (单位) 算子。  
i) 线性算子可能保持范数但不保持内积。

2.2. 证明或证伪: 两个正规算子之和也是正规的。

2.3. 证明一个酉等价于对角矩阵的矩阵是正规的。

2.4. 正交对角化矩阵

$$\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

找出  $A$  的所有平方根, 即找出所有满足  $B^2 = A$  的矩阵  $B$ . **注记:**  $A$  的所有平方根都是自伴随的。

2.5. 判断正误: 任何自伴随矩阵都有一个自伴随的平方根。证明你的结论。

2.6. 正交对角化矩阵

$$A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix},$$

即将其表示为  $A = UDU^*$ , 其中  $D$  是对角矩阵,  $U$  是酉矩阵。

在  $A$  的所有平方根中, 找出具有正特征值的平方根。你可以将  $B$  表示为乘积形式。

2.7. 判断正误:

- a) 两个自伴随矩阵的乘积是自伴随的。  
b) 如果  $A$  是自伴随的, 那么  $A^k$  是自伴随的。证明你的结论。

2.8. 设  $A$  是  $m \times n$  矩阵。证明:

- a)  $A^*A$  是自伴随的。  
b)  $A^*A$  的所有特征值都是非负的。  
c)  $A^*A + I$  是可逆的。

2.9. 证明或证伪:

- a) 如果  $A$  是自伴随的, 那么  $A + iI$  是可逆的。  
b) 如果  $U$  是酉的,  $U + \frac{3}{4}I$  是可逆的。  
c) 如果矩阵  $A$  是实的, 那么  $A - iI$  是可逆的。

2.10. 正交对角化旋转矩阵

$$R_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix},$$

其中  $\alpha$  不是  $\pi$  的整数倍。注意, 在这种情况下你会得到复数特征值。

2.11. 正交对角化矩阵

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

**提示:** 你会得到实数特征值。此外, 三角恒等式  $\sin^2 x = 2 \sin x \cos x$ ,  $\sin^2 x = (1 - \cos 2x)/2$ ,  $\cos^2 x = (1 + \cos 2x)/2$  (应用于  $x = \alpha/2$ ) 将有助于简化特征向量的表达式。

2.12. 你能从几何上描述上一问题中矩阵  $A$  所代表的线性变换吗? 它有一个非常简单的几何解释。

2.13. 证明具有模为 1 的特征值 (即所有特征值满足  $|\lambda_k| = 1$ ) 的正规算子是酉的。

**提示:** 考虑对角化。

2.14. 证明具有实数特征值的正规算子是自伴随的。

2.15. 举例说明定理 2.2 的结论对于复数对称矩阵不成立。即:

- a) 构建一个 (可对角化的)  $2 \times 2$  复数对称矩阵, 它不容许一组正交的特征向量基;  
b) 构建一个  $2 \times 2$  复数对称矩阵, 它不能被对角化。

### 3. 极分解与奇异值分解

#### 3.1. 正定算子 平方根

**定义** 称自伴随算子  $A: X \rightarrow X$  为**正定的**, 如果

$$(A\mathbf{x}, \mathbf{x}) > 0 \quad \forall \mathbf{x} \neq 0,$$

称其为**半正定的**, 如果

$$(A\mathbf{x}, \mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in X.$$

我们将使用记号  $A > 0$  表示正定算子,  $A \geq 0$  表示半正定算子。

下面的定理描述了正定和半正定算子。

**定理 3.1.** 设  $A = A^*$ . 那么

1.  $A > 0$  当且仅当  $A$  的所有特征值都是正的。
2.  $A \geq 0$  当且仅当  $A$  的所有特征值都是非负的。

**证明:** 通过选取一组标准正交基, 使得  $A$  在该基下的矩阵是对角矩阵 (见定理 2.1), 我们可以无损于一般性地证明。要完成证明, 只需注意到, 对于对角矩阵, 当且仅当其对角线元素都为正 (非负) 时, 该矩阵才是正定 (半正定) 的。□

**推论 3.2** 设  $A = A^* \geq 0$  是一个半正定算子。存在唯一一个半正定算子  $B$ , 使得  $B^2 = A$ .

这样的  $B$  被称为  $A$  的 (正) **平方根**, 并记作  $\sqrt{A}$  或  $A^{1/2}$ .

**证明:** 我们来证明  $\sqrt{A}$  的存在性。设  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $A$  的特征向量的标准正交基, 并设  $\lambda_1, \lambda_2, \dots, \lambda_n$  是相应的特征值。注意, 由于  $A \geq 0$ , 所有  $\lambda_k \geq 0$ .

在基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  下,  $A$  的矩阵是对角矩阵  $\text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ , 对角线上是  $\lambda_1, \lambda_2, \dots, \lambda_n$ . 定义  $B$  在同一基下的矩阵为  $\text{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\}$ .

显然,  $B = B^* \geq 0$  且  $B^2 = A$ .

为了证明  $B$  的唯一性, 让我们假设存在一个算子  $C = C^* \geq 0$  使得  $C^2 = A$ . 设  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  是  $C$  的特征向量的标准正交基, 并设  $\mu_1, \mu_2, \dots, \mu_n$  是相应的特征值 (注意  $\mu_k \geq 0 \quad \forall k$ )。  $C$  在该基下的矩阵是对角矩阵  $\text{diag}\{\mu_1, \mu_2, \dots, \mu_n\}$ , 因此  $A = C^2$  在同一基下的矩阵是  $\text{diag}\{\mu_1^2, \mu_2^2, \dots, \mu_n^2\}$ . 这暗示  $A$  的任何特征值  $\lambda$  都必须是  $\mu_k^2$  的形式, 并且, 更重要的是, 如果  $A\mathbf{x} = \lambda\mathbf{x}$ , 那么  $C\mathbf{x} = \sqrt{\lambda}\mathbf{x}$ .

因此, 在上面的基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  下,  $C$  的矩阵是对角矩阵  $\text{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\}$ , 即  $B = C$ . □

### 3.2. 算子的模 奇异值

考虑算子  $A : X \rightarrow Y$ . 它的埃尔米特平方  $A^*A$  是作用在  $X$  上的半正定算子。确实，

$$(A^*A)^* = A^*(A^*)^* = A^*A$$

并且

$$(A^*Ax, x) = (Ax, Ax) = \|Ax\|^2 \geq 0 \quad \forall x \in X.$$

因此，存在（唯一）一个半正定算子  $R = \sqrt{A^*A}$ . 这个算子  $R$  被称为算子  $A$  的模 (modulus)，通常记为  $|A|$ .

$A$  的模显示了算子  $A$  的“大小”：

**命题 3.3** 对于线性算子  $A : X \rightarrow Y$ ,

$$\||A|x\| = \|Ax\| \quad \forall x \in X.$$

**证明：** 对于任何  $x \in X$ ,

$$\begin{aligned} \||A|x\|^2 &= (|A|x, |A|x) = (|A|^*|A|x, x) = (|A|^2x, x) \\ &= (A^*Ax, x) = (Ax, Ax) = \|Ax\|^2. \end{aligned}$$

□

**推论 3.4**

$$\text{Ker } A = \text{Ker } |A| = (\text{Ran } |A|)^\perp.$$

**证明：** 第一个等式直接来自命题 3.3, 第二个等式来自恒等式  $\text{Ker } T = (\text{Ran } T^*)^\perp$  ( $|A|$  是自伴随的)。□

**定理 3.5.** (算子的极分解, polar decomposition) 设  $A : X \rightarrow X$  是一个算子 (方阵)。那么  $A$  可以表示为

$$A = U|A|,$$

其中  $U$  是酉算子。

**注记** 酉算子  $U$  通常不是唯一的。从定理的证明中可以看出,  $U$  仅在  $A$  可逆时才唯一。

**注记** 极分解  $A = U|A|$  也适用于作用在一个空间到另一个空间上的算子  $A : X \rightarrow Y$ . 但在这种情况下, 我们只能保证  $U$  是从  $\text{Ran } |A| = (\text{Ker } A)^\perp$  到  $Y$  的一个等距同构。

如果  $\dim X \leq \dim Y$ , 则此等距同构可以扩展为从整个  $X$  到  $Y$  的等距同构 (如果  $\dim X = \dim Y$ , 则它将是一个酉算子)。

**定理 3.5 的证明** 考虑向量  $\mathbf{x} \in \text{Ran } |A|$ . 那么向量  $\mathbf{x}$  可以表示为  $\mathbf{x} = |A|\mathbf{v}$  对于某个向量  $\mathbf{v} \in X$ .

定义  $U_0\mathbf{x} := A\mathbf{v}$ . 根据命题 3.3

$$\|U_0\mathbf{x}\| = \|A\mathbf{x}\| = \||A|\mathbf{v}\| = \|\mathbf{x}\|,$$

所以看起来  $U$  是从  $\text{Ran } |A|$  到  $X$  的一个等距同构。

但首先我们需要证明  $U_0$  是良好定义的。设  $\mathbf{v}_1$  是另一个使得  $\mathbf{x} = |A|\mathbf{v}_1$  的向量。但是  $\mathbf{x} = |A|\mathbf{v} = |A|\mathbf{v}_1$  意味着  $\mathbf{v} - \mathbf{v}_1 \in \text{Ker } |A| = \text{Ker } A$  (参见推论 3.4), 所以  $A\mathbf{v} = A\mathbf{v}_1$ , 这意味着  $U_0\mathbf{x}$  是良好定义的。

根据构造,  $A = U_0|A|$ . 我们将核验  $U_0$  是一个线性变换的证明留给读者。

为了将  $U_0$  扩展为酉算子  $U$ , 我们找到一个酉变换  $U_1 : \text{Ker } A \rightarrow (\text{Ran } A)^\perp = \text{Ker } A^*$ . 这样做总是可能的, 因为对于方阵,  $\dim \text{Ker } A = \dim \text{Ker } A^*$  (根据秩定理)。

很容易验证,  $U = U_0 + U_1$  是一个酉算子, 并且  $A = U|A|$ . □

### 3.3. 奇异值 施密特分解

**定义**  $|A|$  的特征值被称为  $A$  的**奇异值**(singular value). 换句话说, 如果  $\lambda_1, \lambda_2, \dots, \lambda_n$  是  $A^*A$  的特征值, 那么  $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$  就是  $A$  的奇异值。

**注记** 在许多教材、文献中, 奇异值被定义为  $A^*A$  的特征值的非负平方根, 而不提及算子  $|A|$ .

我认为算子  $|A|$  的概念很重要, 所以已经在前面介绍了。然而, 算子  $|A|$  的概念对于后续内容 (定义施密特和奇异值分解) 不是必需的。此外, 正如下面将要显示的, 算子  $|A|$  可以很容易地从奇异值分解构造出来。

设  $A : X \rightarrow Y$  是一个算子, 并设  $\sigma_1, \sigma_2, \dots, \sigma_n$  是  $A$  的奇异值 (计入重数)。假设  $\sigma_1, \sigma_2, \dots, \sigma_r$  是  $A$  的**非零**奇异值 (计入重数)。这意味着  $\sigma_k = 0$  对于  $k > r$ .

根据奇异值的定义, 数字  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  是  $A^*A$  的特征值。设  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $A^*A$  的特征向量的标准正交基,  $A^*A\mathbf{v}_k = \sigma_k^2\mathbf{v}_k$ .

**命题 3.6 系统**

$$\{\mathbf{w}_k := \frac{1}{\sigma_k} A\mathbf{v}_k, \quad k = 1, 2, \dots, r\}$$

是一个标准正交系统。

**证明:**

$$(A\mathbf{v}_j, A\mathbf{v}_k) = (A^*A\mathbf{v}_j, \mathbf{v}_k) = (\sigma_j^2\mathbf{v}_j, \mathbf{v}_k) = \sigma_j^2(\mathbf{v}_j, \mathbf{v}_k) = \begin{cases} 0, & j \neq k \\ \sigma_j^2, & j = k \end{cases},$$

因为  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  是一个标准正交系统。  $\square$

在上述命题的记号中, 算子  $A$  可以表示为

$$(3.1) \quad A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*,$$

或者等价地

$$(3.2) \quad A\mathbf{x} = \sum_{k=1}^r \sigma_k (\mathbf{x}, \mathbf{v}_k) \mathbf{w}_k.$$

确实, 我们知道  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $X$  的一组标准正交基。那么将  $\mathbf{x} = \mathbf{v}_j$  代入 (3.2) 的右侧, 我们得到

$$\sum_{k=1}^r \sigma_k (\mathbf{v}_j, \mathbf{v}_k) \mathbf{w}_k = \sigma_j (\mathbf{v}_j, \mathbf{v}_j) \mathbf{w}_j = \sigma_j \mathbf{w}_j = A\mathbf{v}_j$$

如果  $j = 1, 2, \dots, r$ , 并且

$$\sum_{k=1}^r \sigma_k (\mathbf{v}_k^* \mathbf{v}_j) \mathbf{w}_k = \mathbf{0} = A\mathbf{v}_j$$

对于  $j > r$ . 所以, (3.1) 中左右两侧的算子在基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  上是相同的, 因此它们是相等的。

**定义** 上述分解 (3.1) (或 (3.2)) 被称为算子  $A$  的**施密特分解**(Schmidt decomposition)。

**注记** 施密特分解不是唯一的。为什么?

**引理 3.7.** 设  $A$  可以被表示为

$$A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*,$$

其中  $\sigma_k > 0$  并且  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}, \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$  是某些标准正交系。

那么这个表示给出了  $A$  的施密特分解。

**证明:** 我们只需要证明  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是  $A^*A$  的特征向量,  $A^*A\mathbf{v}_k = \sigma_k^2 \mathbf{v}_k$ . 由于  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$  是标准正交系统,

$$\mathbf{w}_k^* \mathbf{w}_j = (\mathbf{w}_j, \mathbf{w}_k) = \delta_{k,j} := \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases},$$

因此

$$A^*A = \sum_{k=1}^r \sigma_k^2 \mathbf{v}_k \mathbf{v}_k^*.$$

由于  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  是标准正交系统,

$$A^* A \mathbf{v}_j = \sum_{k=1}^r \sigma_k^2 \mathbf{v}_k \mathbf{v}_k^* \mathbf{v}_j = \sigma_j^2 \mathbf{v}_j,$$

因此  $\mathbf{v}_k$  是  $A^* A$  的特征向量。 □

**推论 3.8** 设

$$A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$$

是  $A$  的施密特分解。那么

$$A^* = \sum_{k=1}^r \sigma_k \mathbf{v}_k \mathbf{w}_k^*$$

是  $A^*$  的施密特分解。

### 3.4. 施密特分解的矩阵表示 奇异值分解

施密特分解可以写成一个很好的矩阵形式。即, 假设  $A: \mathbb{F}^n \rightarrow \mathbb{F}^m$  (这里  $\mathbb{F}$  总是  $\mathbb{C}$  或  $\mathbb{R}$ ; 我们可以通过选取  $X$  和  $Y$  中的标准正交基并处理这些基下的坐标来完成)。设  $\sigma_1, \sigma_2, \dots, \sigma_r$  是  $A$  的非零奇异值 (计入重数), 并设

$$A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$$

是  $A$  的施密特分解。

如你所见, 这个等式可以重写为

$$(3.3) \quad A = \widetilde{W} \widetilde{\Sigma} \widetilde{V}^*,$$

其中  $\widetilde{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}$  并且  $\widetilde{V}$  和  $\widetilde{W}$  是分别以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  和  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$  为列的矩阵。(你能说出每个矩阵的大小吗?)

注意, 由于  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  和  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$  是标准正交系统, 矩阵  $\widetilde{V}$  和  $\widetilde{W}$  是等距同构。还需注意  $r = \text{rank } A$ , 见下面的练习 3.1。

如果矩阵  $A$  是可逆的, 那么  $m = n = r$ , 矩阵  $\widetilde{V}, \widetilde{W}$  是酉的, 并且  $\widetilde{\Sigma}$  是一个可逆的对角矩阵。

事实证明, 总是可以写出一个类似的表示 (3.3), 用酉矩阵  $V$  和  $W$  来代替  $\widetilde{V}$  和  $\widetilde{W}$ , 并且在许多情况下, 处理这样的表示会更方便。为了写出这个表示, 我们首先需要将系统  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  和  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$  补全为  $\mathbb{F}^n$  和  $\mathbb{F}^m$  中的正交基。

回想一下, 要将  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  补全为  $\mathbb{F}^n$  中的标准正交基, 只需找到  $\text{Ker } V^*$  的一组标准正交基  $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ ; 那么系统  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  将是  $\mathbb{F}^n$  中的一组标

准正交基。并且人们总是能通过格拉姆-施密特正交化从任意系统得到一组标准正交基。

然后  $A$  可以表示为

$$(3.4) \quad A = W\Sigma V^*,$$

其中  $V \in M_{n \times n}^{\mathbb{F}}$  和  $W \in M_{m \times m}^{\mathbb{F}}$  分别是以  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  和  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  为列的酉矩阵, 而  $\Sigma \in M_{m \times n}^{\mathbb{R}^+}$  是一个“对角”矩阵

$$(3.5) \quad \sigma_{j,k} = \begin{cases} \sigma_k, & j = k \leq r \\ 0, & \text{其他} \end{cases}.$$

也就是说, 为了得到矩阵  $\Sigma$ , 你需要取对角矩阵  $\text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}$  并通过在“南方”(下方)和“东方”(右侧)添加额外的零将其变成一个  $m \times n$  矩阵。

**定义 3.9** 对于矩阵  $A \in M_{m \times n}^{\mathbb{F}}$  (这里  $\mathbb{F}$  总是  $\mathbb{C}$  或  $\mathbb{R}$ ), 其**奇异值分解** (singular value decomposition, SVD) 是形如 (3.4) 的分解, 即分解  $A = W\Sigma V^*$ , 其中  $W \in M_{m \times m}^{\mathbb{F}}$  和  $V \in M_{n \times n}^{\mathbb{F}}$  是酉矩阵, 而  $\Sigma \in M_{m \times n}^{\mathbb{R}^+}$  是“对角”矩阵 (意思是  $\sigma_{k,k} \geq 0$  对于所有  $k = 1, 2, \dots, \min\{m, n\}$ , 并且  $\sigma_{j,k} = 0$  对于所有  $j \neq k$ )。

(3.3) 这种表示经常称作**约简 SVD** (reduced or compact SVD). 更精确地说, 约简 SVD 是一个表示  $A = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^*$ , 其中  $\widetilde{\Sigma} \in M_{r \times r}^{\mathbb{R}^+}$ ,  $r \leq \min\{m, n\}$  是一个对角矩阵, 其对角线元素严格为正, 而  $\widetilde{W} \in M_{m \times r}^{\mathbb{F}}$ ,  $\widetilde{V} \in M_{n \times r}^{\mathbb{F}}$  是等距同构; 而且, 我们要求  $\widetilde{W}$  和  $\widetilde{V}$  中至少有一个不是方阵。

**注记 3.10** 很容易看出, 如果  $A = W\Sigma V^*$  是  $A$  的奇异值分解, 那么  $\sigma_k := \sigma_{k,k}$  是  $A$  的奇异值, 即  $\sigma_k^2$  是  $A^*A$  的特征值。而且,  $V$  的列  $\mathbf{v}_k$  是  $A^*A$  的相应特征向量,  $A^*A\mathbf{v}_k = \sigma_k^2\mathbf{v}_k$ . 还要注意, 如果  $\sigma_k \neq 0$ , 那么  $\mathbf{w}_k = \frac{1}{\sigma_k}A\mathbf{v}_k$ .

所有这些都意味着任何奇异值分解  $A = W\Sigma V^*$  都可以通过本节上面描述的构造从施密特分解 (3.2) 得到。

对于不可逆矩阵  $A$ , 约简奇异值分解可以解释为施密特分解 (3.2) 的矩阵形式。对于可逆矩阵  $A$ , 施密特分解的矩阵形式给出了奇异值分解。

**注记 3.11** 奇异值分解  $A = W\Sigma V^*$  的另一种解释是,  $\Sigma$  是  $A$  在 (标准正交) 基  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  和  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$  下的矩阵, 即  $\Sigma = [A]_{\mathcal{B}, \mathcal{A}}$ .

我们将在后面使用这个解释。

### 3.4.1. 从奇异值分解到极分解

注意, 如果我们知道方阵  $A$  的奇异值分解  $A = W\Sigma V^*$ , 我们可以写出  $A$  的极分解:

$$(3.6) \quad A = W\Sigma V^* = (WV^*)(V\Sigma V^*) = U|A|$$

其中  $|A| = V\Sigma V^*$  并且  $U = WV^*$ .

为了说明这确实是一个极分解, 让我们注意到  $V\Sigma V^*$  是一个自伴随的、半正定的算子, 并且

$$\begin{aligned} A^*A &= (W\Sigma V^*)^*(W\Sigma V^*) = V\Sigma^*W^*W\Sigma V^* = V\Sigma^*\Sigma V^* \\ &= V(\Sigma^*\Sigma)V^* = (V\Sigma V^*)(V\Sigma V^*) = (|A|)^2. \end{aligned}$$

所以根据  $|A|$  的定义 (它是  $A^*A$  的唯一半正定平方根), 我们可以看出  $|A| = V\Sigma V^*$ . 变换  $WV^*$  显然是一个酉变换, 因为它是由两个酉变换相乘得到的, 所以 (3.6) 确实给出了  $A$  的一个极分解.

请注意, 此推理仅适用于方阵, 因为如果  $A$  不是方阵, 则乘积  $V\Sigma$  是未定义的 (维度不匹配, 你能看出为什么吗?).

### 练习

3.1. 证明矩阵  $A$  的非零奇异值的数量 (计入重数) 与其秩相等.

3.2. 为以下矩阵  $A$  找出施密特分解  $A = \sum_{k=1}^r s_k \mathbf{w}_k \mathbf{v}_k^*$ :

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

3.3. 设  $A$  是一个可逆矩阵, 设  $A = W\Sigma V^*$  是它的奇异值分解. 求  $A^*$  和  $A^{-1}$  的奇异值分解.

3.4. 为以下矩阵  $A$  找出奇异值分解  $A = W\Sigma V^*$ , 其中  $V$  和  $W$  是酉矩阵:

$$\begin{aligned} \text{a) } A &= \begin{pmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{pmatrix}; \\ \text{b) } A &= \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}. \end{aligned}$$

3.5. 找出矩阵

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$$

的奇异值分解. 并用它来找出:

- $\max_{\|\mathbf{x}\| \leq 1} \|A\mathbf{x}\|$  以及最大值对应的向量;
- $\min_{\|\mathbf{x}\| = 1} \|A\mathbf{x}\|$  以及最小值对应的向量;
- $A$  对  $\mathbb{R}^2$  中的闭单位球  $B = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq 1\}$  的像  $A(B)$ . 并在几何上描述  $A(B)$ .

3.6. 证明对于方阵  $A$ ,  $|\det A| = \det |A|$ .

3.7. 判断正误:

- 矩阵的奇异值也是该矩阵的特征值.
- 矩阵  $A$  的奇异值是  $A^*A$  的特征值.
- 如果  $s$  是矩阵  $A$  的一个奇异值, 而  $c$  是一个标量, 那么  $|c|s$  是  $cA$  的奇异值.
- 任何线性算子的奇异值都是非负的.
- 自伴随矩阵的奇异值与其特征值相等.

3.8. 设  $A$  是一个  $m \times n$  矩阵. 证明  $A^*A$  和  $AA^*$  的非零特征值 (计入重数) 是相同的. 你能说出  $A^*A$  的零特征值和  $AA^*$  的零特征值何时具有相同的重数吗?

3.9. 设  $s$  是算子  $A$  的最大奇异值, 设  $\lambda$  是  $A$  具有最大绝对值的特征值. 证明  $|\lambda| \leq s$ .

3.10. 证明矩阵的秩等于其非零奇异值的数量 (计入重数).



3.11. 证明算子范数  $\|A\|$  与弗罗贝尼乌斯范数  $\|A\|_2$  相等当且仅当该矩阵秩为 1。提示：上一个问题可能有所帮助。

3.12. 对于矩阵

$$A = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix},$$

描述单位球的逆像，即所有  $\mathbf{x} \in \mathbb{R}^2$  使得  $\|A\mathbf{x}\| \leq 1$  的集合。使用奇异值分解。

## 4. 奇异值分解的应用

正如我们上面讨论的，奇异值分解 (SVD) 本质上是对两组不同标准正交基的对角化。由于这里有两组不同的基，我们无法从其奇异值分解中得知一个算子的谱的性质。例如，奇异值分解 (3.5) 中的  $\Sigma$  对角线上的项并不是  $A$  的特征值。注意，对于  $A = W\Sigma V^*$  (如 (3.5) 所示)，通常有  $A^n \neq W\Sigma^n V^*$ ，因此这种对角化并不能帮助我们计算矩阵的函数。

然而，正如下面的例子所示，奇异值能够很好地揭示线性变换的所谓**度量性质** (metric property)。

最后说明：进行奇异值分解需要找到埃尔米特 (自伴) 矩阵  $A^*A$  的特征值和特征向量。为了找到特征值，我们通常需要计算特征多项式，找到它的根，等等。这看起来是一个相当复杂的过程，尤其考虑到对于五次及更高次的方程，并没有求根公式。

然而，存在非常有效的数值方法，可以计算出埃尔米特矩阵的特征值和特征向量，精确到任意给定的精度。这些方法不涉及计算特征多项式及其根。它们通过迭代过程直接计算近似的特征值和特征向量。由于埃尔米特矩阵具有标准正交的特征向量基，这些方法效果非常好。

我们在此不讨论这些方法，这超出了本书的范围。但是，你可以相信我，存在非常有效的数值方法来计算埃尔米特矩阵的特征值和特征向量，以及找到奇异值分解。这些方法非常有效，计算量也只会比求解线性方程组略多一些。

### 4.1. 单位球的像

例如，考虑以下问题：设  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  是一个线性变换，令  $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$  是  $\mathbb{R}^n$  中的闭单位球。我们希望描述  $A(B)$ ，即我们想弄清楚单位球在仿射变换下是如何被映射的。

让我们先考虑最简单的情况，即  $A$  是一个对角矩阵  $A = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ ，且  $\sigma_k > 0$  对  $k = 1, 2, \dots, n$  成立。那么对于  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  和  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T = A\mathbf{x}$ ，我们有  $y_k = \sigma_k x_k$  (等价地， $x_k = y_k / \sigma_k$ ) 对  $k = 1, 2, \dots, n$  成立。因

此,

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^T = A\mathbf{x} \text{ 其中 } \|\mathbf{x}\| \leq 1,$$

当且仅当坐标  $y_1, y_2, \dots, y_n$  满足不等式

$$\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} + \dots + \frac{y_n^2}{\sigma_n^2} = \sum_{k=1}^n \frac{y_k^2}{\sigma_k^2} \leq 1$$

(这只是不等式  $\|\mathbf{x}\|^2 = \sum_k |x_k|^2 \leq 1$ )。

满足上述不等式点的集合被称为**椭球体** (ellipsoid)。如果  $n = 2$ , 这是一个半轴长为  $\sigma_1$  和  $\sigma_2$  的椭圆面; 如果  $n = 3$ , 它是一个半轴长为  $\sigma_1, \sigma_2$  和  $\sigma_3$  的椭球体。在  $\mathbb{R}^n$  中, 这个集合的几何形状也容易可视化, 我们称之为半轴长为  $\sigma_1, \sigma_2, \dots, \sigma_n$  的椭球体。向量  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  或更确切地说, 它们对应的直线, 被称为椭球体的**主轴** (principal axes)。

奇异值分解本质上说明了, 在任意内积空间中, 任何算子都可以通过一对标准正交基变得对角化 (见注记 3.11)。即, 考虑奇异值分解 (3.1) 中的正交基  $\mathcal{A} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  和  $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ . 那么  $A$  在这些基下的矩阵是

$$[A]_{\mathcal{B}, \mathcal{A}} = \text{diag}\{\sigma_n : n = 1, 2, \dots, n\}.$$

假设所有  $\sigma_k > 0$ , 并重复上述推理, 很容易证明任何点  $\mathbf{y} = A\mathbf{x} \in A(B)$  当且仅当它满足不等式:

$$\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} + \dots + \frac{y_n^2}{\sigma_n^2} = \sum_{k=1}^n \frac{y_k^2}{\sigma_k^2} \leq 1.$$

其中  $y_1, y_2, \dots, y_n$  是  $\mathbf{y}$  在标准正交基  $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$  下的坐标。类似地,  $(x_1, x_2, \dots, x_n)^T = [\mathbf{x}]_{\mathcal{A}}$ .

但这本质上是同一个椭球体, 只是被“旋转”了 (具有不同但仍正交的主轴)!

还有一个替代的解释, 呈现在下面。

考虑“对角”矩阵  $\Sigma$  的一般情况, 形式如 (3.5)。很容易看出, 单位球  $B$  的像  $\Sigma B$  是一个椭球体 (不是在整个空间中, 而是在  $\text{Ran } \Sigma$  中), 其半轴长为  $\sigma_1, \sigma_2, \dots, \sigma_r$ .

现在考虑一般情况,  $A = W\Sigma V^*$ , 其中  $W, V$  是酉算子。酉变换不改变单位球 (因为它们保持范数), 所以  $V^*(B) = B$ . 我们知道  $\Sigma(B)$  是  $\text{Ran } \Sigma$  中的一个椭球体, 半轴长为  $\sigma_1, \sigma_2, \dots, \sigma_r$ . 酉变换不改变物体的几何形状, 所以  $W(\Sigma(B))$  也是一个椭球体, 且具有相同的半轴长。从分解  $A = W\Sigma V^*$  (利用  $W$  和  $V^*$  都是可逆的事实) 可以很容易看出,  $W$  将  $\text{Ran } \Sigma$  映射到  $\text{Ran } A$ , 因此我们可以得出结论:

闭单位球  $B$  的像  $A(B)$  是  $\text{Ran } A$  中的一个椭球体, 其半轴长为  $\sigma_1, \sigma_2, \dots, \sigma_r$ . 这里  $r$  是非零奇异值的数量, 即  $A$  的秩。

## 4.2. 线性变换的算子范数

给定一个线性变换  $A : X \rightarrow Y$ ，我们考虑以下优化问题：在闭单位球  $B = \{\mathbf{x} \in X : \|\mathbf{x}\| \leq 1\}$  上，求  $\|A\mathbf{x}\|$  的最大值。

再次，奇异值分解允许我们解决这个问题。对于具有非负项的对角矩阵  $A$ ，最大值正好是最大的对角项。确实，设  $s_1, s_2, \dots, s_r$  是  $A$  的非零对角项，设  $s_1$  是它们中最大的。由于对于  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ ,

$$(4.1) \quad A\mathbf{x} = \sum_{k=1}^r s_k x_k \mathbf{e}_k,$$

我们可以得出

$$\|A\mathbf{x}\|^2 = \sum_{k=1}^r s_k^2 |x_k|^2 \leq s_1^2 \sum_{k=1}^r |x_k|^2 = s_1^2 \cdot \|\mathbf{x}\|^2,$$

因此  $\|A\mathbf{x}\| \leq s_1 \|\mathbf{x}\|$ 。另一方面， $\|A\mathbf{e}_1\| = \|s_1 \mathbf{e}_1\| = s_1 \|\mathbf{e}_1\|$ ，因此  $s_1$  确实是闭单位球  $B$  上  $\|A\mathbf{x}\|$  的最大值。注意，在上述推理中，我们没有假设矩阵  $A$  是方阵；我们只假设“主对角线”外的所有元素都为 0，因此公式 (4.1) 成立。

为了处理一般情况，我们考虑奇异值分解 (3.5)， $A = W\Sigma V^*$ ，其中  $W, V$  是酉算子， $\Sigma$  是具有非负项的对角矩阵。由于酉变换不改变范数，我们可以得出

$\|A\mathbf{x}\|$  在单位球  $B$  上的最大值是  $\Sigma$  的最大对角项，即  $A$  的最大奇异值。

**定义** 量  $\max\{\|A\mathbf{x}\| : \mathbf{x} \in X, \|\mathbf{x}\| \leq 1\}$  被称为  $A$  的**算子范数**(operator norm)，记作  $\|A\|$ 。

很容易看出  $\|A\|$  满足范数的所有性质：

1.  $\|\alpha A\| = |\alpha| \cdot \|A\|$ ;
2.  $\|A + B\| \leq \|A\| + \|B\|$ ;
3. 对所有  $A$  都有  $\|A\| \geq 0$ ;
4.  $\|A\| = 0$  当且仅当  $A = \mathbf{0}$ ,

因此它确实是  $X$  到  $Y$  的线性变换空间上的一个范数。

算子范数的一个主要性质是不等式

$$\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|,$$

这很容易从范数  $\|\mathbf{x}\|$  的齐次性推导出来。

事实上，可以证明算子范数  $\|A\|$  是满足

$$\|A\mathbf{x}\| \leq C \|\mathbf{x}\| \quad \forall \mathbf{x} \in X$$

的最佳（最小）非负数  $C$ 。这通常被用作算子范数的定义。

在线性变换空间上, 我们之前已经由内积的概念定义了一个范数, 即弗罗贝尼乌斯 (Frobenius) 范数, 或称希尔伯特-施密特 (Hilbert-Schmidt) 范数  $\|A\|_2$ :

$$\|A\|_2^2 = \text{trace}(A^*A).$$

所以, 让我们来研究这两个范数是如何比较的。

设  $s_1, s_2, \dots, s_r$  是  $A$  的非零奇异值 (计重数), 设  $s_1$  是它们中最大的。那么  $s_1^2, s_2^2, \dots, s_r^2$  是  $A^*A$  的非零特征值 (同样计重数)。回想一下, 迹等于特征值之和, 我们得出

$$\|A\|_2^2 = \text{trace}(A^*A) = \sum_{k=1}^r s_k^2.$$

另一方面, 我们知道算子范数  $\|A\|$  等于其最大奇异值, 即  $\|A\| = s_1$ . 因此, 我们可以得出  $\|A\| \leq \|A\|_2$ , 即

矩阵的算子范数不能大于其弗罗贝尼乌斯范数。

这个陈述也可以用柯西-施瓦茨不等式直接证明, 并且这种证明在一些教材中已经给出。我们这里展示的证明的美妙之处在于, 它不需要任何计算, 并阐明了不等式背后的原因。

### 4.3. 矩阵的条件数

假设我们有一个可逆矩阵  $A$ , 并且我们想解方程  $A\mathbf{x} = \mathbf{b}$ . 解当然是  $\mathbf{x} = A^{-1}\mathbf{b}$ , 但我们想研究如果我们只知道近似数据时会发生什么。

在现实生活中, 数据是通过某些实验获得的。但即使我们有精确的数据, 计算机计算过程中的舍入误差也可能产生类似效应, 将数据扭曲。

让我们考虑最简单的模型, 假设方程的右侧有一个小的误差。这意味着, 我们求解的是

$$A\mathbf{x} = \mathbf{b} + \Delta\mathbf{b},$$

而不是  $A\mathbf{x} = \mathbf{b}$ . 其中  $\Delta\mathbf{b}$  是右侧  $\mathbf{b}$  的一个小扰动。因此, 我们得到近似解  $\mathbf{x} + \Delta\mathbf{x}$ , 而  $A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ . 我们假设  $A$  是可逆的, 所以  $\Delta\mathbf{x} = A^{-1}\Delta\mathbf{b}$ .

我们想知道解中的相对误差  $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$  与右侧的相对误差  $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$  相比有多大。很容易看出:

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A^{-1}\Delta\mathbf{b}\|}{\|\mathbf{x}\|} = \frac{\|A^{-1}\Delta\mathbf{b}\| \|\mathbf{b}\|}{\|\mathbf{b}\| \|\mathbf{x}\|} = \frac{\|A^{-1}\Delta\mathbf{b}\| \|A\mathbf{x}\|}{\|\mathbf{b}\| \|\mathbf{x}\|}.$$

由于  $\|A^{-1}\Delta\mathbf{b}\| \leq \|A^{-1}\| \cdot \|\Delta\mathbf{b}\|$  且  $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$ , 我们可以得出:

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

$\|A\| \cdot \|A^{-1}\|$  这个量被称为矩阵的**条件数** (condition number)。它估计了解的相对误差  $\mathbf{x}$  在多大程度上取决于右侧  $\mathbf{b}$  的相对误差。

让我们看看这个量与奇异值是如何关联的。设  $s_1, s_2, \dots, s_n$  是  $A$  的奇异值, 并且假设  $s_1$  是最大奇异值,  $s_n$  是最小奇异值。我们知道算子范数等于其最大奇异值, 所以

$$\|A\| = s_1, \quad \|A^{-1}\| = \frac{1}{s_n},$$

因此

$$\|A\| \cdot \|A^{-1}\| = \frac{s_1}{s_n}.$$

换句话说,

矩阵的条件数等于最大和最小奇异值之比。

我们上面推导出  $\|\Delta \mathbf{x}\|/\|\mathbf{x}\| \leq \|A^{-1}\| \cdot \|A\| \cdot \|\Delta \mathbf{b}\|/\|\mathbf{b}\|$ . 不难看出, 这个估计是精确的, 即可以选择右侧  $\mathbf{b}$  和误差  $\Delta \mathbf{b}$ , 使得等式成立:

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} = \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}.$$

我们只需取  $\mathbf{b} = \mathbf{w}_1$  且  $\Delta \mathbf{b} = \alpha \mathbf{w}_n$ , 其中  $\mathbf{w}_1$  和  $\mathbf{w}_n$  分别是奇异值分解  $A = W\Sigma V^*$  中  $W$  的第一列和最后一列,  $\alpha \neq 0$  是任意标量。这里, 像往常一样, 奇异值假设为非递增排序  $s_1 \geq s_2 \geq \dots \geq s_n$ , 所以  $s_1$  是最大的, 而  $s_n$  是最小的。

我们将细节留给读者作为练习。

如果一个矩阵的条件数不是太大, 则称该矩阵是**良态的** (well conditioned)。如果条件数很大, 则称该矩阵是**病态的** (ill conditioned)。这里的“大”取决于具体问题: 你能在多大程度上确定你的右侧数据, 对解需要多高的精度等等。

#### 4.4. 矩阵的有效秩

理论上, 矩阵的秩很容易计算: 只需对矩阵进行行变换并计算主元即可。然而, 在实际应用中, 情况并非如此简单。主要原因是, 我们通常不知道精确的矩阵, 只知道其近似值, 精度有限。

此外, 即使我们知道精确的矩阵, 大多数计算机程序在计算过程中也会引入舍入误差, 因此我们实际上无法区分一个零主元和一个非常小的主元。

一种简单粗暴的操作方法是这样的: 在计算秩 (以及与之相关的其他对象, 如列空间、核等) 时, 只需设置一个容差 (一个小的数), 如果主元小于容差, 就将其视为零。这种方法的优点在于它的简单性, 因为它非常容易编程。然而, 主要的缺点是无法看出容差的作用。例如, 如果我们设置容差为  $10^{-6}$ , 我们失去了什么?  $10^{-8}$  会好多少? 虽然上述方法对良态矩阵效果很好, 但在一般情况下并不可靠。

一个更好的方法是使用奇异值。它需要更多的计算，但能给出更好、更易于解释的结果。在这种方法中，我们也设定一个小的数作为容差，然后进行奇异值分解。之后，我们简单地将小于容差的奇异值视为零。这种方法的优点在于我们可以清楚地看到我们在做什么。奇异值是椭球体  $A(B)$  ( $B$  是闭单位球) 的半轴长，因此通过设置容差，我们只是决定了椭球体应该有多“细”才被认为是“扁平”的。

#### 4.5. 摩尔-彭罗斯 (伪) 逆

正如我们在第 5 章第 4 节中所讨论的，在方程  $A\mathbf{x} = \mathbf{b}$  没有解的情况下，最小二乘解给了我们“次优”的解决方案 (当方程有解时，它也给出了  $A\mathbf{x} = \mathbf{b}$  的解)。

注意，最小二乘解并未解决唯一性问题：方程  $A^*A\mathbf{x} = A^*\mathbf{b}$  的解不一定唯一。一个自然区别开来的解是具有最小范数的解；这样的解确实是唯一的，并且可以通过取任意一个解，然后将其投影到  $(\text{Ker } A^*A)^\perp = (\text{Ker } A)^\perp$  上得到 (参见第 5 章的问题 4.5 和 4.6)。

不难看出，如果  $A = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^*$  是  $A$  的约简奇异值分解，那么最小范数的最小二乘解  $\mathbf{x}_0$  由下式给出：

$$(4.2) \quad \mathbf{x}_0 = \widetilde{V}\widetilde{\Sigma}^{-1}\widetilde{W}^*\mathbf{b}.$$

确实， $\mathbf{x}_0$  是  $A\mathbf{x} = \mathbf{b}$  的一个最小二乘解 (即  $A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}$  的解)：

$$A\mathbf{x}_0 = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^*(\widetilde{V}\widetilde{\Sigma}^{-1}\widetilde{W}^*\mathbf{b}) = \widetilde{W}\widetilde{\Sigma}\widetilde{\Sigma}^{-1}\widetilde{W}^*\mathbf{b} = \widetilde{W}\widetilde{W}^*\mathbf{b} = P_{\text{Ran } A}\mathbf{b};$$

在链的最后一个等式中，我们使用了  $\widetilde{W}\widetilde{W}^* = P_{\text{Ran } \widetilde{W}}$  ( $P_{\text{Ran } \widetilde{W}} = \widetilde{W}(\widetilde{W}^*\widetilde{W})^{-1}\widetilde{W}^* = \widetilde{W}\widetilde{W}^*$ ) 并且  $\text{Ran } \widetilde{W} = \text{Ran } A$  (见问题 4.4)。

$A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}$  的一般解由

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{y}, \quad \mathbf{y} \in \text{Ker } A$$

给出，因此  $\mathbf{x}_0$  确实是  $A\mathbf{x} = P_{\text{Ran } A}\mathbf{b}$  的唯一最小范数解，或者等价地，是  $A\mathbf{x} = \mathbf{b}$  的最小范数的最小二乘解。

**定义 4.1.** 算子  $A^+ := \widetilde{V}\widetilde{\Sigma}^{-1}\widetilde{W}^*$ ，其中  $A = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^*$  是  $A$  的约简奇异值分解，称为  $A$  的摩尔-彭罗斯逆 (或摩尔-彭罗斯伪逆) (Moore-Penrose (pseudo)inverse)。换句话说，摩尔-彭罗斯逆是给出  $A\mathbf{x} = \mathbf{b}$  的唯一最小二乘解的算子。

**注记 4.2.** 文献中通常将摩尔-彭罗斯逆定义为一个矩阵  $A^+$ ，它满足：

1.  $AA^+A = A$ ;
2.  $A^+AA^+ = A^+$ ;
3.  $(AA^+)^* = AA^+$ ;
4.  $(A^+A)^* = A^+A$ .

很容易验证算子  $A^+ := \widetilde{V}\widetilde{\Sigma}^{-1}\widetilde{W}^*$  满足上述性质 1-4。

还可以 (尽管有点难) 证明满足性质 1-4 的算子  $A^+$  是唯一的。确实, 通过用  $A^+$  左乘或右乘等式 1, 我们得到  $(A^+A)^2 = A^+A$  和  $(AA^+)^2 = AA^+$ ; 与性质 3 和 4 一起, 可以得出  $A^+A$  和  $AA^+$  是正交投影 (见第 5 章问题 5.6)。

显然,  $\text{Ker } A \subset \text{Ker } A^+A$ . 另一方面, 等式 1 暗示  $\text{Ker } A^+A \subset \text{Ker } A$  (为什么?), 所以  $\text{Ker } A^+A = \text{Ker } A$ . 但这表明  $A^+A$  是  $(\text{Ker } A)^\perp = \text{Ran } A^*$  上的正交投影,

$$A^+A = P_{\text{Ran } A^*}.$$

性质 1 也暗示了  $AA^+y = y$  对所有  $y \in \text{Ran } A$  成立。由于  $AA^+$  是一个正交投影, 我们得出  $\text{Ran } A \subset \text{Ran } AA^+$ . 相反的包含关系  $\text{Ran } AA^+ \subset \text{Ran } A$  是平凡的, 所以  $AA^+$  是  $\text{Ran } A$  上的正交投影,

$$AA^+ = P_{\text{Ran } A}.$$

知道了  $A^+A$  和  $AA^+$ , 我们可以将性质 2 重写为

$$P_{\text{Ran } A^*}A^+ = A^+ \quad \text{或} \quad A^+P_{\text{Ran } A} = A^+.$$

结合上述恒等式, 我们得到

$$P_{\text{Ran } A^*}A^+P_{\text{Ran } A} = A^+.$$

最后, 对于  $A$  的目标空间中的任何  $\mathbf{b}$ , 令

$$\mathbf{x}_0 := A^+\mathbf{b} = P_{\text{Ran } A^*}A^+\mathbf{b} \in \text{Ran } A^*$$

并且

$$A\mathbf{x}_0 = AA^+\mathbf{b} = P_{\text{Ran } A}\mathbf{b},$$

则  $\mathbf{x}_0$  是  $A\mathbf{x} = \mathbf{b}$  的一个最小二乘解。由于  $\mathbf{x}_0 \in \text{Ran } A^* = (\text{Ker } A)^\perp$ ,  $\mathbf{x}_0$  如前所述, 是最小范数的最小二乘解。但是, 如我们之前所示, 这样的最小范数解由 (4.2) 给出, 所以  $A^+ = \widetilde{V}\widetilde{\Sigma}^{-1}\widetilde{W}^*$ 。

### 练习

4.1. 求以下矩阵的范数和条件数:

$$\begin{aligned} \text{a) } A &= \begin{pmatrix} 4 & 0 \\ 1 & 3 \end{pmatrix}. \\ \text{b) } A &= \begin{pmatrix} 5 & 3 \\ -3 & 3 \end{pmatrix}. \end{aligned}$$

对于 a) 部分的矩阵  $A$ , 给出一个右侧  $\mathbf{b}$  和误差  $\Delta\mathbf{b}$  的例子, 使得

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|};$$

这里  $A\mathbf{x} = \mathbf{b}$  且  $A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ .

4.2. 设  $A$  是一个正常算子, 其特征值为  $\lambda_1, \lambda_2, \dots, \lambda_n$  (计重数)。证明  $A$  的奇异值是  $|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|$ 。

## 4.3. 求矩阵

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

的奇异值、范数和条件数。

你基本上可以不经计算完成此题，如果你能回答以下问题：

- 某个子空间  $E$  上的正交投影  $P_E$  的奇异值是多少？
- 跨越向量  $(1, 1, 1)^T$  的子空间的零空间的矩阵是什么？
- 算子  $T$  和  $aT + bI$ （其中  $a$  和  $b$  是标量）的特征值之间有什么关系？

当然，你也可以直接进行计算。

4.4. 设  $A = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^*$  是  $A$  的约简奇异值分解。证明  $\text{Ran } A = \text{Ran } \widetilde{W}$ ，然后通过取伴随矩阵证明  $\text{Ran } A^* = \text{Ran } \widetilde{V}$ 。

4.5. 用奇异值分解  $A = W\Sigma V^*$  表示摩尔-彭罗斯逆  $A^+$  的公式。

4.6. (提霍诺夫正则化, Tychonov's regularization): 证明摩尔-彭罗斯逆  $A^+$  可以计算为极限:

$$A^+ = \lim_{\varepsilon \rightarrow 0^+} (A^*A + \varepsilon I)^{-1}A^* = \lim_{\varepsilon \rightarrow 0^+} A^*(AA^* + \varepsilon I)^{-1}.$$

## 5. 正交矩阵的结构

一个行列式为 1 的正交矩阵  $U$  通常被称为**旋转** (rotation)。下面的定理解释了这个名称。

**定理 5.1.** 设  $U$  是  $\mathbb{R}^n$  中的一个正交算子，且  $\det U = 1$ .<sup>42</sup> 则存在一组标准正交基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ ，使得  $U$  在该基下的矩阵具有分块对角形式：

$$\begin{pmatrix} R_{\phi_1} & & & & \\ & R_{\phi_2} & & & \\ & & \ddots & & \\ & & & R_{\phi_k} & \\ & & & & I_{n-2k} \end{pmatrix},$$

其中  $R_{\phi_k}$  是 2 维旋转矩阵，

$$R_{\phi_k} = \begin{pmatrix} \cos \phi_k & -\sin \phi_k \\ \sin \phi_k & \cos \phi_k \end{pmatrix},$$

而  $I_{n-2k}$  表示  $(n-2k) \times (n-2k)$  的单位矩阵。

**证明：** 我们知道，如果  $p$  是一个实系数多项式，并且  $\lambda$  是它的复根， $p(\lambda) = 0$ ，那么  $\bar{\lambda}$  也是  $p$  的根， $p(\bar{\lambda}) = 0$ （这可以通过将  $\bar{\lambda}$  代入  $p(z) = \sum_{k=0}^n a_k z^k$  来很容易地验证）。

因此，实矩阵  $A$  的所有复特征值可以配对成  $\lambda_k, \bar{\lambda}_k$ 。

<sup>42</sup> 对于一个正交矩阵  $U$ ，它的行列式为  $\pm 1$ 。



我们知道,酉矩阵的特征值绝对值都为 1,所以  $A$  的所有复特征值都可以写成  $\lambda_k = \cos \alpha_k + i \sin \alpha_k$  和  $\bar{\lambda}_k = \cos \alpha_k - i \sin \alpha_k$ .

固定一对复特征值  $\lambda$  和  $\bar{\lambda}$ , 设  $\mathbf{u} \in \mathbb{C}^n$  是  $U$  的特征向量,  $U\mathbf{u} = \lambda\mathbf{u}$ . 那么  $U\bar{\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$ . 现在, 将  $\mathbf{u}$  分解为实部和虚部, 即定义

$$\mathbf{x} := \operatorname{Re} \mathbf{u} = \frac{\mathbf{u} + \bar{\mathbf{u}}}{2}, \quad \mathbf{y} := \operatorname{Im} \mathbf{u} = \frac{\mathbf{u} - \bar{\mathbf{u}}}{2i}$$

(注意,  $\mathbf{x}, \mathbf{y}$  是实向量, 即所有项均为实数的向量)。那么  $\mathbf{u} = \mathbf{x} + i\mathbf{y}$ . 那么

$$U\mathbf{x} = U\frac{\mathbf{u} + \bar{\mathbf{u}}}{2} = \frac{1}{2}(U\mathbf{u} + U\bar{\mathbf{u}}) = \frac{1}{2}(\lambda\mathbf{u} + \bar{\lambda}\bar{\mathbf{u}}) = \operatorname{Re}(\lambda\mathbf{u}).$$

类似地,

$$U\mathbf{y} = U\frac{\mathbf{u} - \bar{\mathbf{u}}}{2i} = \frac{1}{2i}(U\mathbf{u} - U\bar{\mathbf{u}}) = \frac{1}{2i}(\lambda\mathbf{u} - \bar{\lambda}\bar{\mathbf{u}}) = \operatorname{Im}(\lambda\mathbf{u}).$$

由于  $\lambda = \cos \alpha + i \sin \alpha$ , 我们有

$$\lambda\mathbf{u} = (\cos \alpha + i \sin \alpha)(\mathbf{x} + i\mathbf{y}) = ((\cos \alpha)\mathbf{x} - (\sin \alpha)\mathbf{y}) + i((\cos \alpha)\mathbf{y} + (\sin \alpha)\mathbf{x}).$$

所以

$$U\mathbf{x} = \operatorname{Re}(\lambda\mathbf{u}) = (\cos \alpha)\mathbf{x} - (\sin \alpha)\mathbf{y}, \quad U\mathbf{y} = \operatorname{Im}(\lambda\mathbf{u}) = (\cos \alpha)\mathbf{y} + (\sin \alpha)\mathbf{x}.$$

换句话说,  $U$  对由向量  $\mathbf{x}, \mathbf{y}$  生成的二维子空间  $E_\lambda$  保持不变, 即  $E_\lambda$  是  $U$  的不变子空间, 且  $U$  在该子空间上的限制矩阵是旋转矩阵

$$R_{-\alpha} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}.$$

注意, 向量  $\mathbf{u}$  和  $\bar{\mathbf{u}}$  (酉矩阵对应于不同特征值的特征向量) 是正交的, 所以根据勾股定理

$$\|\mathbf{x}\| = \|\mathbf{y}\| = \frac{1}{\sqrt{2}}\|\mathbf{u}\|,$$

很容易核验  $\mathbf{x} \perp \mathbf{y}$ , 所以  $\mathbf{x}, \mathbf{y}$  是  $E_\lambda$  中的一组正交基。如果我们对每个向量  $\mathbf{x}, \mathbf{y}$  乘以相同的非零数, 我们不会改变线性变换的矩阵, 所以不妨假设  $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ , 即  $\mathbf{x}, \mathbf{y}$  是  $E_\lambda$  中的一组标准正交基。

让我们将标准正交向量组  $\mathbf{v}_1 = \mathbf{x}, \mathbf{v}_2 = \mathbf{y}$  补全成  $\mathbb{R}^n$  中的一组标准正交基。由于  $UE_\lambda \subset E_\lambda$ , 即  $E_\lambda$  是  $U$  的不变子空间, 在该基下的  $U$  的矩阵具有分块三角形形式:

$$\left( \begin{array}{c|c} R_{-\alpha} & * \\ \hline \mathbf{0} & U_1 \end{array} \right),$$

其中  $\mathbf{0}$  表示一个  $(n-2) \times 2$  的零块。

由于旋转矩阵  $R_{-\alpha}$  是可逆的, 我们有  $UE_\lambda = E_\lambda$ . 因此

$$U^*E_\lambda = U^{-1}E_\lambda = E_\lambda,$$

所以我们构造的基下的  $U$  的矩阵实际上是分块对角形式:

$$\left( \begin{array}{c|c} R_{-\alpha} & \mathbf{0} \\ \hline \mathbf{0} & U_1 \end{array} \right).$$

由于  $U$  是酉的,

$$I = U^*U = \left( \begin{array}{c|c} I & \mathbf{0} \\ \hline \mathbf{0} & U_1^*U_1 \end{array} \right),$$

所以, 由于  $U_1$  是方阵, 它也是酉的。

如果  $U_1$  有复特征值, 我们可以应用相同的过程将其大小减 2, 直到我们剩下只具有实特征值的块。实特征值只能是  $+1$  或  $-1$ , 所以在一组标准正交基下,  $U$  的矩阵具有以下形式:

$$\begin{pmatrix} R_{-\alpha_1} & & & & & \\ & R_{-\alpha_2} & & & & \\ & & \ddots & & & \\ & & & R_{-\alpha_d} & & \\ & & & & -I_r & \\ & & & & & I_l \end{pmatrix};$$

这里  $I_r$  和  $I_l$  分别是  $r \times r$  和  $l \times l$  的单位矩阵。由于  $\det U = 1$ , 特征值为  $-1$  的重数 (即  $r$ ) 必须是偶数。

注意,  $2 \times 2$  矩阵  $-I_2$  可以解释为通过角度  $\pi$  的旋转。因此, 这一矩阵具有上述定理结论中的形式, 其中  $\phi_k = -\alpha_k$  或  $\phi_k = \pi$ .  $\square$

让我们给出定理 5.1 的另一个解释。定义  $T_j$  为在由向量  $\mathbf{v}_j, \mathbf{v}_{j+1}$  张成的平面中的一次  $\phi_j$  旋转。那么定理 5.1 就可以简单地说成  $U$  是旋转  $T_j, j = 1, 2, \dots, k$  的复合。注意, 由于旋转  $T_j$  在相互正交的平面上作用, 它们是可交换的, 也就是说, 复合的顺序并不重要。因此, 该定理可以解释为:

任何  $\mathbb{R}^n$  中的旋转都可以表示为最多  $n/2$  个可交换的平面中旋转的复合。

如果一个正交矩阵的行列式为  $-1$ , 其结构由以下定理描述。

**定理 5.2.** 设  $U$  是  $\mathbb{R}^n$  中的一个正交算子, 且  $\det U = -1$ . 则存在一组标准正交基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , 使得  $U$  在该基下的矩阵具有分块对角形式:

$$\begin{pmatrix} R_{\phi_1} & & & & \\ & R_{\phi_2} & & & \\ & & \ddots & & \\ & & & R_{\phi_k} & \\ & & & & I_r \\ & & & & & -1 \end{pmatrix},$$

其中  $r = n - 2k - 1$ , 并且  $R_{\phi_k}$  是 2 维旋转矩阵,

$$R_{\phi_k} = \begin{pmatrix} \cos \phi_k & -\sin \phi_k \\ \sin \phi_k & \cos \phi_k \end{pmatrix},$$

而  $I_{n-2k}$  表示  $(n - 2k) \times (n - 2k)$  的单位矩阵。

我们把证明留给读者作为练习。对定理 5.1 证明的修改是很明显的。

注意, 从上述定理可以得出, 一个行列式为  $-1$  的  $2 \times 2$  正交矩阵总是反射。

现在让我们固定一组标准正交基, 例如  $\mathbb{R}^n$  中的标准基。我们称一个**初等旋转** (elementary rotation)<sup>43</sup> 是在  $x_j - x_k$  平面中的一次旋转, 即一个只改变坐标  $x_j$  和  $x_k$  的线性变换, 并且它在这两个坐标上像平面旋转一样作用。

**定理 5.3.** 任何旋转  $U$  (即  $\det U = 1$  的正交变换) 都可以表示为最多  $n(n-1)/2$  个初等旋转的乘积。

为了证明这个定理, 我们需要以下简单的引理。

**引理 5.4.** 设  $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$ . 存在一个  $\mathbb{R}^2$  的旋转  $R_\alpha$ , 它将向量  $\mathbf{x}$  移动到向量  $(a, 0)^T$ , 其中  $a = \sqrt{x_1^2 + x_2^2}$ .

证明是基本的, 我们将其作为读者练习。你可以画一张图或者写出  $R_\alpha$  的公式。

**引理 5.5.** 设  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ . 存在  $n-1$  个初等旋转  $R_1, R_2, \dots, R_{n-1}$ , 使得  $R_{n-1} \dots R_2 R_1 \mathbf{x} = (a, 0, 0, \dots, 0)^T$ , 其中  $a = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ .

**证明:** 引理证明的思路非常简单。我们使用一个初等旋转  $R_1$  (在  $x_{n-1} - x_n$  平面中) 来“消去” $\mathbf{x}$  的最后一个坐标 (引理 5.4 保证了这样的旋转存在)。然后使用一个初等旋转  $R_2$  (在  $x_{n-2} - x_{n-1}$  平面中) 来“消去” $R_1 \mathbf{x}$  的第  $n-1$  个坐标 (旋转  $R_2$  不改变最后一个坐标, 所以  $R_2 R_1 \mathbf{x}$  的最后一个坐标保持为零), 以此类推。

为了进行正式证明, 我们将使用数学归纳法。 $n=1$  的情况是平凡的, 因为  $\mathbb{R}^1$  中的任何向量都具有所需的形状。 $n=2$  的情况由引理 5.4 处理。

<sup>43</sup> 这个术语并没有被广泛接受。

现在假设引理对  $n-1$  成立, 我们将证明对  $n$  成立. 根据引理 5.4, 存在一个  $2 \times 2$  旋转矩阵  $R_\alpha$ , 使得

$$R_\alpha \begin{pmatrix} x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} a_{n-1} \\ 0 \end{pmatrix},$$

其中  $a_{n-1} = \sqrt{x_{n-1}^2 + x_n^2}$ . 那么如果我们定义  $n \times n$  的初等旋转  $R_1$  为:

$$R_1 = \begin{pmatrix} I_{n-2} & \mathbf{0} \\ \mathbf{0} & R_\alpha \end{pmatrix}$$

( $I_{n-2}$  是  $(n-2) \times (n-2)$  的单位矩阵), 那么

$$R_1 \mathbf{x} = (x_1, x_2, \dots, x_{n-2}, a_{n-1}, 0)^T.$$

我们假设引理的结论对  $n-1$  成立, 因此存在  $n-2$  个初等旋转 (我们称它们为  $R_2, R_3, \dots, R_{n-1}$ ), 它们在  $\mathbb{R}^{n-1}$  中 (仅作用于坐标  $x_1, x_2, \dots, x_{n-1}$ ) 将向量  $(x_1, x_2, \dots, x_{n-1}, a_{n-1})^T \in \mathbb{R}^{n-1}$  变换为向量  $(a, 0, \dots, 0)^T \in \mathbb{R}^{n-1}$ . 换句话说,

$$R_{n-1} \dots R_3 R_2 (x_1, x_2, \dots, x_{n-1}, a_{n-1})^T = (a, 0, \dots, 0)^T.$$

我们可以总假设初等旋转  $R_2, R_3, \dots, R_{n-1}$  在  $\mathbb{R}^n$  中作用, 只需假设它们不改变最后一个坐标. 那么

$$R_{n-1} \dots R_3 R_2 R_1 \mathbf{x} = (a, 0, \dots, 0)^T \in \mathbb{R}^n.$$

现在我们来证明  $a = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ . 这可以通过直接计算轻松验证, 但我们现在采用间接推理. 我们知道正交变换保持范数, 并且  $a \geq 0$ . 那么我们就没有任何选择, 唯一的可能性就是  $a = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ .  $\square$

**引理 5.6.** 设  $A$  是一个具有实数项的  $n \times n$  矩阵. 存在初等旋转  $R_1, R_2, \dots, R_N$ ,  $N \leq n(n-1)/2$ , 使得矩阵  $B = R_N \dots R_2 R_1 A$  是上三角的, 并且其所有对角线元素, 除了最后一个  $B_{n,n}$  之外, 都是非负的.

**证明:** 我们将使用数学归纳法.  $n=1$  的情况是平凡的, 因为我们可以说任何  $1 \times 1$  矩阵都具有所需的形状.

让我们考虑  $n=2$  的情况. 设  $\mathbf{a}_1$  是  $A$  的第一列. 根据引理 5.4, 存在一个旋转  $R$ , 它可以“消去” $\mathbf{a}_1$  的第二个坐标, 使得第一个坐标非负. 然后矩阵  $B = RA$  具有所需的形状.

现在假设引理对  $(n-1) \times (n-1)$  矩阵成立, 我们想对  $n \times n$  矩阵证明它. 对于  $n \times n$  矩阵  $A$ , 设  $\mathbf{a}_1$  是它的第一列. 根据引理 5.5, 我们可以找到  $n-1$  个初等旋转

(例如  $R_1, R_2, \dots, R_{n-1}$ ), 它们将  $\mathbf{a}_1$  变换为  $(a, 0, \dots, 0)^T$ . 那么矩阵  $R_{n-1} \dots R_2 R_1 A$  具有以下分块三角形式:

$$R_{n-1} \dots R_2 R_1 A = \begin{pmatrix} a & * \\ \mathbf{0} & A_1 \end{pmatrix},$$

其中  $A_1$  是一个  $(n-1) \times (n-1)$  的块。

我们假设引理对  $n-1$  成立, 所以  $A_1$  可以通过最多  $(n-1)(n-2)/2$  个旋转变换成所需的上三角形式。注意, 这些旋转作用在  $\mathbb{R}^{n-1}$  中(只作用于坐标  $x_2, x_3, \dots, x_n$ ), 但我们可以总假设它们作用在整个  $\mathbb{R}^n$  中, 只需假设它们不改变第一个坐标。那么, 这些旋转不会改变  $R_{n-1} \dots R_2 R_1 A$  的第一列向量  $(a, 0, \dots, 0)^T$ . 因此, 矩阵  $A$  可以通过最多  $n-1 + (n-1)(n-2)/2 = n(n-1)/2$  个初等旋转变换成所需的上三角形式。  $\square$

**定理 5.3 的证明** 根据引理 5.5, 存在初等旋转  $R_1, R_2, \dots, R_N$ , 使得矩阵  $U_1 = R_N \dots R_2 R_1 U$  是上三角的, 并且除最后一个对角线元素  $B_{n,n}$  外, 所有对角线元素都是非负的。

注意, 矩阵  $U_1$  是正交的。任何正交矩阵都是正规的, 我们知道一个上三角矩阵只有在它是对角矩阵时才是正规的。因此,  $U_1$  是一个对角矩阵。

我们知道正交矩阵的特征值只能是 1 或  $-1$ , 所以我们只能在  $U_1$  的对角线上有 1 或  $-1$ . 但是, 我们知道  $U_1$  的所有对角线元素, 除了最后的之外, 都是非负的, 所以  $U_1$  的所有对角线元素, 除了最后的之外, 都是 1. 最后一个对角线元素可以是  $\pm 1$ .

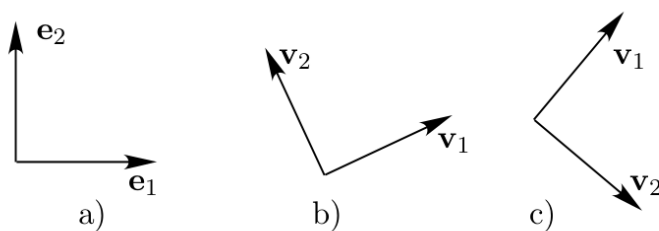
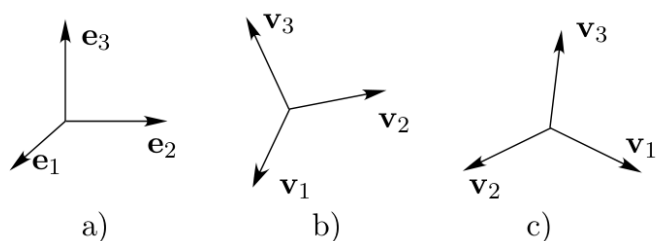
由于初等旋转的行列式为 1, 我们可以得出  $\det U_1 = \det U = 1$ , 所以最后一个对角线元素也必须是 1. 因此  $U_1 = I$ , 所以  $U$  可以表示为初等旋转的乘积  $U = R_1^{-1} R_2^{-1} \dots R_N^{-1}$ . 这里我们使用了初等旋转的逆也是初等旋转这一事实。  $\square$

## 6. 定向

### 6.1. 动机

图0.4 和图0.5 分别展示了  $\mathbb{R}^2$  和  $\mathbb{R}^3$  中的 3 组标准正交基。在每张图中, 基 b) 可以通过一次旋转从标准基 a) 获得, 而不可能通过旋转将标准基 a) 变成基 c) (使得  $\mathbf{e}_k$  变成  $\mathbf{v}_k$  对所有  $k$  成立)。

你可能以前听过“定向”(orientation)这个词, 并且可能知道基 a) 和 b) 具有正定向, 而基 c) 的定向是负的。你可能还知道一些确定定向的规则, 例如物理学中的右手定则。所以, 如果你能“看到”一组基, 比如在  $\mathbb{R}^3$  中, 你大概可以判断它的定向是什么。

图 0.4  $\mathbb{R}^2$  上的定向图 0.5  $\mathbb{R}^3$  上的定向

但如果只给出向量  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  的坐标呢？当然，你可以尝试画一张图来可视化向量，然后看看定向是什么。但这并非总是一件容易的事。更重要的是，你如何“告诉”计算机呢？

事实证明，有一个更简单的方法。让我们来解释一下。我们需要核验是否有可能在  $\mathbb{R}^3$  中通过旋转标准基  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  来得到基  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 。存在唯一一个线性变换  $U$ ，使得

$$U\mathbf{e}_k = \mathbf{v}_k, \quad k = 1, 2, 3;$$

它的矩阵（在标准基下）是由列向量  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  组成的。它是一个正交矩阵（因为它将一组标准正交基变换为另一组标准正交基），所以我们需要看看它何时是旋转。定理 5.1 和 5.2 给出了答案：矩阵  $U$  是旋转当且仅当  $\det U = 1$ 。注意（对于  $3 \times 3$  矩阵），如果  $\det U = -1$ ，那么  $U$  是围绕某个轴的旋转与该旋转平面（即该轴的垂直平面）上的反射的复合。

这为下面的形式化定义提供了动机。

## 6.2. 形式定义

设  $\mathcal{A}$  和  $\mathcal{B}$  是实向量空间  $X$  中的两组基。如果坐标变换矩阵  $[I]_{\mathcal{B}, \mathcal{A}}$  的行列式为正，我们说基  $\mathcal{A}$  和  $\mathcal{B}$  具有**相同**的定向；如果行列式为负，我们说它们具有**不同**的定向。

注意，由于  $[I]_{\mathcal{A}, \mathcal{B}} = [I]_{\mathcal{B}, \mathcal{A}}^{-1}$ ，我们可以在定义中使用矩阵  $[I]_{\mathcal{A}, \mathcal{B}}$ 。

我们通常假设  $\mathbb{R}^n$  中的标准基  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  具有正定向。在一个抽象空间中, 只需要固定一组基, 并令其定向为正。

如果在  $\mathbb{R}^n$  中, 一组标准正交基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  具有正定向 (即与标准基相同的定向), 那么定理 5.1 和 5.2 说明基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是通过一次旋转从标准基获得的。

### 6.3. 基的连续变换与定向

**定义** 我们说基  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  可以**连续地**变换为基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ , 如果存在一组基的连续族  $\mathcal{V}(t) = \{\mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_n(t)\}, t \in [a, b]$ , 使得

$$\mathbf{v}_k(a) = \mathbf{a}_k, \quad \mathbf{v}_k(b) = \mathbf{b}_k, \quad k = 1, 2, \dots, n.$$

“连续的基族”意味着向量函数  $\mathbf{v}_k(t)$  是连续的 (它们在某组基下的坐标是连续函数), 并且, 至关重要, 系统  $\mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_n(t)$  在所有  $t \in [a, b]$  都是一组基。

注意, 通过进行变量替换, 我们总可以假设 (如果需要)  $[a, b] = [0, 1]$ .

**定理 6.1.** 两组基  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  和  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  具有相同定向, 当且仅当其中一组基可以连续地变换为另一组基。

**证明:** 假设基  $\mathcal{A}$  可以连续地变换为基  $\mathcal{B}$ , 设  $\mathcal{V}(t), t \in [a, b]$  是一个连续的基族, 执行这个变换。考虑一个矩阵函数  $V(t)$ , 其列向量是  $\mathbf{v}_k(t)$  在基  $\mathcal{A}$  下的坐标向量  $[\mathbf{v}_k(t)]_{\mathcal{A}}$ .

显然,  $V(t)$  的元素是连续函数, 并且  $V(a) = I, V(b) = [I]_{\mathcal{A}, \mathcal{B}}$ . 注意, 因为  $\mathcal{V}(t)$  始终是一组基,  $\det V(t)$  永远不为零。那么, 介值定理断言  $\det V(a)$  和  $\det V(b)$  具有相同的符号。由于  $\det V(a) = \det I = 1$ , 我们可以得出  $[I]_{\mathcal{A}, \mathcal{B}}$  的行列式

$$\det[I]_{\mathcal{A}, \mathcal{B}} = \det V(b) > 0,$$

所以基  $\mathcal{A}$  和  $\mathcal{B}$  具有相同定向。

为了证明反向蕴含, 即定理的“仅当”部分, 需要证明单位矩阵  $I$  可以通过可逆矩阵连续地变换为任何满足  $\det B > 0$  的矩阵  $B$ . 换句话说, 需要证明存在一个连续的矩阵函数  $V(t)$  在区间  $[a, b]$  上, 使得对所有  $t \in [a, b]$ , 矩阵  $V(t)$  是可逆的, 并且

$$V(a) = I, \quad V(b) = B.$$

我们将证明这个事实留给读者作为练习。有几种方法可以证明这一点, 其中一种在下面的问题 6.2-6.5 中概述。□

练习

6.1. 设  $R_\alpha$  是  $\alpha$  角的旋转, 其在标准基下的矩阵为

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

求  $R_\alpha$  在基  $\mathbf{v}_1, \mathbf{v}_2$ , 其中  $\mathbf{v}_1 = \mathbf{e}_2, \mathbf{v}_2 = \mathbf{e}_1$  下的矩阵。

6.2. 设

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

是旋转矩阵。证明  $2 \times 2$  单位矩阵  $I_2$  可以通过可逆矩阵连续变换为  $R_\alpha$ 。

6.3. 设  $U$  是一个  $n \times n$  正交矩阵, 且  $\det U > 0$ . 证明  $n \times n$  单位矩阵  $I_n$  可以通过可逆矩阵连续变换为  $U$ . **提示:** 使用前一个问题和旋转在  $\mathbb{R}^n$  中的表示 (作为平面旋转的乘积), 见第 5 节。

6.4. 设  $A$  是一个  $n \times n$  正定埃尔米特矩阵,  $A = A^* > \mathbf{0}$ . 证明  $n \times n$  单位矩阵  $I_n$  可以通过可逆矩阵连续变换为  $A$ . **提示:** 对于对角矩阵怎么样?

6.5. 使用极分解和上面问题 6.3、6.4, 完成定理 6.3 的“仅当”部分的证明。



## 第七章 双线性型与二次型

尽管研究**实**二次型（即二次齐次多项式）可能是本章主题的最初动机，但**复**二次型  $(A\mathbf{x}, \mathbf{x}), \mathbf{x} \in \mathbb{C}^n, A = A^*$  也具有重要的意义。因此，除非另有说明，本章中的结论在实数和复数情况下都适用。

为了避免重复书写本质上相同的公式，我们将使用适应复数情况的记号：特别地，在实数情况下，用  $A^*$  代替  $A^T$ 。

### 1. 主要定义

#### 1.1. $\mathbb{R}^n$ 上的双线性型

$\mathbb{R}^n$  上的**双线性型**(bilinear form) 是一个含两参数  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  的函数  $L = L(\mathbf{x}, \mathbf{y})$ ，它在每个参数上都是线性的，即满足：

1.  $L(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2, \mathbf{y}) = \alpha L(\mathbf{x}_1, \mathbf{y}) + \beta L(\mathbf{x}_2, \mathbf{y})$ ;
2.  $L(\mathbf{x}, \alpha\mathbf{y}_1 + \beta\mathbf{y}_2) = \alpha L(\mathbf{x}, \mathbf{y}_1) + \beta L(\mathbf{x}, \mathbf{y}_2)$ .

双线性型的值可以属于任意向量空间，但在本书中，我们只考虑取实数值的向量。

如果  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  且  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ ，则双线性型可以写成

$$L(\mathbf{x}, \mathbf{y}) = \sum_{j,k=1}^n a_{j,k} x_k y_j,$$

或者以矩阵形式表示为

$$L(\mathbf{x}, \mathbf{y}) = (A\mathbf{x}, \mathbf{y}),$$

其中

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}.$$

矩阵  $A$  由双线性型  $L$  唯一确定。

#### 1.2. $\mathbb{R}^n$ 上的二次型

**二次型**(quadratic form) 有几种等价的定义。

可以认为二次型是双线性型  $L$  的“对角线”，即任何二次型  $Q$  由  $Q[\mathbf{x}] = L(\mathbf{x}, \mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x})$  定义。

另一种更代数的方式是，二次型是一个**二次齐次多项式**，即  $Q[\mathbf{x}]$  是一个关于  $n$  个变量  $x_1, x_2, \dots, x_n$  的多项式，只包含二次项。这意味着只允许出现  $ax_k^2$  和  $cx_jx_k$  形式的项。

可以将二次型  $Q[\mathbf{x}]$  写成  $Q[\mathbf{x}] = (\mathbf{A}\mathbf{x}, \mathbf{x})$  的方式有很多（事实上有无限多种）。例如，二次型  $Q[\mathbf{x}] = x_1^2 + x_2^2 - 4x_1x_2$  在  $\mathbb{R}^2$  上可以表示为  $(\mathbf{A}\mathbf{x}, \mathbf{x})$ ，其中  $A$  可以是以下任何一个矩阵：

$$\begin{pmatrix} 1 & -4 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}.$$

事实上，任何形式为

$$\begin{pmatrix} 1 & a \\ -4-a & 1 \end{pmatrix}$$

的矩阵都可以。

但是，如果我们要求矩阵  $A$  是对称的，那么这样的矩阵是唯一的：

$\mathbb{R}^n$  上的任何二次型  $Q[\mathbf{x}]$  都存在唯一的表示  $Q[\mathbf{x}] = (\mathbf{A}\mathbf{x}, \mathbf{x})$ ，其中  $A$  是一个（实）对称矩阵。

例如，对于二次型

$$Q[\mathbf{x}] = x_1^2 + 3x_2^2 + 5x_3^2 + 4x_1x_2 - 16x_1x_3 + 7x_2x_3$$

在  $\mathbb{R}^3$  上，对应的对称矩阵  $A$  是

$$A = \begin{pmatrix} 1 & 2 & -8 \\ 2 & 3 & 3.5 \\ -8 & 3.5 & 5 \end{pmatrix}.$$

### 1.3. $\mathbb{C}^n$ 上的二次型

也可以在  $\mathbb{C}^n$ （或任何复内积空间）上定义一个**二次型**，通过取一个自伴算子  $A = A^*$ ，并定义  $Q$  为  $Q[\mathbf{x}] = (\mathbf{A}\mathbf{x}, \mathbf{x})$ 。虽然后面大多数例子都在  $\mathbb{R}^n$  上，但所有定理在  $\mathbb{C}^n$  上也成立。考虑到这一点，我们将始终使用  $A^*$  而不是  $A^T$ 。

与实数情况的唯一本质区别是，在复数情况下我们没有选择的自由：如果要求二次型中的项均为实数，则对应的矩阵必须是埃尔米特的（自伴随的）。

注意到如果  $A = A^*$ ，那么

$$(\mathbf{A}\mathbf{x}, \mathbf{x}) = (\mathbf{x}, A^*\mathbf{x}) = (\mathbf{x}, \mathbf{A}\mathbf{x}) = \overline{(\mathbf{A}\mathbf{x}, \mathbf{x})},$$

所以  $(A\mathbf{x}, \mathbf{x}) \in \mathbb{R}$ .

其逆命题也成立。

**引理 1.1.** 设  $(A\mathbf{x}, \mathbf{x})$  对所有  $\mathbf{x} \in \mathbb{C}^n$  都是实数。那么  $A = A^*$ .

我们把证明留给读者作为练习，见下面的问题 1.4。

证明引理 1.1 的一种可能方法是使用下面版本的极化恒等式。

**引理 1.2.** 设  $A$  是内积空间  $X$  中的一个算子。

1. 如果  $X$  是一个复空间，则对于任意  $\mathbf{x}, \mathbf{y} \in X$ ,

$$(A\mathbf{x}, \mathbf{y}) = \frac{1}{4} \sum_{\alpha \in \mathbb{C}: \alpha^4=1} \alpha (A(\mathbf{x} + \alpha\mathbf{y}), \mathbf{x} + \alpha\mathbf{y}).$$

2. 如果  $X$  是一个实空间且  $A = A^*$ ，则对于任意  $\mathbf{x}, \mathbf{y} \in X$ ,

$$(A\mathbf{x}, \mathbf{y}) = \frac{1}{4} [(A(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) - (A(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y})].$$

引理 1.2 的证明请参见上面第 5 章的练习 6.3。

### 练习

1.1. 求  $\mathbb{R}^3$  上双线性型  $L$  的矩阵，其中

$$L(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_1y_2 + 14x_1y_3 - 5x_2y_1 + 2x_2y_2 - 3x_2y_3 + 8x_3y_1 + 19x_3y_2 - 2x_3y_3.$$

1.2. 通过

$$L(\mathbf{x}, \mathbf{y}) = \det[\mathbf{x}, \mathbf{y}]$$

在  $\mathbb{R}^2$  上定义双线性型  $L$  (即，计算  $L(\mathbf{x}, \mathbf{y})$  时，我们构造一个以  $\mathbf{x}, \mathbf{y}$  为列的  $2 \times 2$  矩阵并计算其行列式)。

求  $L$  的矩阵。

1.3. 求  $\mathbb{R}^3$  上二次型  $Q$  的矩阵，其中

$$Q[\mathbf{x}] = x_1^2 + 2x_1x_2 - 3x_1x_3 - 9x_2^2 + 6x_2x_3 + 13x_3^2.$$

1.4. 证明上面的引理 1.1。

**提示：**考虑表达式  $(A(\mathbf{x} + z\mathbf{y}), \mathbf{x} + z\mathbf{y})$ ，并证明如果它对所有  $z \in \mathbb{C}$  都是实数，那么  $(A\mathbf{x}, \mathbf{y}) = (\mathbf{y}, A^*\mathbf{x})$ 。

## 2. 二次型的对角化

你可能之前在研究平面上的二次曲线时就遇到过二次型。也许你甚至研究过  $\mathbb{R}^3$  中的二次曲面。

我们想为这些对象的分类提供一个统一的方法。假设我们有一个在  $\mathbb{R}^n$  中的集合，由方程  $Q[\mathbf{x}] = 1$  定义，其中  $Q$  是某个二次型。如果  $Q$  具有某种简单的形式，例如，如果其对应的矩阵是对角矩阵，即如果  $Q[\mathbf{x}] = a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2$ ，那么我们可以很容易地可视化这个集合，特别是当  $n = 2, 3$  时。在更高的维度中，即使不能可视化，也能很好地理解集合的结构。

因此, 如果我们给定一个一般、复杂的二次型, 我们总想尽可能地简化它, 例如, 使其对角化。实现这一目标的标准方法是变量替换。

## 2.1. 正交 (相似) 对角化

设我们有一个在  $\mathbb{F}^n$  ( $\mathbb{F}$  是  $\mathbb{R}$  或  $\mathbb{C}$ ) 中的二次型  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ . 引入新的变量  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{F}^n$ , 其中  $\mathbf{y} = S^{-1}\mathbf{x}$ , 这里  $S$  是某个可逆的  $n \times n$  矩阵, 所以  $\mathbf{x} = S\mathbf{y}$ .

那么,

$$Q[\mathbf{x}] = Q[S\mathbf{y}] = (AS\mathbf{y}, S\mathbf{y}) = (S^*AS\mathbf{y}, \mathbf{y}).$$

所以, 在新变量  $\mathbf{y}$  下, 二次型具有矩阵  $S^*AS$ .

因此, 我们想找到一个可逆矩阵  $S$ , 使得矩阵  $S^*AS$  是对角矩阵。注意, 这与我们之前讨论的矩阵对角化是不同的: 我们试图将矩阵  $A$  表示为  $A = SDS^{-1}$ , 所以对角矩阵  $D = S^{-1}AS$ . 然而, 对于酉矩阵  $U$ , 我们有  $U^* = U^{-1}$ , 我们可以正交对角化对称矩阵。因此, 我们可以将之前研究的**正交对角化**应用于二次型。

具体来说, 我们可以将矩阵  $A$  表示为  $A = UDU^* = UDU^{-1}$ . 回想一下,  $D$  是一个对角矩阵, 对角线上的项是  $A$  的特征值, 而  $U$  是特征向量构成的矩阵 (我们需要选择一组正交的特征向量基)。那么  $D = U^*AU$ , 所以在新变量  $\mathbf{y} = U^{-1}\mathbf{x}$  下, 二次型具有对角矩阵。

让我们分析一下正交对角化的几何意义。酉矩阵  $U$  的列  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  构成了  $\mathbb{F}^n$  中的一组标准正交基, 称之为基  $\mathcal{B}$ . 从这组基到标准基的坐标变换矩阵  $[I]_{\mathcal{S}, \mathcal{B}}$  正好是  $U$ . 我们知道  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T = U^{-1}\mathbf{x}$ , 所以坐标  $y_1, y_2, \dots, y_n$  可以解释为向量  $\mathbf{x}$  在新基  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  下的坐标。

因此, 正交对角化可以让我们非常清晰地可视化集合  $Q[\mathbf{x}] = 1$ , 或者类似的集合, 只要我们能对对角矩阵可视化。

**例子** 考虑一个二次变量的二次型 (即  $\mathbb{R}^2$  上的二次型),  $Q(x, y) = 2x^2 + 2y^2 + 2xy$ . 让我们描述满足  $Q(x, y) = 1$  的点  $(x, y)^T \in \mathbb{R}^2$  的集合。

$Q$  的矩阵是

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

对该矩阵进行正交对角化, 我们可以将其表示为

$$A = U \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} U^*, \quad \text{其中} \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

或者等价地

$$U^*AU = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} =: D.$$

集合  $\{\mathbf{y} : (D\mathbf{y}, \mathbf{y}) = 1\}$  是半轴为  $1/\sqrt{3}$  和 1 的椭圆。因此, 集合  $\{\mathbf{x} \in \mathbb{R}^2 : (A\mathbf{x}, \mathbf{x}) = 1\}$  和它是同一个椭圆, 只是在新基  $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T, (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$  下, 或者说, 前者只是同一个椭圆被旋转了  $\pi/4$ 。

## 2.2. 非正交对角化

正交对角化涉及到计算特征值和特征向量, 所以对于大的  $n$  可能难以用计算机完成。另一方面, 非正交对角化, 即找到一个可逆矩阵  $S$  (不要求  $S^{-1} = S^*$ ) 使得  $D = S^*AS$  是对角矩阵, 这计算起来就要容易得多, 并且只需要代数运算 (加、减、乘、除)。

下面我们介绍两种最常用的非正交对角化方法。

### 2.2.1. 通过配方法对角化

第一种方法基于配方法。我们将以实二次型 ( $\mathbb{R}^n$  上的二次型) 为例来说明这种方法。经过简单的修改, 这种方法也可以用于复数情况, 但我们在此不作讨论。如有必要, 感兴趣的读者应能自行进行适当的修改。

再次考虑一个二次变量的二次型,  $Q[\mathbf{x}] = 2x_1^2 + 2x_1x_2 + 2x_2^2$  (与上面例子中的二次型相同, 只是这里我们称变量为  $x_1, x_2$  而不是  $x, y$ )。由于

$$2(x_1 + \frac{1}{2}x_2)^2 = 2(x_1^2 + 2x_1\frac{1}{2}x_2 + \frac{1}{4}x_2^2) = 2x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2$$

(注意, 前两项与  $Q$  的前两项相同), 我们得到

$$2x_1^2 + 2x_1x_2 + 2x_2^2 = 2(x_1 + \frac{1}{2}x_2)^2 + \frac{3}{2}x_2^2 = 2y_1^2 + \frac{3}{2}y_2^2,$$

其中  $y_1 = x_1 + \frac{1}{2}x_2$ ,  $y_2 = x_2$ 。

同样的方法可以应用于多于两个变量的二次型。例如, 考虑  $\mathbb{R}^3$  中的二次型  $Q[\mathbf{x}]$ :

$$Q[\mathbf{x}] = x_1^2 - 6x_1x_2 + 4x_1x_3 - 6x_2x_3 + 8x_2^2 - 3x_3^2.$$

考虑所有涉及第一个变量  $x_1$  的项 (本例中为前三项), 我们提取一个包含这些项 (加上其他项) 的完全平方或其倍数。

由于

$$(x_1 - 3x_2 + 2x_3)^2 = x_1^2 - 6x_1x_2 + 4x_1x_3 - 12x_2x_3 + 9x_2^2 + 4x_3^2,$$

我们可以将二次型重写为

$$(x_1 - 3x_2 + 2x_3)^2 - x_2^2 + 6x_2x_3 - 7x_3^2.$$

注意, 表达式  $-x_2^2 + 6x_2x_3 - 7x_3^2$  只包含变量  $x_2$  和  $x_3$ . 由于

$$-(x_2 - 3x_3)^2 = -(x_2^2 - 6x_2x_3 + 9x_3^2) = -x_2^2 + 6x_2x_3 - 9x_3^2,$$

我们有

$$-x_2^2 + 6x_2x_3 - 7x_3^2 = -(x_2 - 3x_3)^2 + 2x_3^2.$$

因此, 我们可以将二次型  $Q$  写成

$$Q[\mathbf{x}] = (x_1 - 3x_2 + 2x_3)^2 - (x_2 - 3x_3)^2 + 2x_3^2 = y_1^2 - y_2^2 + 2y_3^2,$$

其中

$$y_1 = x_1 - 3x_2 + 2x_3, \quad y_2 = x_2 - 3x_3, \quad y_3 = x_3.$$

最后, 让我们来解决一个细心的读者可能已经提出的问题: 如果我们某个时候得到了两个变量的乘积, 但没有相应的平方项, 这时该怎么办? 例如, 如何对角化形式  $x_1x_2$ ? 答案直接来自恒等式

$$(2.1) \quad 4x_1x_2 = (x_1 + x_2)^2 - (x_1 - x_2)^2,$$

这给出了表示

$$Q[\mathbf{x}] = y_1^2 - y_2^2, \quad y_1 = (x_1 + x_2)/2, \quad y_2 = (x_1 - x_2)/2.$$

### 2.2.2. 使用行/列运算进行对角化

还有另一种对二次型进行非正交对角化的方法。其思想是对二次型的矩阵  $A$  进行行运算。与高斯-若尔当消元法的区别在于, 在每次行运算后, 我们需要执行相同的列运算, 原因是我们想得到对角矩阵  $S^*AS$ 。

我们以一个例子来说明这个方法是如何进行的。假设我们要对角化矩阵

$$A = \begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & 1 \\ 3 & 1 & 1 \end{pmatrix}$$

的二次型。我们将矩阵  $A$  与单位矩阵进行拼接, 并对增广矩阵  $(A|I)$  执行行/列运

算。在每次行运算后，我们必须对矩阵  $A$  执行相同的列运算。

$$\begin{aligned}
 & \left( \begin{array}{ccc|ccc} 1 & -1 & 3 & 1 & 0 & 0 \\ -1 & 2 & 1 & 0 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{R_2+R_1} \left( \begin{array}{ccc|ccc} 1 & -1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{C_2+C_1} \\
 & \left( \begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 3 & 4 & 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{R_3-3R_1} \left( \begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 0 & 4 & -8 & -3 & 0 & 1 \end{array} \right) \xrightarrow{C_3-3C_1} \\
 & \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 0 & 4 & -8 & -3 & 0 & 1 \end{array} \right) \xrightarrow{R_3-4R_2} \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 & 0 \\ 0 & 0 & -24 & -7 & -4 & 1 \end{array} \right) \xrightarrow{C_3-4C_2} \\
 & \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -24 & -7 & -4 & 1 \end{array} \right)
 \end{aligned}$$

注意，我们只对增广矩阵的左侧执行列运算。

我们得到左侧的对角矩阵  $D$ ，右侧的矩阵  $S^*$ ，所以  $D = S^*AS$ 。

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -24 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -7 & -4 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & 1 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 7 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix}.$$

让我解释一下这个方法为何有效。行运算是通过左乘一个初等矩阵实现的。对应的列运算是通过右乘其转置的初等矩阵实现的。因此，执行行运算  $E_1, E_2, \dots, E_N$  和相同的列运算，我们将矩阵  $A$  转换为

$$(2.2) \quad E_N \dots E_2 E_1 A E_1^* E_2^* \dots E_N^* = E A E^*.$$

至于右侧的单位矩阵，我们只对其执行了行运算，所以单位矩阵变为

$$E_N \dots E_2 E_1 I = EI = E.$$

如果我们现在设  $E^* = S$ ，则我们得到  $S^*AS$  是一个对角矩阵，而矩阵  $E = S^*$  是变换后的增广矩阵的右半部分。

在上面的例子中，我们很幸运，因为我们不需要交换两行。行交换这个操作稍微棘手一些。如果你知道该怎么做，它会很简单，但你可能很难猜到正确的行运算。例如，考虑一个矩阵为

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

的二次型。如果我们想通过行和列运算来对角化它，最简单的想法是交换第 1 行和第 2 行。但我们也必须执行相同的列运算，即交换第 1 列和第 2 列，所以我们会得到相同的矩阵。因此，我们需要一些更不寻常的操作。例如，恒等式 (2.1) 可以用来对角化这个二次型。

然而，一个更简单的想法也奏效：使用行运算来得到对角线上的非零项！例如，如果我们开始使  $a_{1,1}$  非零，下面的连续行（及相应的列）运算是一种可能的选择：

$$\begin{aligned} & \left( \begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{R_1 + \frac{1}{2}R_2} \left( \begin{array}{cc|cc} 1/2 & 1 & 1 & 1/2 \\ 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{C_1 + \frac{1}{2}C_2} \\ & \left( \begin{array}{cc|cc} 1 & 1 & 1 & 1/2 \\ 1 & 0 & 0 & 1 \end{array} \right) \xrightarrow{R_2 - R_1} \left( \begin{array}{cc|cc} 1 & 1 & 1 & 1/2 \\ 0 & -1 & -1 & 1/2 \end{array} \right) \xrightarrow{C_2 - C_1} \\ & \left( \begin{array}{cc|cc} 1 & 0 & 1 & 1/2 \\ 0 & -1 & -1 & 1/2 \end{array} \right) \end{aligned}$$

**注记** 非正交对角化在一组非正交基下提供了对集合  $Q[\mathbf{x}] = 1$  的简单描述。它比正交对角化给出的表示更难可视化。然而，如果我们不关心细节，例如，如果我们只需要知道该集合是否为椭圆（或双曲面等），那么非正交对角化是获得答案的更简单方法。

**注记 2.1** 对于复数二次型（即形式为  $(A\mathbf{x}, \mathbf{x})$ ,  $A = A^*$ ），非正交对角化与实数情况的处理方式相同，唯一的区别是相应的“列运算”具有共轭复数系数。

原因是，如果一个行运算是通过左乘一个初等矩阵  $E_k$  给出的，那么相应的列运算则是通过右乘  $E_k^*$  给出的，见 (2.2)。

注意到公式 (2.2) 在复数和实数情况下都适用：在实数情况下，我们可以写  $E_k^T$  而不是  $E_k^*$ ，但使用埃尔米特伴随允许我们在两种情况下使用相同的公式。

## 练习

### 2.1. 对矩阵

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

的二次型进行对角化。使用两种方法：配方法和行运算。你更喜欢哪一种？你能判断矩阵  $A$  是否是正定的吗？

### 2.2. 对于矩阵

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

正交对角化相应的二次型，即找到一个对角矩阵  $D$  和一个酉矩阵  $U$ ，使得  $D = U^*AU$ 。



### 3. 惯性定理

如上所述, 对角化二次型有许多方法。注意, 最终的对角矩阵不是唯一的。例如, 如果我们得到一个对角矩阵

$$D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\},$$

我们可以取一个对角矩阵

$$S = \text{diag}\{s_1, s_2, \dots, s_n\}, \quad s_k \in \mathbb{R}, \quad s_k \neq 0,$$

并将  $D$  变换为

$$S^*DS = \text{diag}\{s_1^2\lambda_1, s_2^2\lambda_2, \dots, s_n^2\lambda_n\}.$$

这种变换改变了  $D$  的对角线元素。然而, 它**不改变**对角线元素的**符号**。这始终是成立的!

也就是说, 著名的**惯性定理** (Sylvester' s Law of Inertia) 指出:

对于一个埃尔米特矩阵  $A$  (即二次型  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ ) 及其任意对角化  $D = S^*AS$ ,  $D$  的正 (负、零) 对角线元素的数量仅取决于  $A$ , 而不取决于对角化的具体选择。

这里我们当然假设  $S$  是一个可逆矩阵,  $D$  是一个对角矩阵。

惯性定理证明的思路是, 用与  $S$  或  $D$  无关的  $A$  来表达对角化  $D = S^*AS$  的正 (负、零) 对角线元素的数量。

我们将需要以下定义。

**定义** 给定一个  $n \times n$  埃尔米特矩阵  $A = A^*$  ( $\mathbb{F}^n$  上的二次型  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ ), 我们称一个子空间  $E \subset \mathbb{F}^n$  为**正的** (分别, **负的**, **零的**), 如果对所有  $\mathbf{x} \in E, \mathbf{x} \neq 0$ , 都有

$$(A\mathbf{x}, \mathbf{x}) > 0 \quad (\text{分别 } (A\mathbf{x}, \mathbf{x}) < 0, \quad (A\mathbf{x}, \mathbf{x}) = 0.)$$

有时, 为了强调  $A$  的作用, 我们会说  $A$ -正 ( $A$ -负,  $A$ -零)。

**定理 3.1.** 设  $A$  是一个  $n \times n$  埃尔米特矩阵,  $D = S^*AS$  是它的对角化 (通过可逆矩阵  $S$ )。那么  $D$  的正 (分别负) 对角线元素的数量等于  $A$ -正 (分别  $A$ -负) 子空间的**最大维度**。

上面的定理说明, 如果  $r_+$  是  $D$  的正对角线元素的数量, 那么存在一个维度为  $r_+$  的  $A$ -正子空间  $E$ , 但是不可能找到一个维度大于  $r_+$  的正子空间  $E$ 。

我们将需要以下引理, 它可以被视为上述定理的一个特例。

**引理 3.2.** 设  $D$  是一个对角矩阵  $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . 那么  $D$  的正 (分别负) 对角线元素的数量等于  $D$ -正 (分别  $D$ -负) 子空间的**最大维度**。

**证明：**通过重新排列  $\mathbb{F}^n$  中的标准基（改变编号），我们可以无损地假设正对角线元素是前  $r_+$  个对角线元素。

考虑由前  $r_+$  个坐标向量  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{r_+}$  张成的子空间  $E_+$ . 显然  $E_+$  是一个  $D$ -正子空间，且  $\dim E_+ = r_+$ .

现在我们证明对于任何其他的  $D$ -正子空间  $E$ ，都有  $\dim E \leq r_+$ . 考虑正交投影  $P = P_{E_+}$ ,

$$P\mathbf{x} = (x_1, x_2, \dots, x_{r_+}, 0, \dots, 0)^T, \quad \mathbf{x} = (x_1, x_2, \dots, x_n)^T.$$

对于一个  $D$ -正子空间  $E$ ，定义一个算子  $T: E \rightarrow E_+$  为

$$T\mathbf{x} = P\mathbf{x}, \quad \forall \mathbf{x} \in E.$$

换句话说， $T$  是投影  $P$  的**限制 (restriction)** ( $P$  定义在整个空间上，但我们将它的定义域限制到  $E$ ，目标空间限制到  $E_+$ )。我们得到一个从  $E$  到  $E_+$  的算子，并使用另一个字母来区分它与  $P$ 。

注意到  $\text{Ker } T = \{\mathbf{0}\}$ . 确实，设  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in E$  且  $T\mathbf{x} = P\mathbf{x} = \mathbf{0}$ . 那么，根据  $P$  的定义，

$$x_1 = x_2 = \dots = x_{r_+} = 0.$$

因此

$$(D\mathbf{x}, \mathbf{x}) = \sum_{k=r_++1}^n \lambda_k x_k^2 \leq 0 \quad (\lambda_k \leq 0 \text{ 对 } k > r_+).$$

但是  $\mathbf{x}$  属于一个  $D$ -正子空间  $E$ ，所以不等式  $(D\mathbf{x}, \mathbf{x}) \leq 0$  仅对  $\mathbf{x} = \mathbf{0}$  成立。

现在我们应用秩定理(第二章定理 7.1)。首先， $\text{rank } T = \dim \text{Ran } T \leq \dim E_+ = r_+$ ，因为  $\text{Ran } T \subset E_+$ . 根据秩定理， $\dim \text{Ker } T + \text{rank } T = \dim E$ . 但是我们刚刚证明了  $\text{Ker } T = \{\mathbf{0}\}$ ，即  $\dim \text{Ker } T = 0$ ，所以

$$\dim E = \text{rank } T \leq \dim E_+ = r_+.$$

要证明关于负对角线元素的命题，只需对矩阵  $-D$  应用上述推理。 □

**定理 3.1 的证明** 设  $D = S^*AS$  是  $A$  的一个对角化。由于

$$(D\mathbf{x}, \mathbf{x}) = (S^*AS\mathbf{x}, \mathbf{x}) = (AS\mathbf{x}, S\mathbf{x}),$$

可知对于任一  $D$ -正子空间  $E$ ，子空间  $SE$  是一个  $A$ -正子空间。同样的恒等式暗示，对于任一  $A$ -正子空间  $F$ ，子空间  $S^{-1}F$  是  $D$ -正的。

由于  $S$  和  $S^{-1}$  是可逆变换， $\dim E = \dim SE$  且  $\dim F = \dim S^{-1}F$ . 因此，对于任何  $D$ -正子空间  $E$ ，我们可以找到一个相同维度的  $A$ -正子空间（即  $SE$ ），反

之亦然：对于任何  $A$ -正子空间  $F$ ，我们可以找到一个相同维度的  $D$ -正子空间（即  $S^{-1}F$ ）。因此， $A$ -正子空间和  $D$ -正子空间的最大维度是相同的，定理得证。

关于负对角线元素的处理方式类似，细节留作读者练习。  $\square$

## 4. 正定二次型 极小极大特征值刻画与塞尔维斯特正定性判据

**定义** 一个二次型  $Q$  被称为

- **正定**，如果对所有  $\mathbf{x} \neq \mathbf{0}$ ，有  $Q[\mathbf{x}] > 0$ 。
- **半正定**，如果对所有  $\mathbf{x}$ ，有  $Q[\mathbf{x}] \geq 0$ 。
- **负定**，如果对所有  $\mathbf{x} \neq \mathbf{0}$ ，有  $Q[\mathbf{x}] < 0$ 。
- **半负定**，如果对所有  $\mathbf{x}$ ，有  $Q[\mathbf{x}] \leq 0$ 。
- **不定**，如果它既能取到正值，又能取到负值，即存在向量  $x_1$  和  $x_2$  使得  $Q[x_1] > 0$  且  $Q[x_2] < 0$ 。

**定义** 一个埃尔米特矩阵  $A = A^*$  被称为是**正定**（**负定**，……）的，如果相应的二次型  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$  是正定（负定，……）的。

**定理 4.1.** 设  $A = A^*$ 。那么

1.  $A$  是正定的，当且仅当  $A$  的所有特征值都是正的。
2.  $A$  是半正定的，当且仅当  $A$  的所有特征值都是非负的。
3.  $A$  是负定的，当且仅当  $A$  的所有特征值都是负的。
4.  $A$  是半负定的，当且仅当  $A$  的所有特征值都是非正的。
5.  $A$  是不定的，当且仅当它同时具有正特征值和负特征值。

**证明：**证明可以平凡地从正交对角化得出。事实上，存在一组标准正交基，使得  $A$  在该基下的矩阵是对角矩阵，而对于对角矩阵，定理是显而易见的。  $\square$

**注记** 注意到，为了判断一个矩阵（二次型）是否是正定的（负定的等），不必计算特征值。根据惯性定理，只需执行任意的（不一定是正交的）对角化  $D = S^*AS$ ，然后查看  $D$  的对角线元素。

### 4.1. 塞尔维斯特正定性判据

很容易看出，一个  $2 \times 2$  矩阵

$$A = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$$

是正定的，当且仅当

$$(4.1) \quad a > 0 \quad \text{且} \quad \det A = ac - |b|^2 > 0.$$

事实上, 如果  $a > 0$  且  $\det A = ac - |b|^2 > 0$ , 那么  $c > 0$ , 所以  $\text{trace } A = a + c > 0$ . 因此, 我们知道如果  $\lambda_1, \lambda_2$  是  $A$  的特征值, 那么  $\lambda_1 \lambda_2 > 0$  ( $\det A > 0$ ) 且  $\lambda_1 + \lambda_2 = \text{trace } A > 0$ . 这只有可能在两个特征值都为正时发生. 因此, 我们证明了条件 (4.1) 蕴含  $A$  是正定的. 反向蕴含很简单, 留作读者练习.

这个结果可以推广到  $n \times n$  矩阵. 也就是说, 对于矩阵

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix},$$

我们考虑它的所有左上子矩阵 (upper left submatrices)<sup>44</sup>

$$A_1 = (a_{1,1}), \quad A_2 = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}, \quad A_3 = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}, \dots, \quad A_n = A.$$

**定理 4.2.** 塞尔维斯特正定性判据 (Sylvester's Criterion of Positivity)

矩阵  $A = A^*$  是正定的, 当且仅当

$$\det A_k > 0 \quad \text{对所有} \quad k = 1, 2, \dots, n.$$

首先, 我们注意到, 如果  $A > 0$ , 那么  $A_k > 0$  成立 (你能解释为什么吗?). 因此, 由于正定矩阵的所有特征值都是正的, 参见定理 4.1,  $\det A_k > 0$  对所有  $k = 1, 2, \dots, n$ .

我们可以证明, 如果  $\det A_k > 0 \forall k$ , 那么  $A$  的所有特征值都是正的. 证明通过分析使用行和列运算进行的二次型对角化, 该方法在第 2.2 节中已有所描述. 关键在于观察到, 如果我们按自然顺序执行行/列运算 (即先从所有其他行/列减去第一行/列, 然后从第 3, 4, ...,  $n$  行/列减去第二行/列, 依此类推), 并且我们不进行任何行交换, 那么我们也自动地对角化了子二次型  $A_k$ . 也就是说, 在减去第一行和第二行 (以及列) 后, 我们得到了  $A_2$  的对角化; 在减去第三行/列后, 我们得到了  $A_3$  的对角化, 依此类推.

由于我们只执行行替换, 我们不会改变行列式. 而且, 由于我们不执行行交换并以正确的顺序执行运算, 我们保留了  $A_k$  的行列式. 因此, 条件  $\det A_k > 0$  保证了每个新的对角线元素都是正的.

当然, 必须确保我们只能使用行替换, 并按正确的顺序执行运算, 即我们不会遇到任何病态情况. 如果分析算法, 可以看出唯一可能发生的非正常情况是在

<sup>44</sup>译者注: 它们的行列式被称为“顺序主子式”。

某个步骤中主元位置为零。换句话说，如果我们减去前  $k$  行和列并得到  $A_k$  的对角化，那么第  $k+1$  行和第  $k+1$  列的元素为 0。我们把证明该情况不可能发生留给读者作为练习。

上面概述的证明很简单。然而，让我更详细地介绍另一种证明，这种证明可以在更高级的教科书中找到。我个人更喜欢第二个证明，因为它展示了一些重要的联系。

我们将需要以下埃尔米特矩阵特征值的一个刻画。

## 4.2. 特征值的极小极大刻画

回想一下，子空间  $E \subset X$  的**余维度 (codimension)** 被定义为其正交补的维度， $\text{codim } E = \dim(E^\perp)$ 。由于对于子空间  $E \subset X$ ， $\dim X = n$ ，我们有  $\dim E + \dim E^\perp = n$ ，所以  $\text{codim } E = \dim X - \dim E$ 。

回想一下，平凡子空间  $\{0\}$  的维度为零，因此整个空间  $X$  的余维度为 0。

**定理 4.3.** (特征值的极小极大刻画)

设  $A = A^*$  是一个  $n \times n$  矩阵，设  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$  是其特征值（按降序排列）。那么

$$\lambda_k = \max_{\substack{E: \\ \dim E = k}} \min_{\substack{\mathbf{x} \in E \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \min_{\substack{F: \\ \text{codim } F = k-1}} \max_{\substack{\mathbf{x} \in F \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}).$$

我们更详细地解释一下像  $\max \min$  和  $\min \max$  这样的表达式的含义。为了计算第一个表达式，我们需要考虑所有维度为  $k$  的子空间  $E$ 。对于每个这样的子空间  $E$ ，我们考虑所有范数为 1 的  $\mathbf{x} \in E$  的集合，并找到  $(A\mathbf{x}, \mathbf{x})$  在所有这些  $\mathbf{x}$  上的最小值。因此，对于每个子空间，我们得到一个数字，我们需要选择一个子空间  $E$  使得该数字最大。这就是  $\max \min$ 。

$\min \max$  的定义类似。

**注记** 一个敏锐的读者可能会注意到一个问题：为什么这里最大值和最小值存在？众所周知，最大值和最小值往往不存在：例如，函数  $f(x) = x$  在开区间  $(0, 1)$  上既没有最大值也没有最小值。

然而，在这种情况下，最大值和最小值确实存在。存在两种关于  $(A\mathbf{x}, \mathbf{x})$  会达到最大值和最小值的解释：第一种需要对分析的基本概念有一定的熟悉度：只需说明  $E$  中的单位球面，即集合  $\{\mathbf{x} \in E : \|\mathbf{x}\| = 1\}$  是紧致的，而一个连续函数（在这种情况下，我们的  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ ）在紧致集上会达到其最大值和最小值。

另一种解释是注意到函数  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ ， $\mathbf{x} \in E$ ，是  $E$  上的一个二次型。在  $E$  的某组标准正交基下计算该二次型的矩阵并不难，但让我们指出该矩阵并不是  $A$ ：它必须是一个  $k \times k$  矩阵，其中  $k = \dim E$ 。

容易看出, 对于二次型, 在单位球体上的最大值和最小值是其矩阵的最大和最小特征值。

至于在所有子空间上的推广, 我们将在下面证明最大值和最小值确实存在。

**定理 7.1 的证明** 首先, 通过选择适当的标准正交基, 我们可以无损地假设矩阵  $A$  是对角矩阵,  $A = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ .

选择维度为  $k$  的子空间  $E$  和余维度为  $k-1$  的子空间  $F$  (即  $\dim F = n - (k-1) = n - k + 1$ )。由于  $\dim E + \dim F = k + n - k + 1 = n + 1 > n$ , 存在一个非零向量  $\mathbf{x}_0 \in E \cap F$ . 通过归一化它可以无损地假设  $\|\mathbf{x}_0\| = 1$ . 我们可以总是将特征值按降序排列, 所以假设  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ .

由于  $\mathbf{x}$  属于  $E$  和  $F$  两个子空间,

$$\min_{\substack{\mathbf{x} \in E, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) \leq (A\mathbf{x}_0, \mathbf{x}_0) \leq \max_{\substack{\mathbf{x} \in F, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}).$$

我们没有对子空间  $E$  和  $F$  做出任何除了维度之外的假设, 所以上述不等式

$$(4.2) \quad \min_{\substack{\mathbf{x} \in E, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) \leq \max_{\substack{\mathbf{x} \in F, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x})$$

对所有适当维度的  $E$  和  $F$  的对都成立。

定义

$$E_0 := \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k\}, \quad F_0 := \text{span}\{\mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}.$$

由于对于自伴矩阵  $B$ ,  $(B\mathbf{x}, \mathbf{x})$  在单位球体  $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$  上的最大值和最小值分别是最大和最小特征值 (在对角矩阵上很容易验证), 我们得到

$$\min_{\substack{\mathbf{x} \in E_0, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \max_{\substack{\mathbf{x} \in F_0, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \lambda_k.$$

从 (4.2) 可知, 对于任何子空间  $E$ ,  $\dim E = k$ ,

$$\min_{\mathbf{x} \in E, \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}) \leq \max_{\mathbf{x} \in F_0, \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}) = \lambda_k.$$

同理, 对于任何余维度为  $k-1$  的子空间  $F$ ,

$$\max_{\mathbf{x} \in F, \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}) \geq \min_{\mathbf{x} \in E_0, \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}) = \lambda_k.$$

但是在子空间  $E_0$  和  $F_0$  上, 最大值和最小值都是  $\lambda_k$ , 所以  $\min \max = \max \min = \lambda_k$ .  $\square$

**推论 4.4.** (特征值的交错性质) 设  $A = A^* = \{a_{j,k}\}_{j,k=1}^n$  是一个自伴矩阵, 令  $\tilde{A} = \{a_{j,k}\}_{j,k=1}^{n-1}$  为它  $(n-1) \times (n-1)$  的子矩阵 (移除最后一行和最后一列)。设  $\lambda_1, \lambda_2, \dots, \lambda_n$  和  $\mu_1, \mu_2, \dots, \mu_{n-1}$  分别是  $A$  和  $\tilde{A}$  的特征值 (按降序排列)。那么

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n,$$

即

$$\lambda_k \geq \mu_k \geq \lambda_{k+1}, \quad k = 1, 2, \dots, n-1.$$

**证明：** 设  $\tilde{X} \subset \mathbb{F}^n$  是由前  $n-1$  组基向量张成的子空间,  $\tilde{X} = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}\}$ . 由于  $(\tilde{A}\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, \mathbf{x})$  对所有  $\mathbf{x} \in \tilde{X}$  成立, 定理 4.3 暗示

$$\mu_k = \max_{\substack{E \subset \tilde{X}, \\ \dim E = k}} \min_{\substack{\mathbf{x} \in E, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}).$$

为了得到  $\lambda_k$ , 我们需要在  $\mathbb{F}^n$  的所有维度为  $k$  的子空间上取最大值 (任何  $\tilde{X}$  的子空间都是  $\mathbb{F}^n$  的子空间)。因此,

$$\mu_k \leq \lambda_k.$$

(在更大集合上的最大值不会变得更小)。

另一方面,  $\tilde{X}$  中余维度为  $k-1$  的任何子空间  $E$  (这里指的是在  $\tilde{X}$  中的余维度) 的维度是  $n-1-(k-1) = n-k$ , 因此它在  $\mathbb{F}^n$  中的余维度是  $k$ . 所以

$$\mu_k = \min_{\substack{E \subset \tilde{X}, \\ \dim E = n-k}} \max_{\substack{\mathbf{x} \in E, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) \leq \min_{\substack{E \subset \mathbb{F}^n, \\ \dim E = n-k}} \max_{\substack{\mathbf{x} \in E, \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \lambda_{k+1}$$

(在更大集合上的最小值不会变得更大)。

□

**定理 4.2 的证明** 如果  $A > 0$ , 那么  $A_k > 0$  对  $k = 1, 2, \dots, n$  也成立 (你能解释为什么吗?). 由于正定矩阵的所有特征值都是正的 (见定理 4.1),  $\det A_k > 0$  对所有  $k = 1, 2, \dots, n$ .

现在我们来证明另一个蕴含。设  $\det A_k > 0$  对所有  $k$ . 我们将通过归纳法证明 (对  $k$ ), 所有  $A_k$  (因此  $A = A_n$ ) 都是正定的。

显然  $A_1$  是正定的 (它是一个  $1 \times 1$  矩阵, 所以  $A_1 = \det A_1$ )。假设  $A_{k-1} > 0$  (且  $\det A_k > 0$ ), 我们来证明  $A_k$  是正定的。设  $\lambda_1, \lambda_2, \dots, \lambda_k$  和  $\mu_1, \mu_2, \dots, \mu_{k-1}$  分别是  $A_k$  和  $A_{k-1}$  的特征值。根据推论 4.4,

$$\lambda_j \geq \mu_j > 0, \quad j = 1, 2, \dots, k-1.$$

由于  $\det A_k = \lambda_1 \lambda_2 \dots \lambda_{k-1} \lambda_k > 0$ , 最后一个特征值  $\lambda_k$  也必须是正的。因此, 由于其所有特征值都为正, 矩阵  $A_k$  是正定的。

□

### 4.3. 一些注记

首先, 请注意, 塞尔维斯特正定性判据不能推广到半正定矩阵 (当  $n \geq 3$  时), 这意味着对于  $n \times n$  矩阵,  $n \geq 3$ , 条件  $\det A_k \geq 0$  并不蕴含  $A$  是半正定的, 参见下面的问题 4.4。

然而, 对于  $2 \times 2$  矩阵, 条件  $\det A_k \geq 0$  蕴含  $A$  是半正定的, 参见下面的问题 4.3。这有时会导致对  $n \times n$  矩阵的错误结论。

最后, 我们应该简单地谈谈负定矩阵。这个典型的错误, 初学者常常会犯: 他们认为条件  $\det A_k < 0$  意味着  $A$  是负定的。但这却是错误的!

要核验矩阵  $A$  是否是负定的, 只需核验矩阵  $-A$  是否是正定的。将塞尔维斯特正定性判据应用于  $-A$ , 我们可以看到  $A$  是负定的, 当且仅当  $(-1)^k \det A_k > 0$  对所有  $k = 1, 2, \dots, n$ .

### 练习

#### 4.1. 使用塞尔维斯特正定性判据核验矩阵

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 4 & -2 \\ 2 & -2 & 1 \end{pmatrix}$$

否是正定的。

矩阵  $-A$ ,  $A^3$  和  $A^{-1}$ ,  $A + B^{-1}$ ,  $A + B$ ,  $A - B$  是否是正定的?

#### 4.2. 判断正误:

- 如果  $A$  是正定的, 那么  $A^5$  是正定的。
- 如果  $A$  是负定的, 那么  $A^8$  是负定的。
- 如果  $A$  是负定的, 那么  $A^{12}$  是正定的。
- 如果  $A$  是正定的, 且  $B$  是半负定的, 那么  $A - B$  是正定的。
- 如果  $A$  是不定的, 且  $B$  是正定的, 那么  $A + B$  是不定的。

#### 4.3. 设 $A$ 是一个 $2 \times 2$ 埃尔米特矩阵, 满足 $a_{1,1} \geq 0$ , $\det A \geq 0$ . 证明 $A$ 是半正定的。

#### 4.4. 找一个 $n \times n$ 实对称矩阵 $A$ , 使得 $\det A_k \geq 0$ 对所有 $k = 1, 2, \dots, n$ , 但是矩阵 $A$ 不是半正定的。注意 $n$ 至少为 3, 参见上面的问题 4.3。

#### 4.5. 设 $A$ 是一个 $n \times n$ 埃尔米特矩阵, 使得对所有 $k = 1, 2, \dots, n-1$ , 都有 $\det A_k > 0$ , 并且 $\det A \geq 0$ . 证明 $A$ 是半正定的。

#### 4.6. 找到一个 $3 \times 3$ 实对称矩阵 $A$ , 使得 $a_{1,1} > 0$ , 对 $k = 2, 3$ 有 $\det A_k \geq 0$ , 但矩阵 $A$ 不是半正定的。

## 5. 正定二次型与内积

设  $V$  是一个内积空间, 设  $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $V$  中的一组基 (不一定是正交的)。设  $G = \{g_{j,k}\}_{j,k=1}^n$  是由

$$g_{j,k} = (\mathbf{v}_k, \mathbf{v}_j)$$

定义的矩阵。

如果  $\mathbf{x} = \sum_{k=1}^n x_k \mathbf{v}_k$  且  $\mathbf{y} = \sum_{j=1}^n y_j \mathbf{v}_j$ , 那么

$$\begin{aligned} (\mathbf{x}, \mathbf{y}) &= \left( \sum_{k=1}^n x_k \mathbf{v}_k, \sum_{j=1}^n y_j \mathbf{v}_j \right) = \sum_{k,j=1}^n x_k y_j (\mathbf{v}_k, \mathbf{v}_j) \\ &= \sum_{j=1}^n \sum_{k=1}^n g_{j,k} x_k y_j = (G[\mathbf{x}]_{\mathcal{B}}, [\mathbf{y}]_{\mathcal{B}})_{\mathbb{C}^n}, \end{aligned}$$

其中  $(\cdot, \cdot)_{\mathbb{C}^n}$  代表  $\mathbb{C}^n$  中的标准内积。可以立即看出  $G$  是一个正定矩阵 (为什么?)。



因此, 当在内积空间中的任意 (不一定是正交的) 基下处理坐标时, 内积 (用坐标表示) 不是像在  $\mathbb{C}^n$  中那样通过标准内积计算, 而是通过如上所述的正定矩阵  $G$  来计算。

注意, 这个  $G$ -内积当且仅当  $G = I$  时才与  $\mathbb{C}^n$  中的标准内积重合, 这当且仅当基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是标准正交的时成立。

反之, 给定一个正定矩阵  $G$ , 可以在  $\mathbb{C}^n$  中定义一个非标准的内积 ( $G$ -内积) 为

$$(\mathbf{x}, \mathbf{y})_G := (G\mathbf{x}, \mathbf{y})_{\mathbb{C}^n}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{C}^n.$$

可以很容易地核验  $(\mathbf{x}, \mathbf{y})_G$  确实是一个内积, 即满足第 5 章第 1.3 节性质 1-4.



## 第八章 对偶空间与张量

本章中讨论的所有向量空间都是有限维的。

### 1. 对偶空间

#### 1.1. 线性泛函与对偶空间 对偶空间中的坐标变换

**定义 1.1** 向量空间  $V$  (域为  $\mathbb{F}$ ) 上的**线性泛函**(linear functional) 是一个线性变换  $L : V \rightarrow \mathbb{F}$ .

这类特殊的线性变换足够重要, 值得一个单独的名称。

如果我们把向量看作是某种物理对象, 比如力和速度, 那么线性泛函可以被看作是一个(线性的)测量, 它给你一个标量作为结果: 可以想象一下给定定向上的力和速度。

**定义 1.2** 有限维<sup>45</sup>向量空间  $V$  上所有线性泛函的集合被称为  $V$  的**对偶空间**(dual space), 通常记作  $V'$  或  $V^*$ .

正如我们在第一章第四节中讨论过的, 从  $V$  到  $W$  的所有线性变换的集合  $\mathcal{L}(V, W)$  (具有自然定义加法和标量乘法) 是一个向量空间。因此, 对偶空间  $V' = \mathcal{L}(V, \mathbb{F})$  是一个向量空间。

让我们来看一个例子。设空间  $V$  是  $\mathbb{R}^n$ , 那么它的对偶是什么? 我们知道, 线性变换  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  由一个  $m \times n$  矩阵表示, 所以  $\mathbb{R}^n$  上的线性泛函(即线性变换  $L : \mathbb{R}^n \rightarrow \mathbb{R}$ ) 由一个  $1 \times n$  矩阵(行向量)给出, 我们记之为  $[L]$ . 所有这些行向量的集合与  $\mathbb{R}^n$  同构(同构是通过取转置  $[L] \rightarrow [L]^T$  给出的)。

因此,  $\mathbb{R}^n$  的对偶就是  $\mathbb{R}^n$  本身。对于复数空间  $\mathbb{C}^n$  也是如此, 当然, 对于任意域  $\mathbb{F}$  上的  $\mathbb{F}^n$  也是如此。由于域  $\mathbb{F}$  (这里我们主要关心  $\mathbb{F} = \mathbb{R}$  或  $\mathbb{F} = \mathbb{C}$  的情况) 上  $n$  维空间  $V$  与  $\mathbb{F}^n$  同构, 而  $\mathbb{F}^n$  的对偶与  $\mathbb{F}^n$  同构, 我们可以得出对偶空间  $V'$  与  $V$  同构。

因此, 对偶空间的定义在一开始显得有些“愚蠢”, 因为它似乎没有提供任何新的东西。

---

<sup>45</sup>我们这里只考虑有限维空间, 因为对于无限维空间, 对偶空间并不完全由所谓的**有界**(bounded)线性泛函组成。在不给出精确定义的情况下, 我们只提一句: 在有限维情况下(域和目标空间都是有限维的), 所有线性变换都是有界的, 因此我们不需要提及“有界”这个词。

然而，事实并非如此！如果我们仔细观察，就会发现对偶空间确实是一个新对象。为了说明这一点，让我们分析一下当我们在  $V$  中改变基时，矩阵  $[L]$  的项（我们称之为  $L$  的坐标）是如何变化的。

### 1.1.1. 坐标变换公式

设

$$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}, \quad \mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$$

是  $V$  中的两组基，设  $[L]_{\mathcal{A}} = [L]_{S, \mathcal{A}}$  和  $[L]_{\mathcal{B}} = [L]_{S, \mathcal{B}}$  分别是  $L$  在基  $\mathcal{A}$  和  $\mathcal{B}$  下的矩阵（我们假设标量目标空间中的基总是标准基，所以我们可以省略下标  $S$ ）。然后，回忆第二章 8.4 节中的坐标变换规则，我们得到

$$[L]_{\mathcal{B}} = [L]_{\mathcal{A}} [I]_{\mathcal{A}, \mathcal{B}}.$$

回忆一下，对于向量  $\mathbf{v} \in V$ ，其不同基下的坐标由公式

$$[\mathbf{v}]_{\mathcal{B}} = [I]_{\mathcal{B}, \mathcal{A}} [\mathbf{v}]_{\mathcal{A}}$$

相关联，并且

$$[I]_{\mathcal{A}, \mathcal{B}} = [I]_{\mathcal{B}, \mathcal{A}}^{-1}.$$

如果我们令  $S := [I]_{\mathcal{B}, \mathcal{A}}$ ，那么  $[\mathbf{v}]_{\mathcal{B}} = S[\mathbf{v}]_{\mathcal{A}}$ 。那么  $[L]_{\mathcal{B}}^T$  和  $[L]_{\mathcal{A}}^T$  的项由公式

$$(1.1) \quad [L]_{\mathcal{B}}^T = (S^{-1})^T [L]_{\mathcal{A}}^T$$

相关联。（由于我们通常将向量表示为其坐标的列向量，所以我们使用  $[L]_{\mathcal{A}}^T$  和  $[L]_{\mathcal{B}}^T$  而不是  $[L]_{\mathcal{A}}$  和  $[L]_{\mathcal{B}}$ ）。

用文字来说，

如果  $S$  是  $X$  中的坐标变换矩阵（从旧坐标到新坐标），那么对偶空间  $X'$  中的坐标变换矩阵是  $(S^{-1})^T$ 。

因此，对偶空间  $V'$  虽然与  $V$  同构，但实际上指向不同的对象：区别在于当改变  $V$  中的基时， $V$  和  $V'$  中的坐标如何变化。

**注记** 有人可能会问：为什么我们不能在  $X$  中选择一组基，而在对偶空间  $X'$  中选择一组与之完全不相关的基呢？当然，我们可以这样做，但试想一下，如果我们知道  $\mathbf{x}$  在某组基下的坐标，以及  $L$  在另外一组与之完全不相关的基下的坐标，那该如何计算  $L(\mathbf{x})$ ？

因此，如果我们想（已知向量  $\mathbf{x}$  在某组基下的坐标）用矩阵代数的标准规则来计算线性泛函  $L$  的作用，即用行（泛函）乘以列（向量），那么我们别无选择：线性泛函  $L$  的“坐标”应该是它在（相同基下的）矩阵中对应的项。事实上，我们稍后会在下面第 1.3 节看到，线性泛函的项（“坐标”）确实是某些基（所谓的**对偶基**(dual basis)）下的坐标。

## 1.1.2. 唯一性定理

**引理 1.3.** 设  $\mathbf{v} \in V$ . 如果对所有  $L \in V'$  都有  $L(\mathbf{v}) = 0$ , 那么  $\mathbf{v} = \mathbf{0}$ . 其推论是, 如果对所有  $L \in V'$  都有  $L(\mathbf{v}_1) = L(\mathbf{v}_2)$ , 那么  $\mathbf{v}_1 = \mathbf{v}_2$ .

**证明:** 固定  $V$  中的一组基  $\mathcal{B}$ . 则

$$L(\mathbf{v}) = [L]_{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

通过选取不同的矩阵 (即不同的  $L$ ), 我们可以轻易看出  $[\mathbf{v}]_{\mathcal{B}} = \mathbf{0}$ . 的确, 如果

$$L_k = [0, \dots, 0, \frac{1}{k}, 0, \dots, 0],$$

那么等式

$$L_k[\mathbf{v}]_{\mathcal{B}} = 0$$

暗示  $[\mathbf{v}]_{\mathcal{B}}$  的第  $k$  个坐标为 0.

对所有  $k$  使用这个等式, 我们得出  $[\mathbf{v}]_{\mathcal{B}} = \mathbf{0}$ , 所以  $\mathbf{v} = \mathbf{0}$ . □

## 1.2. 二次对偶空间

正如我们上面讨论的, 对偶空间  $V'$  是一个向量空间, 因此我们可以考虑它的对偶  $V'' = (V')'$ . 看起来我们也可以考虑  $V''$  的对偶  $V''' \dots$  以此类推. 然而, 有意义的讨论将在  $V''$  处停止, 因为

二次对偶空间  $V''$  在概念上 (即以一种自然的方式) 同构于  $V$ .

让我们解读一下这个陈述. 任何向量  $\mathbf{v} \in V$  在概念上定义了  $V'$  上的一个线性泛函  $L_{\mathbf{v}}$  (即二次对偶空间  $V''$  的一个元素), 其规则是

$$L_{\mathbf{v}}(f) = f(\mathbf{v}) \quad \forall f \in V'.$$

可以很容易地验证映射  $T: V \rightarrow V'', T\mathbf{v} = L_{\mathbf{v}}$  是一个线性变换.

注意,  $\text{Ker } T = \{\mathbf{0}\}$ . 的确, 如果  $T\mathbf{v} = \mathbf{0}$ , 那么

$$f(\mathbf{v}) = 0 \quad \forall f \in V',$$

根据上面的引理 1.3, 我们得到  $\mathbf{v} = \mathbf{0}$ .

由于  $\dim V'' = \dim V' = \dim V$ , 条件  $\text{Ker } T = \{\mathbf{0}\}$  暗示  $T$  是一个可逆变换 (同构).

这个同构  $T$  是非常自然的 (至少对数学家而言). 特别是, 它没有使用基来定义, 因此它不依赖于基的选择. 所以, 非正式地说, 我们说  $V''$  在概念上同构于  $V$ . 更严谨的陈述是, 上面描述的映射  $T$  (我们认为它是自然和概念上的) 是从  $V$  到  $V''$  的一个同构.

### 1.3. 对偶基，又称双正交基

在前面的章节中，我们多次提到将线性泛函的项看作“坐标”。但通常“坐标”是指某组基下的坐标。那么线性泛函的“坐标”真的是某组基下的坐标吗？事实证明答案是“是”，所以术语保持一致。让我们找出与  $L \in V'$  的坐标对应的基。

设  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是  $V$  中的一组基。对于  $L \in V'$ ，设  $[L]_{\mathcal{B}} = [L_1, L_2, \dots, L_n]$  是其在基  $\mathcal{B}$  下的矩阵（行向量）。考虑线性泛函  $\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n \in V'$ ，它们由

$$(1.2) \quad \mathbf{b}'_k(\mathbf{b}_j) = \delta_{k,j}$$

定义。其中  $\delta_{k,j}$  是克罗内克符号 (Kronecker delta)：

$$\delta_{k,j} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}.$$

回忆一下，一个线性变换由其在基上的作用定义，因此泛函  $\mathbf{b}'_k$  是明确定义的。

人们可以很容易地看到，泛函  $L$  可以表示为

$$L = \sum_{k=1}^n L_k \mathbf{b}'_k.$$

确实，取任意  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{b}_k \in V$ ，其在基  $\mathcal{B}$  下的坐标为  $[\mathbf{v}]_{\mathcal{B}} = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ 。根据线性和  $\mathbf{b}'_k$  的定义：

$$\mathbf{b}'_k(\mathbf{v}) = \mathbf{b}'_k \left( \sum_{j=1}^n \alpha_j \mathbf{b}_j \right) = \sum_{j=1}^n \alpha_j \mathbf{b}'_k(\mathbf{b}_j) = \alpha_k.$$

因此，

$$L(\mathbf{v}) = [L]_{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = \sum_{k=1}^n L_k \alpha_k = \sum_{k=1}^n L_k \mathbf{b}'_k(\mathbf{v}).$$

由于这个恒等式对所有  $\mathbf{v} \in V$  都成立，我们得出  $L = \sum_{k=1}^n L_k \mathbf{b}'_k$ 。

因为我们没有对  $L \in V'$  做任何假设，我们刚才已经证明了任何线性泛函  $L$  都可以表示为  $\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n$  的线性组合，所以系统  $\{\mathbf{b}'_k\}_{k=1}^n$  是生成集。

现在我们证明这个系统是线性无关的（因此它是一组基）。设  $\mathbf{0} = \sum_{k=1}^n L_k \mathbf{b}'_k$ 。那么对于任意  $j = 1, 2, \dots, n$ ，

$$0 = \mathbf{0b}_j = \left( \sum_{k=1}^n L_k \mathbf{b}'_k \right) (\mathbf{b}_j) = \sum_{k=1}^n L_k \mathbf{b}'_k(\mathbf{b}_j) = L_j,$$

所以  $L_j = 0$ 。因此，所有的  $L_k$  都为 0，并且该系统是线性无关的。

所以，系统  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  确实是  $V'$  中的一组基，并且  $[L]_{\mathcal{B}}$  的项是  $L$  相对于基  $\mathcal{B}$  的坐标。

**定义 1.4** 设  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是  $V$  中的一组基。由方程 (1.2) 唯一确定的向量组

$$\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\} \subset V'$$

被称为与  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  **对偶的 (或双正交的) 基**。

注意, 我们已经证明了基的对偶系统也是一组基。还请注意, 如果  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  是基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  的对偶系统, 那么  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  也是基  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  的对偶系统。

### 1.3.1. 抽象非正交傅里叶分解

对偶系统可用于计算基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  下向量的坐标。

设  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  是  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  的双正交系统, 设  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{b}_k$ . 那么, 正如之前所示:

$$\mathbf{b}'_j(\mathbf{v}) = \mathbf{b}'_j \left( \sum_{k=1}^n \alpha_k \mathbf{b}_k \right) = \sum_{k=1}^n \alpha_k \mathbf{b}'_j(\mathbf{b}_k) = \alpha_j \mathbf{b}'_j(\mathbf{b}_j) = \alpha_j,$$

所以  $\alpha_k = \mathbf{b}'_k(\mathbf{v})$ . 那么我们可以写成:

$$(1.3) \quad \mathbf{v} = \sum_{k=1}^n \mathbf{b}'_k(\mathbf{v}) \mathbf{b}_k.$$

换句话说,

第  $k$  个坐标 (在基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  下) 是  $\mathbf{b}'_k(\mathbf{v})$ , 其中  $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  是对偶基。

这个公式被称为  $\mathbf{v}$  的 (一个简化的) **抽象非正交傅里叶分解** (abstract non-orthogonal Fourier decomposition) (在基  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  下)。之所以这样命名, 稍后在 2.3 节中会清楚。

**注记 1.5** 设  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  和  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\}$  分别是  $X$  和  $Y$  中的基, 设  $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_m\}$  是  $\mathcal{B}$  的对偶基。那么变换  $T$  在基  $\mathcal{A}, \mathcal{B}$  下的矩阵  $[T]_{\mathcal{B}, \mathcal{A}} =: A = \{a_{k,j}\}_{k=1}^m \sum_{j=1}^n$  由下式给出:

$$a_{k,j} = \mathbf{b}'_k(T\mathbf{a}_j), \quad j = 1, 2, \dots, n, \quad k = 1, 2, \dots, m.$$

## 1.4. 对偶系统的例子

我们考虑的第一个例子是平凡的。设  $V$  为  $\mathbb{R}^n$  (或  $\mathbb{C}^n$ ), 设  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是那里的标准基。对偶空间将是  $n$  维行向量的空间, 它与  $\mathbb{R}^n$  (或复数情况下的  $\mathbb{C}^n$ ) 同构, 那里的标准基是对偶于  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  的。  $(\mathbb{R}^n)'$  (或  $(\mathbb{C}^n)'$ ) 中的标准基是将  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  分别取转置得到的  $\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_n^T$ 。

### 1.4.1. 泰勒公式

下一个例子更有趣。让我们考虑次数最多为  $n$  的多项式空间  $\mathbb{P}_n$ . 我们知道, 幂  $\{\mathbf{e}_k\}_{k=0}^n, \mathbf{e}(t) = t^n$  构成了该空间中的标准基。而这组基的对偶是什么?

这个答案可能很难猜测, 但一旦你知道了, 验证起来就非常容易。也就是说, 考虑线性泛函  $\mathbf{e}'_k \in (\mathbb{P}_n)', k = 0, 1, \dots, n$ , 它们对多项式  $p$  的作用如下:

$$\mathbf{e}'_k(p) := \frac{1}{k!} \frac{d^k}{dt^k} p(t) \Big|_{t=0} = \frac{1}{k!} p^{(k)}(0);$$

这里我们使用常规约定  $0! = 1$  和  $d^0 f / dt^0 = f$ .

由于

$$\frac{d^k}{dt^k} t^j = \begin{cases} j(j-1)\dots(j-k+1)t^{j-k}, & k \leq j \\ 0, & k > j \end{cases}$$

我们可以很容易地看出系统  $\{\mathbf{e}'_k\}_{k=0}^n$  是幂集  $\{\mathbf{e}_k\}_{k=0}^n$  的对偶。

将 (1.3) 应用于上述系统  $\{\mathbf{e}_k\}_{k=0}^n$  及其对偶, 我们得到次数最多为  $n$  的任何多项式  $p$  都可以表示为:

$$(1.4) \quad p(t) = \sum_{k=0}^n \frac{p^{(k)}(0)}{k!} t^k$$

这个公式在微积分中作为多项式的泰勒公式, 是众所周知的。更确切地说, 这是泰勒公式的一个特殊情况, 即所谓的麦克劳林公式。一般的泰勒公式

$$p(t) = \sum_{k=0}^n \frac{p^{(k)}(a)}{k!} (t-a)^k$$

可以通过对多项式  $p(\tau - a)$  应用 (1.4) 然后令  $t := \tau - a$  来得到。它也可以通过考虑幂  $(t-a)^k, k = 0, 1, \dots, n$  并以我们为  $t^k$  找到对偶系统的相同的方式来得到。<sup>46</sup>

### 1.4.2. 拉格朗日插值

我们的下一个例子涉及所谓的拉格朗日插值公式。设  $a_1, a_2, \dots, a_{n+1}$  是互不相同的点 (在  $\mathbb{R}$  或  $\mathbb{C}$  中), 设  $\mathbb{P}_n$  是次数最多为  $n$  的多项式空间。定义泛函  $\mathbf{f}_k \in \mathbb{P}'_n$  为:

$$\mathbf{f}_k(p) = p(a_k) \quad \forall p \in \mathbb{P}_n.$$

这个泛函系统的对偶是什么? 注意, 虽然证明泛函  $\mathbf{f}_k$  是线性无关的 (因此, 因为  $\dim(\mathbb{P}_n)' = \dim \mathbb{P}_n = n+1$ , 它们构成  $(\mathbb{P}_n)'$  中的一组基) 并不难, 但我们不需

<sup>46</sup>一般的泰勒公式比这里得到的用于多项式的公式包含了更多信息: 它说明任何  $n$  次可微的函数都可以用其泰勒多项式在点  $a$  附近进行近似。更重要的是, 如果函数是  $n+1$  次可微的, 它能让我们对误差进行估计。上面多项式的公式可作为一般情况的动机和起点。



要这样做。我们将直接构造对偶系统，然后就能看出系统  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{n+1}$  确实是一组基。

也就是说，让我们定义多项式  $p_k$ ,  $k = 1, 2, \dots, n+1$  为：

$$p_k(t) = \prod_{j:j \neq k} (t - a_j) / \prod_{j:j \neq k} (a_k - a_j)$$

其中乘积中的  $j$  从 1 遍历到  $n+1$ . 显然,  $p_k(a_k) = 1$ , 且如果  $j \neq k$ , 则  $p_k(a_j) = 0$ . 因此, 系统  $\{p_1, p_2, \dots, p_{n+1}\}$  确实是对  $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{n+1}\}$  的对偶。

这里有一个小细节，因为对偶系统的概念只对基定义，而我们没有证明这两个系统中的任何一个是一组基。但人们可以立即看出系统  $\{p_1, p_2, \dots, p_{n+1}\}$  是线性无关的（你能解释为什么吗？），并且由于它包含  $n+1 = \dim \mathbb{P}_n$  个向量，它是一组基。因此，泛函系统  $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{n+1}\}$  也是  $(\mathbb{P}_n)'$  对偶空间中的一组基。

**注记** 注意，我们在这并不走运，这是一个普遍现象。也就是说，正如练习 1.1 所断言的，任何拥有“对偶”系统的向量系统都必须是线性无关的。因此，构造一个对偶系统是证明线性无关性的一种方法（如果你能像上面的例子那样轻易做到，那么这种方法就很简单）。

应用公式 (1.3) 到上面的例子，我们可以看出满足

$$(1.5) \quad p(a_k) = y_k, \quad k = 1, 2, \dots, n+1$$

$\deg p \leq n$  的（唯一）多项式  $p$ ，可以由公式

$$(1.6) \quad p(t) = \sum_{k=1}^{n+1} y_k p_k(t).$$

重构。这个公式在数学中作为“拉格朗日插值公式”是众所周知的。

### 练习

1.1. 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是  $X$  中的一个向量系统，使得存在一个线性泛函系统  $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r$  满足

$$\mathbf{v}'_k(\mathbf{v}_j) = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

a) 证明系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  是线性无关的。

b) 证明如果系统  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$  不是生成集，那么“双正交”系统  $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_r$  不是唯一的。

**提示：**可能最简单的证明方法是将其扩展为基，见第二章命题 5.4.

1.2. 证明对于给定的互不相同的点  $a_1, a_2, \dots, a_{n+1}$  和值  $y_1, y_2, \dots, y_{n+1}$ （不一定互不相同），满足 (1.5) 的多项式  $p$ ,  $\deg p \leq n$ ，是唯一的。尝试使用线性代数的思想来证明，而不是你所知道的多项式知识。

## 2. 内积空间的对偶

让我们回顾一下, 任意域上的内积空间并不存在, 我们所有的内积空间都定义在实数或复数域中。

### 2.1. 里斯表示定理

**定理 2.1.** (里斯 (Riesz) 表示定理) 设  $H$  是一个内积空间。给定  $H$  上的一个线性泛函  $L$ , 则存在唯一一个向量  $\mathbf{y} \in H$  使得

$$(2.1) \quad L(\mathbf{v}) = (\mathbf{v}, \mathbf{y}) \quad \forall \mathbf{v} \in H.$$

**证明:** 在  $H$  中固定一组标准正交基  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ , 设

$$[L] = [L_1, L_2, \dots, L_n]$$

是  $L$  在这组基下的矩阵。定义向量  $\mathbf{y}$  为:

$$(2.2) \quad \mathbf{y} := \sum_{k=1}^n \bar{L}_k \mathbf{e}_k$$

其中  $\bar{L}_k$  表示  $L_k$  的复共轭。在实数空间的情况下, 共轭运算不起作用, 可以简单地忽略。

我们声称  $\mathbf{y}$  满足 (2.1)。

事实上, 取任意向量  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{e}_k$ . 那么

$$[\mathbf{v}] = [\alpha_1, \alpha_2, \dots, \alpha_n]^T,$$

并且

$$L(\mathbf{v}) = [L][\mathbf{v}] = \sum_{k=1}^n L_k \alpha_k.$$

另一方面,<sup>47</sup>

$$(\mathbf{v}, \mathbf{y}) = \sum_{k=1}^n \alpha_k \bar{L}_k = \sum_{k=1}^n \alpha_k L_k.$$

所以 (2.1) 成立。

为了证明向量  $\mathbf{y}$  是唯一的, 我们假设  $\mathbf{y}$  满足 (2.1)。那么对于  $k = 1, 2, \dots, n$ ,

$$(\mathbf{e}_k, \mathbf{y}) = L(\mathbf{e}_k) = L_k,$$

所以  $(\mathbf{y}, \mathbf{e}_k) = \bar{L}_k$ . 然后, 使用在标准正交基下的分解公式, 见第五章 2.1 节, 我们得到:

$$\mathbf{y} = \sum_{k=1}^n (\mathbf{y}, \mathbf{e}_k) \mathbf{e}_k = \sum_{k=1}^n \bar{L}_k \mathbf{e}_k$$

<sup>47</sup>回忆一下, 如果我们知道两个向量在标准正交基下的坐标, 我们就可以通过取这些坐标并计算  $\mathbb{C}^n$  (或  $\mathbb{R}^n$ ) 中的标准内积来计算内积。

这意味着任何满足 (2.1) 的向量必须由 (2.2) 表示。□

**注记** 虽然定理的陈述不要求基，但这里提出的证明利用了  $H$  中的一组标准正交基，尽管得到的向量  $\mathbf{y}$  不依赖于基的选择。<sup>48</sup> 这个证明的一个优点是它给出了表示向量  $\mathbf{y}$  的计算公式。

## 2.2. 内积空间自身是其对偶吗？

对于内积空间  $H$  中的一个向量  $\mathbf{y}$ ，可以通过

$$L_{\mathbf{y}}(\mathbf{v}) := (\mathbf{v}, \mathbf{y})$$

定义一个线性泛函。很容易看出映射  $\mathbf{y} \mapsto L_{\mathbf{y}}$  是一个从  $H$  到其对偶  $H^*$  的单射映射。上面的定理 2.1 断言这个映射是一个满射，所以人们倾向于说内积空间  $H$  的对偶是空间  $H$  本身（的规范同构），其中规范同构由  $\mathbf{y} \mapsto L_{\mathbf{y}}$  给出。

这对于**实**内积空间  $H$  确实是如此，并且很容易证明映射  $\mathbf{y} \mapsto L_{\mathbf{y}}$  是一个**线性**变换。我们已经讨论过这个映射是单射和满射，所以它是一个可逆线性变换，即**同构**。

然而，如果  $H$  是一个**复**空间，则需要更加谨慎。即，映射  $\mathbf{y} \mapsto L_{\mathbf{y}}$ ，它将向量  $\mathbf{y} \in H$  映射到线性泛函  $L_{\mathbf{y}}$ ，如定理 2.1 所示 ( $L_{\mathbf{y}}(\mathbf{v}) = (\mathbf{v}, \mathbf{y})$ )，就不是线性的。更准确地说，虽然很容易证明：

$$(2.3) \quad L_{\mathbf{y}_1 + \mathbf{y}_2} = L_{\mathbf{y}_1} + L_{\mathbf{y}_2},$$

然而，从  $L_{\mathbf{y}}$  的定义和内积的性质可得：

$$(2.4) \quad L_{\alpha \mathbf{y}}(\mathbf{v}) = (\mathbf{v}, \alpha \mathbf{y}) = \bar{\alpha}(\mathbf{v}, \mathbf{y}) = \bar{\alpha} L_{\mathbf{y}}(\mathbf{v}),$$

所以  $L_{\alpha \mathbf{y}} = \bar{\alpha} L_{\mathbf{y}}$ 。

换句话说，我们可以说，一个复数内积空间的对偶是该空间本身，但具有**不同的线性结构**：两个向量的加法等同于相应线性泛函的加法，但一个向量乘以  $\alpha$  等同于相应泛函乘以  $\bar{\alpha}$ 。

一个满足  $T(\alpha \mathbf{x} + \beta \mathbf{y}) = \bar{\alpha} T\mathbf{x} + \bar{\beta} T\mathbf{y}$  的变换有时被称为**共轭线性**变换。

因此，对于复数内积空间  $H$ ，其对偶可以通过一个共轭线性同构（即可逆共轭线性变换）与  $H$  规范对应。

当然，对于实内积空间，复共轭可以简单地忽略（因为  $\alpha$  是实数），所以映射  $\mathbf{y} \mapsto L_{\mathbf{y}}$  是线性的。在这种情况下，我们确实可以说内积空间  $H$  的对偶就是其本身。

<sup>48</sup> 另一种需要基的证明也是可能的。这个替代的证明（在无限维情况下有效）利用了单位球在内积空间中的强凸性，以及来自分析学的完备性思想。

在实数和复数两种情况下，我们仍然可以认为内积空间  $H$  的对偶可以规范对应为空间  $H$  本身。

### 2.3. 双正交系统与标准正交基

**定义 2.2** 设  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是内积空间  $H$  中的一组基。在  $H$  中由

$$(\mathbf{b}_j, \mathbf{b}'_k) = \delta_{j,k},$$

定义的唯一系统  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$ ，其中  $\delta_{j,k}$  是克罗内克符号，被称为与基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  **双正交或对偶**。

这个定义显然与定义 1.4 一致，如果我们像上面讨论的那样将对偶  $H'$  与  $H$  对应。那么从 1.3 节的讨论中可以立即得出，基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  的对偶系统  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  是唯一确定的，并且构成一组基，并且  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  的对偶是  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ 。

抽象非正交傅里叶分解公式 (1.3) 可以重写为：

$$\mathbf{v} = \sum_{k=1}^n (\mathbf{v}, \mathbf{b}'_k) \mathbf{b}_k.$$

注意，一组标准正交基是它自身的对偶。所以，如果  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  是一组标准正交基，那么上面的公式可重写为：

$$\mathbf{v} = \sum_{k=1}^n (\mathbf{v}, \mathbf{e}_k) \mathbf{e}_k,$$

这就是经典的（正交的）抽象傅里叶分解，见第五章 2.1 节公式 (2.2)。

### 3. 对偶（伴随）变换与转置，基本子空间再回顾（又一次）

类比内积空间的情况，（见定理 2.1），通常将  $L(\mathbf{v})$  写成类似于内积的形式，其中  $L$  是一个线性泛函（即  $L \in V', \mathbf{v} \in V$ ）：

$$L(\mathbf{v}) = \langle \mathbf{v}, L \rangle$$

注意，表达式  $\langle \mathbf{v}, L \rangle$  在两个参数上都是线性的，这与内积不同，后者在复数情况下是第一个参数为线性的，第二个参数为共轭线性的。所以，为了区分它与内积，我们使用尖括号。<sup>49</sup>

还请注意，在内积中两个向量属于同一空间，而上面的  $\mathbf{v}$  和  $L$  属于不同的空间：特别地，我们不能将它们相加。

<sup>49</sup>这个记号虽然被广泛使用，但并不标准。有时也使用  $(\mathbf{v}, L)$ ，有时尖括号用于内积。因此，在文本中遇到这样的表达式时，必须非常小心地区分线性泛函的作用与内积。

### 3.1. 对偶（伴随）变换

**定义 3.1** 设  $A: X \rightarrow Y$  是一个线性变换。变换  $A': Y' \rightarrow X'$ （其中  $X'$  和  $Y'$  分别是  $X$  和  $Y$  的对偶空间），使得

$$\langle Ax, y' \rangle = \langle x, A'y' \rangle \quad \forall x \in X, y' \in Y'$$

被称为  $A$  的**对偶（伴随）变换**。

当然，事先我们并不清楚为什么变换  $A'$  存在。下面我们将表明，确实存在这样的变换，而且它是唯一的。

#### 3.1.1. $A: \mathbb{F}^n \rightarrow \mathbb{F}^m$ 情况下的对偶变换

让我们首先考虑  $X = \mathbb{F}^n, Y = \mathbb{F}^m$  的情况（这里的  $\mathbb{F}$  通常是  $\mathbb{R}$  或  $\mathbb{C}$ ，但一切都适用于任意域）。

像往常一样，我们将  $\mathbb{F}^n$  中的向量  $\mathbf{v}$  与其坐标列向量进行对应，并将线性变换与其矩阵（在标准基下）进行对应。

如上所述， $\mathbb{F}^n$  的对偶是大小为  $n$  的行向量空间，所以我们可以将其与  $\mathbb{F}^n$  进行对应。同样，我们将  $(\mathbb{F}^n)'$  中的元素视为其坐标的列向量。

在这些约定下，我们对于  $\mathbf{x} \in \mathbb{F}^n$  和  $\mathbf{x}' \in (\mathbb{F}^n)'$  有：

$$\mathbf{x}'(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x}' \rangle = (\mathbf{x}')^T \mathbf{x}$$

其中右侧是矩阵乘法（或行向量乘以列向量）。那么，对于任意  $\mathbf{x} \in X = \mathbb{F}^n$  和  $\mathbf{y}' \in Y' = (\mathbb{F}^m)'$ ,

$$\langle A\mathbf{x}, \mathbf{y}' \rangle = (\mathbf{y}')^T A\mathbf{x} = (A^T \mathbf{y}')^T \mathbf{x} = \langle \mathbf{x}, A^T \mathbf{y}' \rangle.$$

（中间的表达式是矩阵乘法）。

所以我们已经证明了对偶变换存在。下面让我们证明它是唯一的。假设存在某个变换  $B$  使得

$$\langle A\mathbf{x}, \mathbf{y}' \rangle = \langle \mathbf{x}, B\mathbf{y}' \rangle \quad \forall \mathbf{x} \in \mathbb{F}^n, \forall \mathbf{y}' \in (\mathbb{F}^m)'$$

这意味着对于任意  $\mathbf{x}$  和  $\mathbf{y}'$ ,

$$\langle \mathbf{x}, (A^T - B)\mathbf{y}' \rangle = 0 \quad \forall \mathbf{x} \in \mathbb{F}^n, \forall \mathbf{y}' \in (\mathbb{F}^m)'$$

通过选择  $\mathbf{x}$  和  $\mathbf{y}'$  分别为  $\mathbb{F}^n$  和  $(\mathbb{F}^m)' \cong \mathbb{F}^m$  中的标准基向量，我们得到矩阵  $B$  和  $A^T$  是相同的。

所以，对于  $X = \mathbb{F}^n, Y = \mathbb{F}^m$ ,

对偶变换  $A'$  存在且唯一。而且，其矩阵（在标准基下）等于  $A^T$ （ $A$  矩阵的转置）。

### 3.1.2. 抽象设置下的对偶变换

现在, 让我们考虑一般情况。事实上, 我们不需要做太多, 因为一切都可以归约到  $\mathbb{F}^n$  的情况。

也就是说, 我们固定  $X$  中的基  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  和  $Y$  中的基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\}$ , 以及它们对应的对偶基  $\mathcal{A}' = \{\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_n\}$  和  $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_m\}$  (分别在  $X'$  和  $Y'$  中)。对于一个向量  $\mathbf{v}$  (来自空间或其对偶), 我们像往常一样用  $[\mathbf{v}]_{\mathcal{B}}$  表示其在基  $\mathcal{B}$  下的坐标。那么

$$\langle \mathbf{x}, \mathbf{x}' \rangle = ([\mathbf{x}']_{\mathcal{A}'})^T [\mathbf{x}]_{\mathcal{A}}, \quad \forall \mathbf{x} \in X, \forall \mathbf{x}' \in X',$$

也就是说, 与  $\mathbf{x} \in X$  和  $\mathbf{x}' \in X'$  上的作用相比, 我们可以使用它们坐标的列向量, 以绝对相同的方式操作, 就像在  $\mathbb{F}^n$  的情况下一样。当然, 对于  $Y$  也是如此, 所以通过使用坐标列向量然后将一切平移回抽象设置, 我们得到在这种情况下对偶变换也存在且唯一。而且, 利用 (我们刚刚证明的) 对于  $A: \mathbb{F}^n \rightarrow \mathbb{F}^m$ ,  $A'$  的矩阵是  $A^T$  的事实, 我们得到:

$$(3.1) \quad [A']_{\mathcal{A}', \mathcal{B}'} = ([A]_{\mathcal{B}, \mathcal{A}})^T,$$

或者用通俗的语言说:

对偶变换在对偶基下的矩阵是变换在原始基下的矩阵的转置。

**注记 3.2** 请注意, 虽然我们使用基来构造对偶变换, 但得到的变换不依赖于基的选择。

### 3.1.3. 定义对偶变换的无坐标方法

现在, 让我们提出另一种更“高深”的方法来定义线性变换的对偶。也就是说, 对于  $\mathbf{x} \in X, \mathbf{y}' \in Y'$ , 让我们暂时固定  $\mathbf{y}'$ , 并将表达式  $\langle A\mathbf{x}, \mathbf{y}' \rangle = \mathbf{y}'(A\mathbf{x})$  看作是  $\mathbf{x}$  的函数。很容易看出这是两个线性变换 (哪两个?) 的复合, 因此它是  $\mathbf{x}$  的线性函数, 即  $X$  上的一个线性泛函, 即  $X'$  中的一个元素。

让我们称这个线性泛函为  $B(\mathbf{y}')$ , 以强调它依赖于  $\mathbf{y}'$ . 由于我们可以对每个  $\mathbf{y}' \in Y'$  执行此操作, 则可定义一个变换  $B: Y' \rightarrow X'$  使得

$$\langle A\mathbf{x}, \mathbf{y}' \rangle = \langle \mathbf{x}, B(\mathbf{y}') \rangle$$

我们的下一步是证明  $B$  是一个线性变换。请注意, 由于变换  $B$  是以一种相当间接的方式定义的, 我们无法立即从定义中看出它是线性的。为了证明  $B$  的线性, 让

我们取  $\mathbf{y}'_1, \mathbf{y}'_2 \in Y'$ . 对于  $\mathbf{x} \in X$ :

$$\begin{aligned}\langle \mathbf{x}, B(\alpha \mathbf{y}'_1 + \beta \mathbf{y}'_2) \rangle &= \langle A\mathbf{x}, \alpha \mathbf{y}'_1 + \beta \mathbf{y}'_2 \rangle \quad (\text{根据 } B \text{ 的定义}) \\ &= \alpha \langle A\mathbf{x}, \mathbf{y}'_1 \rangle + \beta \langle A\mathbf{x}, \mathbf{y}'_2 \rangle \quad (\text{根据线性性质}) \\ &= \alpha \langle \mathbf{x}, B(\mathbf{y}'_1) \rangle + \beta \langle \mathbf{x}, B(\mathbf{y}'_2) \rangle \quad (\text{根据 } B \text{ 的定义}) \\ &= \langle \mathbf{x}, \alpha B(\mathbf{y}'_1) + \beta B(\mathbf{y}'_2) \rangle \quad (\text{根据线性性质})\end{aligned}$$

由于这个恒等式对所有  $\mathbf{x}$  都成立, 我们得出  $B(\alpha \mathbf{y}'_1 + \beta \mathbf{y}'_2) = \alpha B(\mathbf{y}'_1) + \beta B(\mathbf{y}'_2)$ , 即  $B$  是线性的。

这种方法的主要优点是它不需要基, 因此可以 (并且被) 用于无限维情况。然而, 我们在 3.1.1 和 3.1.2 节中给出的证明提供了一种构造性计算对偶变换的方法, 所以我们使用了那个证明而不是更通用的无坐标证明。

**注记 3.3** 注意, 上面的无坐标方法可以用来定义内积空间中算子的埃尔米特伴随。与上面呈现的推理相比, 唯一需要再添加的是使用里斯表示定理 (定理 2.1)。我们把细节留给读者作为练习, 见下面的练习 3.2。

### 3.2. 零化子与基本子空间之间的关系

**定义 3.4** 设  $X$  是一个向量空间, 设  $E \subset X$ .  $E$  的**零化子** (annihilator), 记作  $E^\perp$ , 是所有  $\mathbf{x}' \in X'$  的集合, 使得  $\langle \mathbf{x}, \mathbf{x}' \rangle = 0$  对所有  $\mathbf{x} \in E$ .

利用  $X''$  与  $X$  规范同构的这一事实 (见 1.2 节), 我们说对于  $E \subset X'$ , 其**零化子**  $E^\perp$  由所有满足  $\langle \mathbf{x}, \mathbf{x}' \rangle = 0$  对所有  $\mathbf{x}' \in E$  的向量  $\mathbf{x} \in X$  组成。

**注记 3.5** 严格来说, 对于  $E \subset X'$ , 集合  $E^\perp$  应该定义为所有  $\mathbf{x}'' \in X''$  的集合, 使得  $\langle \mathbf{x}', \mathbf{x}'' \rangle = 0$  对所有  $\mathbf{x}' \in E$ . 符号  $E^\perp$  通常用于定义 3.4 的第二部分中的零化子。然而, 由于  $X''$  和  $X$  之间的自然同构, 这两种情况之间没有真正的区别, 所以我们总是使用  $E^\perp$ .

区分  $E \subset X$  和  $E \subset X'$  的情况在无限维情况下非常有意义, 其中  $X''$  并不总是与  $X$  规范同构。

满足  $X''$  与  $X$  规范同构的空间被称为**自反空间** (reflexive space)。

**命题 3.6** 设  $E$  是  $X$  的一个子空间。那么  $(E^\perp)^\perp = E$ .

这个命题看起来完全像第五章命题 3.6。然而, 它的证明有点复杂, 因为第五章命题 3.6 的建议证明严重依赖于内积空间结构: 它使用了分解  $X = E \oplus E^\perp$ , 这在现在的情况下是不成立的, 因为例如,  $E$  和  $E^\perp$  属于不同的空间。

**证明:** 设  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$  是  $E$  的一组基 (回忆一下本章中的所有空间都是有限维的), 所以  $E = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ .

根据第二章命题 5.4, 该系统可以扩展为  $X$  的一组基, 也就是说, 我们可以找到向量  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$  ( $n = \dim X$ ), 使得  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  是  $X$  的一组基。

设  $\{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n\}$  是与  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  对偶的基。根据问题 3.3,  $E^\perp = \text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\}$ . 再次将这个问题应用于  $E^\perp$ , 我们得到

$$(E^\perp)^\perp = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = E.$$

□

下面的定理与第五章定理 5.1 很类似。

**定理 3.7.** 设  $A: X \rightarrow Y$  是从一个向量空间到另一个向量空间的算子。那么:

- a)  $\text{Ker } A' = (\text{Ran } A)^\perp$ ;
- b)  $\text{Ker } A = (\text{Ran } A')^\perp$ ;
- c)  $\text{Ran } A = (\text{Ker } A')^\perp$ ;
- d)  $\text{Ran } A' = (\text{Ker } A)^\perp$ .

**证明:** 首先, 让我们注意到, 由于对于子空间  $E$ , 我们有  $(E^\perp)^\perp = E$ , 所以命题 1 和 3, 命题 2 和 4 是等价的。最后, 命题 2 正是将算子  $A$  替换为  $A'$  的命题 1 (我们使用了  $(A')' = A$  的平凡事实, 例如转置的相应性质)。

因此, 为了证明定理, 我们只需要证明命题 1。

回忆一下,  $A': Y' \rightarrow X'$ . 包含  $\mathbf{y}' \in (\text{Ran } A)^\perp$  意味着  $\mathbf{y}'$  零化了所有形式为  $A\mathbf{x}$  的向量, 即

$$\langle A\mathbf{x}, \mathbf{y}' \rangle = 0 \quad \forall \mathbf{x} \in X.$$

由于  $\langle A\mathbf{x}, \mathbf{y}' \rangle = \langle \mathbf{x}, A'\mathbf{y}' \rangle$ , 最后一个恒等式等价于

$$\langle \mathbf{x}, A'\mathbf{y}' \rangle = 0 \quad \forall \mathbf{x} \in X.$$

但这表示  $A'\mathbf{y}' = \mathbf{0}$  ( $A'\mathbf{y}'$  是零泛函)。

所以我们证明了  $\mathbf{y}' \in (\text{Ran } A)^\perp$  当且仅当  $A'\mathbf{y}' = \mathbf{0}$ , 或者等价地, 当且仅当  $\mathbf{y}' \in \text{Ker } A'$ . □

### 练习

3.1. 证明如果对于线性变换  $T, T_1: X \rightarrow Y$ ,

$$\langle T\mathbf{x}, \mathbf{y}' \rangle = \langle T_1\mathbf{x}, \mathbf{y}' \rangle$$

对所有  $\mathbf{x} \in X$  和所有  $\mathbf{y}' \in Y'$  成立, 那么  $T = T_1$ .

也许最简单的证明方法是使用引理 1.3。

3.2. 结合里斯表示定理 (定理 2.1) 和上面 3.1.3 节的推理, 给出一个内积空间中算子的埃尔米特伴随的无坐标定义。

下一个问题给出了证明命题 3.6 的一种方法。



3.3. 设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  是  $X$  中的一组基, 设  $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n$  是它的对偶基. 设  $E := \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ ,  $r < n$ . 证明  $E^\perp = \text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\}$ .

3.4. 使用前一个问题来证明对于子空间  $E \subset X$ ,

$$\dim E + \dim E^\perp = \dim X.$$

## 4. 空间与其对偶之间的区别

我们知道对偶空间  $X'$  与  $X$  具有相同的维度, 所以空间与其对偶是同构的. 因此, 人们可能会认为空间与其对偶之间实际上没有区别. 然而, 正如我们在 1.1 节中讨论过的, 当我们在空间  $X$  中改变基时,  $X$  中的坐标和  $X'$  中的坐标根据不同的规则变化, 见上面公式 (1.1).

另一方面, 利用  $X$  和  $X''$  的自然同构, 我们可以说  $X$  是  $X'$  的对偶. 从这个角度来看,  $X$  和  $X'$  之间没有区别: 我们可以从  $X$  开始, 说  $X'$  是它的对偶, 或者我们可以反过来, 从  $X'$  开始.

我们已经在上面使用了这种观点, 例如在定理 3.7 的证明中.

还请注意, 坐标变换公式 (1.1) (也见它下方的方框语句) 与这种观点一致: 如果  $\tilde{S} := (S^{-1})^T$ , 那么  $(\tilde{S}^{-1})^T = S$ , 所以我们通过相同的规则从  $X'$  中的坐标变换公式得到了  $X$  中的坐标变换公式!

### 4.1. $X$ 与 $X'$ 之间的同构

定义  $X$  和  $X'$  之间的同构存在无数种可能性.

如果  $X = \mathbb{F}^n$ , 那么最自然地对应  $X$  和  $X'$  的方法是将  $\mathbb{F}^n$  中的标准基与  $(\mathbb{F}^n)'$  中的标准基进行对应. 在这种情况下, 线性泛函的作用将由“内积类型”的表达式

$$\langle \mathbf{v}, \mathbf{v}' \rangle = (\mathbf{v}')^T \mathbf{v}$$

给出. 为了将其推广到一般情况, 必须固定  $X$  中的一组基  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  并考虑其对偶基  $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$ , 并定义一个同构  $T: X \rightarrow X'$  为  $T\mathbf{b}_k = \mathbf{b}'_k$ ,  $k = 1, 2, \dots, n$ .

该同构在一定意义上可以被视为自然的, 然而它依赖于基的选择, 因此在一般情况下无法找到自然的方式来对应  $X$  和  $X'$ .

唯一的例外是当  $X$  是一个实内积空间时: 里斯表示定理 (定理 2.1) 提供了一种自然的方式将线性泛函与  $X$  中的向量进行对应. 请注意, 这种方法仅适用于实内积空间. 对于复数情况, 里斯表示定理给出了  $X$  和  $X'$  的自然对应, 但这种对应不是线性的, 而是共轭线性的.

## 4.2. 例子：速度（微分算子），微分形式作为向量，线性泛函

为了说明向量和线性泛函之间的关系，让我们考虑一个来自多变量微积分的例子，它引出了微分几何中的重要思想，如切线丛和余切线丛。

让我们回忆一下微积分中第二类路径积分的概念： $\mathbb{R}^n$  中的一条路径  $\gamma$  由其参数化定义，即由一个从区间  $[a, b]$  到  $\mathbb{R}^n$  的函数

$$t \mapsto \mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))^T.$$

如果  $\omega$  是所谓的**微分形式**（一阶微分形式），

$$\omega = f_1(\mathbf{x}) dx_1 + f_2(\mathbf{x}) dx_2 + \cdots + f_n(\mathbf{x}) dx_n,$$

那么**路径积分**

$$\int_{\gamma} \omega = \int_{\gamma} f_1 dx_1 + f_2 dx_2 + \cdots + f_n dx_n$$

是通过将  $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$  代入表达式来计算的，即  $\int_{\gamma} \omega$  计算为：

$$\int_a^b \left( f_1(\mathbf{x}(t)) \frac{dx_1(t)}{dt} + f_2(\mathbf{x}(t)) \frac{dx_2(t)}{dt} + \cdots + f_n(\mathbf{x}(t)) \frac{dx_n(t)}{dt} \right) dt.$$

换句话说，在每个时刻  $t$ ，我们必须计算速度

$$\mathbf{v} = \frac{d\mathbf{x}(t)}{dt} = \left( \frac{dx_1(t)}{dt}, \frac{dx_2(t)}{dt}, \dots, \frac{dx_n(t)}{dt} \right)^T,$$

将其应用于线性泛函  $\mathbf{f} = (f_1, f_2, \dots, f_n)$ ， $\mathbf{f}(\mathbf{v}) = \sum_{k=1}^n f_k v_k$ （这里  $f_k = f_k(\mathbf{x}(t))$ ）但对于固定的  $t$ ，每个  $f_k$  只是一个数字，所以我们只写  $f_k$ ），然后将结果（它依赖于  $t$ ）对  $t$  进行积分。

### 4.2.1. 速度作为向量

让我们固定  $t$  并分析  $\mathbf{f}(\mathbf{v})$ . 我们将根据微积分的规则说明  $\mathbf{v}$  的坐标如何变化，以及  $\mathbf{f}$  的坐标如何变化。

假设正如微积分中的惯例， $x_k$  是  $\mathbb{R}^n$  标准基下的坐标，设  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是  $\mathbb{R}^n$  中的另一组基。我们将使用记号  $\tilde{x}_k$  来表示向量  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  的坐标，即  $[\mathbf{x}]_{\mathcal{B}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)^T$ 。

设  $A = \{a_{k,j}\}_{k,j=1}^n$  是坐标变换矩阵， $A = [I]_{\mathcal{B}, \mathcal{S}}$ ，所以新的坐标  $\tilde{x}_k$  用旧的坐标  $x_j$  表示为：

$$\tilde{x}_k = \sum_{j=1}^n a_{k,j} x_j, \quad k = 1, 2, \dots, n.$$

所以向量  $\mathbf{v}$  的新坐标  $\tilde{v}_k$  是从其旧坐标  $v_k$  得到的：

$$\tilde{v}_k = \sum_{j=1}^n a_{k,j} v_j, \quad k = 1, 2, \dots, n.$$

## 4.2.2. 微分形式作为线性泛函（余向量）

现在让我们用新的坐标  $\tilde{x}_k$  来计算微分形式

$$(4.1) \quad \omega = \sum_{k=1}^n f_k dx_k.$$

从旧坐标到新坐标的坐标变换矩阵是  $A^{-1}$ . 设  $A^{-1} = \{\tilde{a}_{k,j}\}_{k,j=1}^n$ , 所以

$$x_k = \sum_{j=1}^n \tilde{a}_{k,j} \tilde{x}_j, \text{ 并且 } dx_k = \sum_{j=1}^n \tilde{a}_{k,j} d\tilde{x}_j, \quad k = 1, 2, \dots, n.$$

将此代入 (4.1) 中, 我们得到:

$$\begin{aligned} \omega &= \sum_{k=1}^n f_k \sum_{j=1}^n \tilde{a}_{k,j} d\tilde{x}_j \\ &= \sum_{j=1}^n \left( \sum_{k=1}^n \tilde{a}_{k,j} f_k \right) d\tilde{x}_j \\ &= \sum_{j=1}^n \tilde{f}_j d\tilde{x}_j, \end{aligned}$$

其中

$$\tilde{f}_j = \sum_{k=1}^n \tilde{a}_{k,j} f_k.$$

但这正是对偶空间中坐标的变换规则! 所以

根据微积分的规则, 一阶微分形式的系数按照与对偶空间中的坐标相同的规则进行变换。

因此, 根据被接受的微积分的规则, 速度  $\mathbf{v}$  的坐标像向量的坐标一样变化, 而一阶微分形式的系数 (坐标) 像线性泛函的系数一样变化。在微分几何中, 所有“速度”的集合被称为**切空间 (tangent space)**, 而所有一阶微分形式的集合是其对偶, 被称为**余切空间 (cotangent space)**。

## 4.2.3. 微分算子作为向量

如前所述, 在微分几何中, 向量通常用“速度”来表示, 即通过导数  $d\mathbf{x}(t)/dt$  的形式来体现。这种观点简单且直观, 但有时被认为有点初等。

一种更“高深”的观点, 在微分几何中 (和其他更高级的文本中) 也有使用, 即用**微分算子**表示向量:

$$(4.2) \quad D = \sum_{k=1}^n v_k \frac{\partial}{\partial x_k}.$$

我们选择非正式表示的原因是, 假设我们想沿着由函数  $t \mapsto \mathbf{x}(t)$  给出的路径计算函数  $\Phi$  的导数, 即导数

$$\frac{d\Phi(\mathbf{x}(t))}{dt}.$$

根据链式法则, 在给定时间  $t$ :

$$\frac{d\Phi(\mathbf{x}(t))}{dt} = \sum_{k=1}^n \left( \frac{\partial \Phi}{\partial x_k} \Big|_{\mathbf{x}=\mathbf{x}(t)} \right) \mathbf{x}'_k(t) = D\Phi \Big|_{\mathbf{x}=\mathbf{x}(t)},$$

其中微分算子  $D$  由 (4.2) 给出,  $v_k = x'_k(t)$ .

当然, 我们需要根据向量坐标的坐标变换规则来表明微分形式的系数  $\mathbf{v}_k$  如何变化。这是直观清晰的, 并且可以通过使用多变量链式法则轻松证明。我们将其留给读者作为练习, 见下面的问题 4.1.

### 4.3. 实内积空间的情况

正如我们上面已经讨论过的, 根据里斯表示定理 (定理 2.1), 实内积空间  $X$  及其对偶  $X'$  是规范同构的。因此, 我们可以说向量和泛函存在于同一个空间中, 这将使得事情变得既更简单也更混乱。

**注记** 首先, 让我们注意到, 如果坐标变换矩阵  $S$  是正交的 ( $S^{-1} = S^T$ ), 那么  $(S^{-1})^T = S$ . 因此, 对于正交坐标变换矩阵, 向量和线性泛函的坐标根据相同的规则变化, 所以人们无法真正区分向量和泛函。

坐标变换矩阵是正交的, 例如, 当我们从一组标准正交基改变到另一组标准正交基时。

#### 4.3.1. 爱因斯坦求和约定, 度量张量

设  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  是实内积空间  $X$  中的一组基, 并设  $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$  是它的对偶基 (我们通过里斯表示定理将对偶空间  $X'$  与  $X$  对应, 所以  $\mathbf{b}'_k$  可以在  $X$  中)。

这里我们介绍微分几何中处理这些基下坐标的标准记法 (所谓的**爱因斯坦求和约定**(Einstein notation)<sup>50</sup>)。由于我们只处理坐标, 因此可以假设我们在  $\mathbb{R}^n$  空间中, 其非标准内积为  $(\cdot, \cdot)_G$ , 由正定矩阵  $G = \{g_{j,k}\}_{j,k=1}^n$  定义, 其中  $g_{j,k} = (\mathbf{b}_k, \mathbf{b}_j)_X$ , 该矩阵通常被称为**度量张量** (metric tensor)。

$$(4.3) \quad (\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})_G = \sum_{j=1}^n \sum_{k=1}^n g_{j,k} x_j y_k, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

(见第七章第 5 节)。

为了区分向量和线性泛函 (余向量), 约定将向量的坐标写成上标, 线性泛函的坐标写成下标: 因此  $x^j, j = 1, 2, \dots, n$  表示向量  $\mathbf{x}$  的坐标, 而  $f_k, k = 1, 2, \dots, n$  表示线性泛函  $\mathbf{f}$  的坐标。

<sup>50</sup>译者注: 直译为“爱因斯坦记法”, 但“爱因斯坦求和约定” (Einstein summation convention) 的术语更常用。

**注记** 将下标写成上标可能会令人困惑，因为需要将其与幂区分开。然而，这是一个标准且广泛使用的记法，所以我们需要熟悉它。虽然我个人，以及许多数学家，更喜欢使用无坐标记法，但所有最终的计算都是在坐标中进行的，所以坐标记法必须被使用。而且你将会发现坐标记法在具体处理时相当方便。

爱因斯坦求和约定的另一个约定是，每当乘积中出现相同的下标和上标，就意味着需要对该下标进行求和。因此， $x^j f_j$  表示  $\sum_j x^j f_j$ ，所以我们可以写成  $\mathbf{f}(\mathbf{x}) = x^j f_j$ 。同样的约定适用于我们有多多个求和下标的情况，所以 (4.3) 可以重写为：

$$(4.4) \quad (\mathbf{x}, \mathbf{y}) = g_{j,k} x^k y^j \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

(数学家很懒，总是试图避免编额外的符号，只要他们能做到)。

最后，爱因斯坦求和约定的最后一个约定是**位置的保持**：如果我们不对某个下标求和，它将保持与之前相同的位置（下标或上标）。因此，我们可以写  $y^j = a_k^j x^k$ ，但不能写  $f_j = a_k^j x^k$ ，因为下标  $j$  必须保持为下标。

注意，为了计算两个向量的内积，仅知道它们的坐标是不够的。你还需要知道矩阵  $G$ （通常称为度量张量）。这与爱因斯坦求和约定一致：如果我们试图将  $(\mathbf{x}, \mathbf{y})$  写成标准内积，那么表达式  $x^k y_k$  仅仅意味着它是坐标的乘积，因为为了求和，我们需要相同的索引指标，它同时作为下标和上标。另一方面，表达式 (4.4) 完全符合这个约定。

#### 4.3.2. 协变和逆变坐标 上/下指标的升降

让我们回忆一下，在实内积空间中，我们有一组基  $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ ，以及它的对偶基  $\{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n\}$ ， $\mathbf{b}'_k \in X$ （我们通过里斯表示定理将对偶空间  $X'$  与  $X$  对应，所以  $\mathbf{b}'_k$  可以在  $X$  中）。给定向量  $\mathbf{x} \in X$ ，它可以表示为：

$$(4.5) \quad \mathbf{x} = \sum_{k=1}^n (\mathbf{x}, \mathbf{b}'_k) \mathbf{b}_k =: \sum_{k=1}^n x^k \mathbf{b}_k,$$

以及：

$$(4.6) \quad \mathbf{x} = \sum_{k=1}^n (\mathbf{x}, \mathbf{b}_k) \mathbf{b}'_k =: \sum_{k=1}^n x_k \mathbf{b}'_k.$$

坐标  $x_k$  被称为向量  $\mathbf{x}$  的**协变** (covariant) 坐标，而坐标  $x^k$  被称为**逆变** (contravariant) 坐标。

现在问自己一个问题：如何从逆变坐标  $x^k$  得到向量的协变坐标  $x_k$ ？

根据爱因斯坦求和约定，我们使用逆变坐标来处理向量，而协变坐标用于线性泛函（即当我们把向量  $\mathbf{x} \in X$  解释为线性泛函时）。我们知道  $x_k = (\mathbf{x}, \mathbf{b}_k)$ （见

(4.6)), 因此:

$$x_k = (\mathbf{x}, \mathbf{b}_k) = \left( \sum_j x^j \mathbf{b}_j, \mathbf{b}_k \right) = \sum_j x^j (\mathbf{b}_j, \mathbf{b}_k) = \sum_j g_{k,j} x^j$$

或者用爱因斯坦求和约定:

$$x_k = g_{k,j} x^j$$

换句话说,

度量张量  $G$  是从逆变坐标  $x^j$  到协变坐标  $x_k$  的变换矩阵。

从逆变坐标获得协变坐标的操作被称为**指标的降** (lowering of the indices)。

注意公式 (4.4) 对于内积的解释: 正如我们所知, 对于向量  $\mathbf{x}$ , 我们得到其协变坐标为  $x_j = g_{j,k} x^k$ . 因此,  $(\mathbf{x}, \mathbf{y}) = x_j y^j$ . 类似地, 由于  $G$  是对称的, 我们可以说  $y_k = g_{j,k} y^j$  并且  $(\mathbf{x}, \mathbf{y}) = x^k y_k$ . 换句话说,

为了计算两个向量的内积, 首先需要使用度量张量  $G$  来降低一个向量的指标, 然后将其视为一个泛函, 计算它在另一个向量上的值.

当然, 我们也可以从协变坐标  $x_j$  变到逆变坐标  $x^j$  (**指标的升**)。由于

$$(x_1, x_2, \dots, x_n)^T = G(x^1, x^2, \dots, x^n)^T,$$

我们得到

$$(x^1, x^2, \dots, x^n)^T = G^{-1}(x_1, x_2, \dots, x_n)^T,$$

所以这种情况下的坐标变换矩阵是  $G^{-1}$ .

我们知道, 由于坐标变换矩阵就是度量张量, 我们可以立即得出  $G^{-1}$  是协变 (坐标下的) 度量张量, 即如果  $G^{-1} = \{g^{k,j}\}_{j,k=1}^n$ , 那么

$$(\mathbf{x}, \mathbf{y}) = g^{k,j} x_j y_k.$$

**注记** 注意, 如果从宏观角度来看, 协变和逆变坐标是完全可互换的。这仅仅取决于我们选择哪一组基作为“主要”基, 哪一组作为“对偶”基。

选择什么作为“主要”对象, 什么作为“对偶”对象, 主要取决于公认的约定。

**注记 4.1** 爱因斯坦求和约定通常用于微分几何, 特别是黎曼几何, 其中向量被对应为速度, 而余向量 (线性泛函) 被对应为一阶微分形式, 见上面 4.2 节。这里的向量和余向量是明显不同的对象, 构成了所谓的**切空间**和**余切空间**。

在黎曼几何中, 我们接下来可以在切空间上引入内积 (即度量张量, 如果从坐标的角度来看), 这允许我们对向量和余向量 (线性泛函)。在坐标表示中, 这种对应是通过升降指标来完成的, 如上所述。

## 4.4. 结论

让我们总结一下上面关于空间与其对偶是否不同的讨论。

简而言之，答案是“是的”，它们是不同的对象。虽然在本书所讨论的有限维情况下，它们是同构的，但将空间与其对偶进行对应通常没有什么好处。

即使在最简单的  $\mathbb{F}^n$  情况下，认为  $\mathbb{F}^n$  的元素是列向量，而其对偶的元素是行向量（尽管在处理对偶空间元素时，我们经常将行向量垂直放置）也是有用的。更显著的例子是 1.4.1 和 1.4.2 节中关于泰勒公式和拉格朗日插值的内容。在那里，你可以清楚地看到线性泛函确实与多项式是完全不同的对象，并且通过对应泛函与多项式几乎没有什么好处。

对于内积空间，情况有所不同，因为这样的空间可以**规范地**与其对偶进行对应。这种对应对于实内积空间是线性的，所以一个实内积空间与其对偶是规范同构的。对于复数空间，这种对应只是**共轭线性的**，但它仍然非常有助于将线性泛函与向量进行对应，并利用内积空间结构和正交性、自伴性、正交投影等思想。

然而，有时即使在实内积空间的情况下，考虑空间及其对偶作为不同的对象也更自然。例如，在黎曼几何中，见上面注记 4.1，向量和余向量来自不同的对象，分别是速度和一阶微分形式。尽管引入度量张量允许我们对向量余向量和余向量，但有时记住它们的起源，将它们视为不同的对象更为方便。

### 练习

4.1. 设

$$D = \sum_{k=1}^n v_k \frac{\partial}{\partial x_k}$$

是一个微分算子。利用链式法则，证明当我们改变基并用新坐标写出  $D$  时，其系数  $v_k$  按照向量的坐标变换规则变化。

## 5. 多线性函数 张量

### 5.1. 多线性函数

**定义 5.1** 设  $V_1, V_2, \dots, V_p, V$  是向量空间（在同一个域  $\mathbb{F}$  上）。一个值在  $V$  上的**多线性**(multilinear)( $p$ -线性)映射是一个函数  $F$ ，具有  $p$  个向量变量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p, \mathbf{v}_k \in V_k$ ，其目标空间为  $V$ ，在每个变量  $\mathbf{v}_k$  上都是线性的。换句话说，这意味着如果我们固定除  $\mathbf{v}_k$  之外的所有变量，我们会得到一个线性映射，并且这对所有  $k = 1, 2, \dots, p$  都成立。我们将使用符号  $L(V_1, V_2, \dots, V_p; V)$  表示所有这些多线性函数的集合。

如果目标空间  $V$  是标量域  $\mathbb{F}$ ，我们称  $F$  为**多线性泛函**(multilinear functional)，或**张量**(tensor)。数字  $p$  被称为多线性泛函（张量）的**价**(或**阶**)(valency)。因此，价

为 1 的张量是线性泛函，价为 2 的张量称为**双线性型**。

**例子** 设  $\mathbf{f}_k \in (V_k)'$ . 定义一个多线性泛函  $F = \mathbf{f}_1 \otimes \mathbf{f}_2 \otimes \cdots \otimes \mathbf{f}_p$  通过泛函相乘  $\mathbf{f}_k$ :

$$(5.1) \quad \mathbf{f}_1 \otimes \mathbf{f}_2 \otimes \cdots \otimes \mathbf{f}_p(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) = \mathbf{f}_1(\mathbf{v}_1)\mathbf{f}_2(\mathbf{v}_2) \cdots \mathbf{f}_p(\mathbf{v}_p),$$

其中  $\mathbf{v}_k \in V_k$ ,  $k = 1, 2, \dots, p$ . 多线性泛函  $\mathbf{f}_1 \otimes \mathbf{f}_2 \otimes \cdots \otimes \mathbf{f}_p$  被称为泛函  $\mathbf{f}_k$  的**张量积** (tensor product)。

### 5.1.1. 多线性函数构成向量空间

注意到，在空间  $L(V_1, V_2, \dots, V_p; V)$  中可以引入自然的加法和标量乘法运算：

$$(F_1 + F_2)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) := F_1(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) + F_2(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p)$$

$$(\alpha F_1)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) := \alpha F_1(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p)$$

其中  $F_1, F_2 \in L(V_1, V_2, \dots, V_p; V)$ ,  $\alpha \in \mathbb{F}$ .

装备了这些运算后，空间  $L(V_1, V_2, \dots, V_p; V)$  就是一个向量空间。

为了看出这一点，我们首先需要证明  $F_1 + F_2$  和  $\alpha F_1$  是多线性函数。由于“多线性”意味着它在每个参数上都是线性的（固定其他变量），这来源于线性变换的相应事实；即，线性变换的和以及线性变换的标量倍数是线性变换，参见第一章第 4 节。

然后很容易证明  $L(V_1, V_2, \dots, V_p; V)$  满足所有向量空间的公理；我们只需要使用  $V$  满足这些公理的事实。我们将细节留给读者作为练习。他/她可以参考第一章第 4 节，其中证明了线性变换的集合满足公理 7。所有其他公理也得到满足的证明非常相似。

### 5.1.2. $L(V_1, V_2, \dots, V_p; V)$ 的维数

设  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_p$  分别是  $V_1, V_2, \dots, V_p$  中的基。由于线性变换由其在基上的作用定义，多线性函数  $F \in L(V_1, V_2, \dots, V_p; V)$  由其在所有元组 (tuples)

$$\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p, \quad \mathbf{b}_{j_k}^k \in \mathcal{B}_k$$

上的值定义。由于恰好有

$$(\dim V_1)(\dim V_2) \cdots (\dim V_p)$$

个这样的元组，并且每个  $F(\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p)$  在 (某组基下) 由  $\dim V$  个坐标确定。因此，我们可以得出  $F \in L(V_1, V_2, \dots, V_p; V)$  由  $(\dim V_1)(\dim V_2) \cdots (\dim V_p)(\dim V)$  个项确定。换句话说，

$$\dim L(V_1, V_2, \dots, V_p; V) = (\dim V_1)(\dim V_2) \cdots (\dim V_p)(\dim V).$$



特别地, 如果目标空间是标量域  $\mathbb{F}$  (即, 如果我们处理多线性泛函),

$$\dim L(V_1, V_2, \dots, V_p; \mathbb{F}) = (\dim V_1)(\dim V_2) \dots (\dim V_p).$$

在  $L(V_1, V_2, \dots, V_p; \mathbb{F})$  中找到一组基很容易。也就是说, 设  $\mathcal{B}_k = \{\mathbf{b}_j^k\}_{j=1}^{\dim V_k}$  是  $V_k$  中的一组基, 设  $\mathcal{B}' = \{\tilde{\mathbf{b}}_j^k\}_{j=1}^{\dim V_k}$  是其对偶系统,  $\tilde{\mathbf{b}}_j^k \in V'_k$ .

### 命题 5.2 系统

$$\tilde{\mathbf{b}}_{j_1}^1 \otimes \tilde{\mathbf{b}}_{j_2}^2 \otimes \dots \otimes \tilde{\mathbf{b}}_{j_p}^p, \quad 1 \leq j_k \leq \dim V_k, \quad k = 1, 2, \dots, p$$

是空间  $L(V_1, V_2, \dots, V_p; \mathbb{F})$  中的一组基。这里  $\tilde{\mathbf{b}}_{j_1}^1 \otimes \tilde{\mathbf{b}}_{j_2}^2 \otimes \dots \otimes \tilde{\mathbf{b}}_{j_p}^p$  是泛函的张量积, 如 (5.1) 定义。

**证明:** 我们想将  $F$  表示为:

$$(5.2) \quad F = \sum_{j_1, j_2, \dots, j_p} \alpha^{j_1, j_2, \dots, j_p} \tilde{\mathbf{b}}_{j_1}^1 \otimes \tilde{\mathbf{b}}_{j_2}^2 \otimes \dots \otimes \tilde{\mathbf{b}}_{j_p}^p$$

由于  $\tilde{\mathbf{b}}_j(\mathbf{b}_l) = \delta_{j,l}$ , 我们得到:

$$(5.3) \quad \tilde{\mathbf{b}}_{j_1}^1 \otimes \tilde{\mathbf{b}}_{j_2}^2 \otimes \dots \otimes \tilde{\mathbf{b}}_{j_p}^p (\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p) = 1$$

并且

$$(5.4) \quad \tilde{\mathbf{b}}_{j_1}^1 \otimes \tilde{\mathbf{b}}_{j_2}^2 \otimes \dots \otimes \tilde{\mathbf{b}}_{j_p}^p (\mathbf{b}_{j'_1}^1, \mathbf{b}_{j'_2}^2, \dots, \mathbf{b}_{j'_p}^p) = 0$$

对于任何不同于  $j_1, j_2, \dots, j_p$  的指标集合  $j'_1, j'_2, \dots, j'_p$ .

因此, 将 (5.2) 应用于元组  $\{\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p\}$ , 我们得到:

$$\alpha_{j_1, j_2, \dots, j_p} = F(\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p),$$

所以表示 (5.2) 是唯一的 (如果存在的话)。

另一方面, 定义  $\alpha_{j_1, j_2, \dots, j_p} := F(\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p)$  并使用 (5.3) 和 (5.4), 我们可以看到等式 (5.2) 对所有形式为  $\{\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p\}$  的元组都成立。因此, 表示 (5.2) 成立, 所以我们确实得到了一组基。□

## 5.2. 张量积

**定义** 设  $V_1, V_2, \dots, V_p$  是向量空间。空间

$$V_1 \otimes V_2 \otimes \dots \otimes V_p$$

的张量积就是  $V'_1, V'_2, \dots, V'_p$  的对偶空间的张量积  $L(V'_1, V'_2, \dots, V'_p; \mathbb{F})$  的集合; 这里  $V'_k$  是  $V_k$  的对偶。

**注记 5.3** 根据命题 5.2, 如果  $\mathcal{B}_k = \{\mathbf{b}_j^k\}_{j=1}^{\dim V_k}$  是  $V_k$  中的一组基, 那么系统:

$$(5.5) \quad \mathbf{b}_{j_1}^1 \otimes \mathbf{b}_{j_2}^2 \otimes \dots \otimes \mathbf{b}_{j_p}^p, \quad 1 \leq j_k \leq \dim V_k, \quad k = 1, 2, \dots, p$$

是  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  中的一组基。

这里我们将向量  $\mathbf{v}_k \in V_k$  看作是  $V'_k$  上的一个线性泛函；向量的张量积  $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  是根据 (5.1) 定义的。

**注记** 向量的张量积  $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  在每个参数  $\mathbf{v}_k$  上显然是线性的。换句话说，映射  $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) \mapsto \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  是一个取值于  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  的多线性泛函。我们将证明留给读者作为练习，见下面的练习 5.1。

**注记** 注意，向量张量积的集合  $\{\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p : \mathbf{v}_k \in V_k\}$  严格小于  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$ ，见下面的练习 5.2。

### 5.2.1. 将多线性函数提升到张量积上的线性变换

**命题 5.4** 对于任何多线性函数  $F \in L(V_1, V_2, \dots, V_p; V)$ ，存在唯一一个线性变换  $T : V_1 \otimes V_2 \otimes \cdots \otimes V_p \rightarrow V$  扩展  $F$ ，即满足：

$$(5.6) \quad F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) = T\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p,$$

对于所有向量  $\mathbf{v}_k \in V_k$ ,  $1 \leq k \leq p$  的选择。

**注记** 如果  $T : V_1 \otimes V_2 \otimes \cdots \otimes V_p \rightarrow V$  是一个线性变换，那么显然函数  $F$ ,

$$F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) := T\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p,$$

是  $L(V_1, V_2, \dots, V_p; V)$  中的一个多线性函数。这直接源于表达式  $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  在每个变量  $\mathbf{v}_k$  上是线性的。

**定理 5.4 的证明** 在基 (5.5) 上定义  $T$  为：

$$T\mathbf{b}_{j_1}^1 \otimes \mathbf{b}_{j_2}^2 \otimes \cdots \otimes \mathbf{b}_{j_p}^p = F(\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p)$$

然后通过线性将其扩展到整个空间  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$ 。为了完成证明，我们需要证明 (5.6) 对所有向量  $\mathbf{v}_k \in V_k, 1 \leq k \leq p$  的选择都成立（我们现在知道只有当每个  $\mathbf{v}_k$  是  $\mathbf{b}_{j_k}^k$  之一时才成立）。

为了证明这一点，让我们将  $\mathbf{v}_k$  分解为：

$$\mathbf{v}_k = \sum_{j_k} \alpha_{j_k}^k \mathbf{b}_{j_k}^k, \quad k = 1, 2, \dots, p.$$

使用每个变量的线性性质，我们得到：

$$\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p = \sum_{j_1, j_2, \dots, j_p} \alpha_{j_1}^1 \alpha_{j_2}^2 \cdots \alpha_{j_p}^p \mathbf{b}_{j_1}^1 \otimes \mathbf{b}_{j_2}^2 \otimes \cdots \otimes \mathbf{b}_{j_p}^p,$$

$$F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) = \sum_{j_1, j_2, \dots, j_p} \alpha_{j_1}^1 \alpha_{j_2}^2 \cdots \alpha_{j_p}^p F(\mathbf{b}_{j_1}^1, \mathbf{b}_{j_2}^2, \dots, \mathbf{b}_{j_p}^p)$$

所以根据  $T$  的定义，恒等式 (5.6) 成立。  $\square$

### 5.2.2. 张量积的对偶

容易看到, 张量积  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  的对偶是张量积  $V'_1 \otimes V'_2 \otimes \cdots \otimes V'_p$ .

事实上, 根据命题 5.4 和其后的注记, 在多线性泛函  $L(V_1, V_2, \dots, V_p, \mathbb{F})$  (即  $V'_1 \otimes V'_2 \otimes \cdots \otimes V'_p$  的元素) 与线性变换  $T: V_1 \otimes V_2 \otimes \cdots \otimes V_p \rightarrow \mathbb{F}$  (即  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  的对偶的元素) 之间存在自然的一一对应关系。

注意, 注记 5.3 和命题 5.2 中的基是对偶基 (分别为  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  和  $V'_1 \otimes V'_2 \otimes \cdots \otimes V'_p$ )。了解对偶基可以让我们轻松地计算空间  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$  和  $V'_1 \otimes V'_2 \otimes \cdots \otimes V'_p$  之间的**对偶性**(duality), 即表达式  $\langle \mathbf{x}, \mathbf{x}' \rangle, \mathbf{x} \in V_1 \otimes V_2 \otimes \cdots \otimes V_p, \mathbf{x}' \in V'_1 \otimes V'_2 \otimes \cdots \otimes V'_p$ 。

### 5.3. 协变和逆变张量

设  $X_1, X_2, \dots, X_p$  是向量空间, 设  $V_k$  是  $X_k$  或  $X'_k, k = 1, 2, \dots, p$ . 对于多线性函数  $F \in L(V_1, V_2, \dots, V_p; V)$ , 我们说它对于变量  $\mathbf{v}_k \in V_k$  是**协变的**, 如果  $V_k = X_k$ ; 对于这个变量是**逆变的**, 如果  $V_k = X'_k$ .

如果一个多线性函数在所有变量上都是协变的 (逆变的), 我们称该多线性函数是协变的 (逆变的)。一般地, 如果一个函数在  $r$  个变量上是协变的, 在  $s$  个变量上是逆变的, 我们称该多线性函数是  $r$ -协变  $s$ -逆变的 (或简单地称为  $(r, s)$  多线性函数, 或称其价为  $(r, s)$ )。

因此, 线性泛函可以解释为 1-协变张量 (回忆一下, 我们用**张量**这个词来指代目标空间是标量域  $\mathbb{F}$  的情况)。根据对偶性, 向量可以解释为 1-逆变张量。

**注记** 一开始, 这个术语可能看起来有点令人困惑: 如果一个变量是向量 (而不是泛函), 它是一个协变变量, 但却是一个逆变对象 (张量)。但是请注意, 我们这里说的不是 “协变变量”: 我们说的是, 如果  $\mathbf{v}_k \in X_k$ , 那么该**多线性函数在变量  $\mathbf{v}_k$  上是协变的**。所以, 我说的协变对象不是  $\mathbf{v}_k$ , 而是张量中我们放入它的 “槽” (slot)!

所以没有矛盾, 我们将逆变对象放入协变槽, 反之亦然。

有时, 稍微滥用术语, 人们会谈论协变 (逆变) 变量或参数。但通常的意思是相应的张量中的 “槽” 是协变的 (逆变的), 而不是作为对象的变量。

### 5.3.1. 线性变换作为张量

一个线性变换  $T: X_1 \rightarrow X_2$  可以被解释为一个 1-协变 1-逆变张量。也就是说，双线性泛函  $F$ ，

$$F(\mathbf{x}_1, \mathbf{x}'_2) := \langle T\mathbf{x}_1, \mathbf{x}'_2 \rangle, \quad \mathbf{x}_1 \in X_1, \quad \mathbf{x}'_2 \in X'_2$$

在第一个变量  $\mathbf{x}_1$  上是协变的，在第二个变量  $\mathbf{x}'_2$  上是逆变的。

反之，

**命题 5.5** 给定一个 1-1- 张量  $F \in L(X_1, X'_2; \mathbb{F})$ ，存在唯一一个线性变换  $T: X_1 \rightarrow X_2$  使得

$$(5.7) \quad F(\mathbf{x}_1, \mathbf{x}'_2) := \langle T\mathbf{x}_1, \mathbf{x}'_2 \rangle$$

对于所有  $\mathbf{x}_1 \in X_1$ ,  $\mathbf{x}'_2 \in X'_2$  的选择成立。

**证明：**首先，请注意，由于引理 1.3，唯一性是平凡的推论，参见上面问题 3.1。所以我们只需要证明  $T$  的存在性。

设  $B_k = \{\mathbf{b}_j^k\}_{j=1}^{\dim X_k}$  是  $X_k$  中的一组基，设  $B'_k = \{\tilde{\mathbf{b}}_j^k\}_{j=1}^{\dim X_k}$  是  $X'_k$  中的对偶基， $k = 1, 2$ 。然后定义矩阵  $A = \{a_{k,j}\}_{k=1}^{\dim X_2} \{j=1}^{\dim X_1}$  为：

$$a_{k,j} = F(\mathbf{b}_j^1, \tilde{\mathbf{b}}_k^{2,k})$$

将  $T$  定义为矩阵  $[T]_{B_2, B_1} = A$  对应的算子。显然（参见注记 1.5）：

$$(5.8) \quad \langle T\mathbf{b}_j^1, \tilde{\mathbf{b}}_k^2 \rangle = a_{k,j} = F(\mathbf{b}_j^1, \tilde{\mathbf{b}}_k^2)$$

这暗示了等式 (5.7)。通过将  $\mathbf{x}_1 = \sum_j \alpha_j \mathbf{b}_j^1$  和  $\mathbf{x}'_2 = \sum_k \beta_k \tilde{\mathbf{b}}_k^2$  分解，并利用每个参数的线性，可以很容易地看出这一点。

另一种更“高深”的解释是，(5.7) 两边的张量在  $X_1 \otimes X'_2$  的基上（参见注记 5.3 中关于基的内容）是相同的，所以它们是相等的。更准确地说，人们应该将双线性型升为线性变换（泛函） $X_1 \otimes X'_2 \rightarrow \mathbb{F}$ （参见命题 5.4），并且由于变换在基上是相同的，所以它们是相等的。

也可以提出  $T$  存在性的一种替代的、无坐标的证明，沿着对偶空间（见 3.1.3 节）的无坐标定义的思路。也就是说，如果我们固定  $\mathbf{x}_1$ ，函数  $F(\mathbf{x}_1, \mathbf{x}'_2)$  在  $X'_2$  上是线性的，所以它是一个  $X'_2$  上的线性泛函，即  $X_2$  中的一个向量。

让我们称这个向量为  $T(\mathbf{x}_1)$ 。所以我们定义了一个变换  $T: X_1 \rightarrow X_2$ 。可以很容易地通过基本上重复 3.1.3 节中的推理来证明  $T$  是一个线性变换。等式 (5.7) 从  $T$  的定义中自动得出。□

**注记** 注意，我们也说  $F$  从命题 5.5 定义的不是变换  $T$ ，而是它的伴随。先验地，不假设任何东西（如变量的顺序及其解释），我们就无法区分一个变换和它的伴随。

**注记** 注意，如果我们想遵循爱因斯坦求和约定，变换  $T$  的矩阵  $A = [T]_{\mathcal{B}_2, \mathcal{B}_1}$  的项  $a_{j,k}$  应该写成  $a_k^j$ ，那么，如果  $x^k, k = 1, 2, \dots, \dim X_1$  是向量  $\mathbf{x} \in X_1$  的坐标，那么  $\mathbf{y} = T\mathbf{x}$  的第  $j$  个坐标由下式给出：

$$y^j = a_k^j x^k.$$

(这里我们跳过了求和符号，但我们指的是  $k$  上的求和)。还请注意，我们保持了指标的位置，所以  $j$  指标留在上面。指标  $k$  没有出现在等式左侧，因为它在右侧被求和消掉 (kill) 了。

类似地，如果  $x'_j, j = 1, 2, \dots, \dim X_2$  是向量  $\mathbf{x}' \in X'_2$  的坐标，那么  $\mathbf{y}' = T'\mathbf{x}'$  的第  $k$  个坐标由下式给出：

$$y_k = a_k^j x_j.$$

(再次，跳过  $j$  上的求和)。同样，由于我们保持了指标的位置，所以在  $y_k$  中的指标  $k$  是下标。

注意，由于  $\mathbf{x} \in X_1$  且  $\mathbf{y} = T\mathbf{x} \in X_2$  是向量，根据爱因斯坦求和约定，其坐标中的指标确实应该写成上标。

类似地， $\mathbf{x}' \in X'_2$  且  $\mathbf{y}' = T'\mathbf{x}' \in X'_1$  是余向量，所以其坐标中的指标应该写成下标。

爱因斯坦求和约定强调了上一注记中提到的事实，即一个 1-协变 1-逆变张量同时给了我们一个线性变换及其伴随：表达式  $a_k^j x^k$  给出了  $T$  的作用，而  $a_k^j x_j$  给出了其伴随  $T'$  的作用。

### 5.3.2. 多线性变换作为张量

更一般地，任何多线性变换都可以被解释为一个张量。也就是说，给定一个多线性变换  $F \in L(V_1, V_2, \dots, V_p; V)$ ，我们可以定义张量  $\tilde{F} \in L(V_1, V_2, \dots, V_p, V'; \mathbb{F})$  为：

$$(5.9) \quad \tilde{F}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p, \mathbf{v}') = \langle F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p), \mathbf{v}' \rangle, \quad \mathbf{v}_k \in V_k, \mathbf{v}' \in V'.$$

反之，

**命题 5.6** 给定一个张量  $\tilde{F} \in L(V_1, V_2, \dots, V_p, V'; \mathbb{F})$ ，存在唯一一个多线性变换  $F \in L(V_1, V_2, \dots, V_p; V)$  使得 (5.9) 成立。

**证明：**根据命题 5.4，张量  $\tilde{F}$  可以扩展为一个线性变换（泛函） $\tilde{T} : V_1 \otimes V_2 \otimes \cdots \otimes V_p \otimes V' \rightarrow \mathbb{F}$ ，使得

$$\tilde{F}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p, \mathbf{v}') = \tilde{T}(\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}')$$

对于所有  $\mathbf{v}_k \in V_k, \mathbf{v}' \in V'$ 。

如果  $\mathbf{w} \in W := V_1 \otimes V_2 \otimes \cdots \otimes V_p$  且  $\mathbf{v}' \in V'$ ，那么

$$\mathbf{w} \otimes \mathbf{v}' \in V_1 \otimes V_2 \otimes \cdots \otimes V_p \otimes V'.$$

因此，我们可以定义一个双线性泛函（张量） $G \in L(W, V'; \mathbb{F})$  为：

$$G(\mathbf{w}, \mathbf{v}') := \tilde{T}(\mathbf{w} \otimes \mathbf{v}')$$

根据命题 5.5， $G$  产生一个线性变换，即存在唯一一个线性变换  $T : W \rightarrow V'$  使得

$$G(\mathbf{w}, \mathbf{v}') = \langle T\mathbf{w}, \mathbf{v}' \rangle \quad \forall \mathbf{w} \in W, \forall \mathbf{v}' \in V'$$

而线性变换  $T$  通过

$$F \in L(V_1, V_2, \dots, V_p; V)$$

定义为

$$F(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p) = T(\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p),$$

见命题 5.4 后的注记。

变换  $F$  的唯一性，如同命题 5.5 中一样，是引理 1.3 的一个平凡推论。我们将细节留给读者作为练习。  $\square$

本节展示了

张量是多线性代数中的通用对象，因为任何多线性变换都可以被解释为一个张量，反之亦然。

### 练习

- 5.1. 证明向量的张量积  $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  在每个参数  $\mathbf{v}_k$  上是线性的。
- 5.2. 证明向量张量积的集合  $\{\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p : \mathbf{v}_k \in V_k\}$  严格小于  $V_1 \otimes V_2 \otimes \cdots \otimes V_p$ 。
- 5.3. 证明命题 5.6 中的变换  $F$  是唯一的。

## 6. 张量的坐标变换公式

区分多线性协变和逆变变量的主要原因是，在改变基时，它们的坐标根据不同的规则变化。因此，协变和逆变向量的项也根据不同的规则变化。

在本节中，我们将详细研究这一点。请注意，坐标表示极其重要，其原因是，在实践上所有数值计算（与理论研究不同）都是在某一坐标系下进行的。

## 6.1. 张量的坐标表示

设  $F$  是一个  $r$ -协变  $s$ -逆变张量,  $r + s = p$ . 设  $\mathbf{x}_1, \dots, \mathbf{x}_r$  是协变变量 ( $\mathbf{x}_k \in X_k$ ),  $\mathbf{f}_1, \dots, \mathbf{f}_s$  是逆变变量 ( $\mathbf{f}_k \in X'_k$ ). 让我们先写协变变量, 所以张量将被写成  $F(\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{f}_1, \dots, \mathbf{f}_s)$ . 对于  $k = 1, 2, \dots, p$ , 固定  $X_k$  中的基  $\mathcal{B}_k = \{\mathbf{b}_j^{(k)}\}_{j=1}^{\dim X_k}$ , 设  $\mathcal{B}'_k = \{\tilde{\mathbf{b}}_j^{(k)}\}_{j=1}^{\dim X_k}$  为  $X'_k$  中的对偶基。

对于向量  $\mathbf{x}_k \in X_k$ , 设  $x_{(k)}^j, j = 1, 2, \dots, \dim X_k$  是其在基  $\mathcal{B}_k$  下的坐标, 类似地, 如果  $\mathbf{f}_k \in X'_k$ , 设  $f_k^{(j)}, j = 1, 2, \dots, \dim X_k$  是其在对偶基  $\mathcal{B}'_k$  下的坐标 (注意, 为了与爱因斯坦求和约定保持一致, 向量的坐标用上标索引, 协向量的坐标用下标索引)。

**命题 6.1** 记:

$$(6.1) \quad \phi_{k_1, \dots, k_s}^{j_1, \dots, j_r} := F(\mathbf{b}_{j_1}^{(1)}, \dots, \mathbf{b}_{j_r}^{(r)}, \tilde{\mathbf{b}}_{k_1}^{(r+1)}, \dots, \tilde{\mathbf{b}}_{k_s}^{(r+s)})$$

那么, 在爱因斯坦求和约定下:

$$(6.2) \quad F(\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{f}_1, \dots, \mathbf{f}_s) = \phi_{k_1, \dots, k_s}^{j_1, \dots, j_r} x_{(1)}^{j_1} \dots x_{(r)}^{j_r} f_{k_1}^{(1)} \dots f_{k_s}^{(s)}$$

(这里的求和是指对指标  $j_1, \dots, j_r$  和  $k_1, \dots, k_s$ )。

注意我们使用符号  $(1), \dots, (r)$  和  $(1), \dots, (s)$  来强调它们不是指标: 括号中的数字仅仅表示参数的顺序。因此, (6.2) 的右侧没有剩下任何指标 (所有指标都在求和中使用了), 所以它只是一个数字 (对于固定的  $\mathbf{x}_k$  和  $\mathbf{f}_k$ )。

**命题 6.1 的证明** 为了证明 (6.1) 意味着 (6.2), 我们首先注意到 (6.1) 意味着 (6.2) 在  $\mathbf{x}_j$  和  $\mathbf{f}_k$  是对应基的元素时成立。通过将每个参数  $\mathbf{x}_j$  和  $\mathbf{f}_k$  分解到相应的基中, 并使用每个参数的线性, 我们可以很容易地得到 (6.2)。这个计算相当简单, 但由于指标很多, 公式可能会非常大, 看起来可能非常吓人。

为了避免写出太多巨大的公式, 我们将这个计算留给读者作为练习。

我们不希望读者感到被欺骗, 所以我们提供了另一种更“高深”(抽象)的解释, 它不需要任何计算! 也就是说, 让我们注意到 (6.2) 等式两边的表达式定义了张量。根据命题 5.4, 它们可以提升为张量积  $X_1 \otimes \dots \otimes X_r \otimes X'_{r+1} \otimes \dots \otimes X'_{r+s}$  上的线性函数。

重新表述我们在此证明的开头讨论的内容, 我们可以说 (6.1) 意味着函数在基

$$\mathbf{b}_{j_1}^{(1)} \otimes \dots \otimes \mathbf{b}_{j_r}^{(r)} \otimes \tilde{\mathbf{b}}_{k_1}^{(r+1)} \otimes \dots \otimes \tilde{\mathbf{b}}_{k_s}^{(r+s)}$$

上是相同的, 所以函数 (因此, 张量) 是相等的。  $\square$

张量  $F$  的项  $\phi_{k_1, \dots, k_s}^{j_1, \dots, j_r}$  被称为张量  $F$  在基  $\mathcal{B}_k, k = 1, 2, \dots, p$  下的项。

现在, 设  $\mathcal{A}_k$  (以及  $\mathcal{A}'_k$ ) 分别是  $X_k$  (以及  $X'_k$ ) 中的一组基。我们想研究当基从  $\mathcal{B}_k$  改变到  $\mathcal{A}_k$  时, 张量  $F$  的项如何变化。

## 6.2. 爱因斯坦求和约定中的坐标变换公式

首先, 让我们考虑更熟悉的 8.1.1.1 节中向量和线性泛函的情况, 但使用爱因斯坦求和约定将其写下来。设  $X$  中有两组基  $\mathcal{B}$  和  $\mathcal{A}$ , 设

$$A = [I]_{\mathcal{A}, \mathcal{B}}$$

是从  $\mathcal{B}$  到  $\mathcal{A}$  的坐标变换矩阵。对于向量  $\mathbf{x} \in X$ , 设  $x^k$  是其在基  $\mathcal{B}$  下的坐标, 设  $\tilde{x}^k$  是其在基  $\mathcal{A}$  下的坐标。类似地, 对于  $\mathbf{f} \in X'$ , 设  $f_k$  是其在基  $\mathcal{B}'$  下的坐标, 设  $\tilde{f}_k$  是其在基  $\mathcal{A}'$  下的坐标 ( $\mathcal{B}'$  和  $\mathcal{A}'$  分别是  $\mathcal{B}$  和  $\mathcal{A}$  的对偶基)。

设  $(A)_k^j$  为矩阵  $A$  的项: 为了与爱因斯坦求和约定保持一致, 上标  $j$  表示行的编号。那么我们可以将坐标变换公式写成:

$$(6.3) \quad \tilde{x}^j = (A)_k^j x^k.$$

类似地, 设  $(A^{-1})_j^k$  为  $A^{-1}$  的项: 再次, 上标表示行的编号。那么我们可以将对偶空间的坐标变换公式写成:

$$(6.4) \quad \tilde{f}_j = (A^{-1})_j^k f_k;$$

这里的求和是在指标  $k$  上 (即沿着  $A^{-1}$  的列), 所以这种情况下的坐标变换矩阵确实是  $(A^{-1})^T$ 。

让我们强调我们在这里没有证明任何东西: 我们只是用爱因斯坦求和约定重写了本章 1.1.1 节中的公式 (1.1)。

**注记** 虽然在接下来的内容中不需要, 但让我们多玩玩爱因斯坦求和约定。也就是说,

$$A^{-1}A = I \quad \text{和} \quad AA^{-1} = I$$

这两个方程可以分别用爱因斯坦求和约定重写为:

$$(A)_k^j (A^{-1})_l^k = \delta_{j,l} \quad \text{和} \quad (A^{-1})_j^k (A)_l^j = \delta_{k,l}.$$

## 6.3. 张量的坐标变换公式

现在我们准备给出一般张量的坐标变换公式。

对于  $k = 1, 2, \dots, p := r + s$ , 设  $A_k = [I]_{\mathcal{A}, \mathcal{B}}$  为坐标变换矩阵, 设  $A_k^{-1}$  为其逆矩阵。

像在 6.2 节中一样, 我们用  $(A)_k^j$  表示矩阵  $A$  的项, 约定上标给出行的编号。



**命题 6.2** 给定一个  $r$ -协变  $s$ -逆变张量  $F$ , 设

$$\phi_{j_1, \dots, j_r}^{k_1, \dots, k_s} \quad \text{和} \quad \tilde{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s}$$

分别是其在 (旧的) 基  $\mathcal{B}_k$  和 (新的)  $\mathcal{A}_k$  下的项。在上面的记法中:

$$\tilde{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s} = \phi_{j'_1, \dots, j'_r}^{k'_1, \dots, k'_s} (A_1^{-1})_{j_1}^{j'_1} \dots (A_r^{-1})_{j_r}^{j'_r} (A_{r+1})_{k'_1}^{k_1} \dots (A_{r+s})_{k'_s}^{k_s}$$

(这里的求和是在指标  $j'_1, \dots, j'_r$  和  $k'_1, \dots, k'_s$  上)。

由于公式中有许多指标, 这个命题看起来非常复杂。然而, 如果理解了主要思想, 公式就会变得相当简单且易于记忆。

为了解释主要思想, 让我们稍微滥用词汇, 用通俗的话来表达这个公式:

为了用“旧”张量项  $\phi_{j_1, \dots, j_r}^{k_1, \dots, k_s}$  来表示“新”张量项  $\tilde{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s}$ , 对于每个**协变**指标 (下标) 需要应用协变规则 (6.4), 对于每个**逆变**指标 (上标) 需要应用逆变规则 (6.3)。

**命题 6.2 的证明** 非正式地, 证明的思路非常简单: 我们一次只改变一组基, 每次应用坐标变换公式 (6.3) 或者 (6.4), 这取决于张量在相应变量上是协变还是逆变。

为了写出严格的正式证明, 我们将使用关于  $r$  和  $s$  (张量的协变和逆变变量的数量) 的归纳法。命题在  $r = 1, s = 0$  和  $r = 0, s = 1$  时成立, 分别参见 (6.4) 与 (6.3)。

现在假设命题对某些  $p$  和  $s$  已经证明, 我们来证明  $r + 1, s$  和  $r, s + 1$  的情况。

我们来处理后者, 前者类似。主要思想是, 我们首先改变  $p = r + s$  组基并使用归纳假设; 然后我们改变最后一组基并使用 (6.3)。

也就是说, 设  $\hat{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}}$  是一个  $(r, s + 1)$  价张量  $F$  在基  $\mathcal{A}_1, \dots, \mathcal{A}_p, \mathcal{B}_{p+1}$  下的项,  $p = r + s$ 。

让我们固定指标  $k_{s+1}$ , 并考虑张量  $F(\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{f}_1, \dots, \mathbf{f}_s, \tilde{\mathbf{b}}_{k_{s+1}}^{(r+s+1)})$  的  $r$ -协变  $s$ -逆变函数 (其中  $\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{f}_1, \dots, \mathbf{f}_s$  是变量)。显然

$$\phi_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}} \quad \text{和} \quad \hat{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}}$$

是这个函数在基  $\mathcal{B}_1, \dots, \mathcal{B}_p$  和  $\mathcal{A}_1, \dots, \mathcal{A}_p$  下的项 (你能看出为什么吗?)。这里指标  $k_{s+1}$  是固定的。

根据归纳假设:

$$(6.5) \quad \hat{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}} = \phi_{j'_1, \dots, j'_r}^{k'_1, \dots, k'_s, k_{s+1}} (A_1^{-1})_{j_1}^{j'_1} \dots (A_r^{-1})_{j_r}^{j'_r} (A_{r+1})_{k'_1}^{k_1} \dots (A_{r+s})_{k'_s}^{k_s}$$

注意, 我们没有对指标  $k_{s+1}$  做任何假设, 所以 (6.5) 对所有  $k_{s+1}$  都成立。

现在让我们固定指标  $j_1, \dots, j_r, k_1, \dots, k_s$ , 并考虑变量  $\mathbf{f}_{s+1}$  的 1-逆变张量

$$F(\mathbf{a}_{j_1}^{(1)}, \dots, \mathbf{a}_{j_r}^{(r)}, \tilde{\mathbf{a}}_{k_1}^{(r+1)}, \dots, \tilde{\mathbf{a}}_{k_s}^{(r+s)}, \mathbf{f}_{s+1}).$$

这里  $\mathbf{a}_j^{(k)}$  是基  $\mathcal{A}_k$  中的向量,  $\tilde{\mathbf{a}}_j^{(k)}$  是对偶基  $\mathcal{A}'_k$  中的向量。

再一次, 很容易看出

$$\hat{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}} \quad \text{和} \quad \tilde{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}}$$

$j_{s+1} = 1, 2, \dots, \dim X_{p+1}$ , 是这个泛函在基  $\mathcal{B}_{p+1}$  和  $\mathcal{A}_{p+1}$  下的指标。根据 (6.3):

$$\tilde{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k_{s+1}} = \hat{\phi}_{j_1, \dots, j_r}^{k_1, \dots, k_s, k'_{s+1}} (A_{p+1})_{k'_{s+1}}^{k_{s+1}},$$

由于我们没有对指标  $j_1, \dots, j_r, k_1, \dots, k_s$  做任何假设, 所以上述恒等式对它们的所有组合都成立。将此与 (6.5) 相结合, 我们得到该命题对  $(r, s+1)$  价张量成立。

$(r+1, s)$  价张量的情况也是完全一样的处理方法: 唯一的区别是最后我们得到一个 1-协变张量, 并使用 (6.4) 而不是 (6.3)。□

## 第九章 高级谱理论

### 1. 凯莱-哈密顿定理

**定理 1.1.** (凯莱-哈密顿, Cayley–Hamilton) 设  $A$  是一个方阵, 其特征多项式为  $p(\lambda) = \det(A - \lambda I)$ . 那么,  $p(A) = \mathbf{0}$ .

**一个错误的证明** 这个证明看起来异常简单: 将  $\lambda$  替换为  $A$  并代入特征多项式 (译者吐槽: 正常人怎么会想要用一个矩阵去替换一个数的啊…这也不是特征值的定义啊…), 我们得到

$$p(A) = \det(A - AI) = \det(\mathbf{0}) = 0.$$

□

但这这是一个错误的证明! 为了明白为什么错误, 让我们分析一下定理的内容。定理说明, 如果我们计算特征多项式

$$\det(A - \lambda I) = p(\lambda) = \sum_{k=0}^n c_k \lambda^k,$$

然后用矩阵  $A$  替换  $\lambda$  得到

$$p(A) := \sum_{k=0}^n c_k A^k = c_0 I + c_1 A + \cdots + c_n A^n,$$

那么结果将是零矩阵。

我们并不清楚, 为何仅仅是将  $A$  而不是将  $\lambda$  代入行列式  $\det(A - \lambda I)$ , 就得到了相同的结果。而且, 很容易看出, 除了  $1 \times 1$  矩阵的平凡情况外, 我们得到的对象是不同的。即,  $A - AI$  是零矩阵, 它的行列式只是数字 0。

但是  $p(A)$  是一个矩阵, 而定理声称这个矩阵是零矩阵。因此, 我们刚刚其实只是在比较苹果和橘子。尽管两种情况下我们都得到了零, 但它们是不同的零: 数字零和零矩阵!

让我们给出另一组基于分析思想的证明。

#### 一个“连续的”证明 <sup>51</sup>

---

<sup>51</sup> 这个证明阐述了一个重要的思想, 即通常只考虑典型、一般的状况就足够了。虽然这超出了本书的范围, 但我们仍提及一下, 而不深入细节: 一个一般的 (即典型的) 矩阵是可对角化的。

该证明基于几个观察。首先，对于对角矩阵，该定理是平凡的，因此对于与对角矩阵相似的矩阵（即可对角化矩阵）也是平凡的（见下文问题 1.1）。

第二个观察是，任何矩阵都可以被可对角化矩阵近似（任意精确）。由于任何算子在某组标准正交基下都可以表示为上三角矩阵（见第六章定理 1.1），我们可以不失一般性地假设  $A$  是一个上三角矩阵。

我们可以通过微扰  $A$  的对角线元素（任意小）来使它们全部不同，因此扰动后的矩阵  $\tilde{A}$  是可对角化的（三角矩阵的特征值是其对角线元素，见第四章第 1.7 节，并且根据第四章推论 2.3，一个具有  $n$  个不同特征值的  $n \times n$  矩阵是可对角化的）。

如我刚才提到的，我们可以任意小地扰动  $A$  的对角线元素，所以弗罗贝尼乌斯范数  $\|A - \tilde{A}\|_2$  可以任意小。因此，我们可以找到一个可对角化矩阵序列  $A_k$  使得  $A_k \rightarrow A$  当  $k \rightarrow \infty$ （例如，使得  $\|A_k - A\|_2 \rightarrow 0$  当  $k \rightarrow \infty$ ）。可以证明，特征多项式  $p_k(\lambda) = \det(A_k - \lambda I)$  收敛于  $A$  的特征多项式  $p(\lambda) = \det(A - \lambda I)$ 。因此，

$$p(A) = \lim_{k \rightarrow \infty} p_k(A_k).$$

但是正如我们上面讨论的，对于可对角化矩阵，凯莱-哈密顿定理是平凡的，所以  $p_k(A_k) = \mathbf{0}$ 。因此， $p(A) = \lim_{k \rightarrow \infty} \mathbf{0} = \mathbf{0}$ 。□

这个证明是为那些熟悉分析（即连续性、收敛性等概念的严格处理）中的想法的读者准备的。<sup>52</sup> 这样的读者应该能够填补所有细节，并且对他/她来说，这个证明应该看起来非常简单和自然。

然而，对于还不熟悉这些想法的读者来说，这个证明肯定会显得奇怪。它甚至可能看起来像是在作弊，但让我重申，这是一个完全正确且严谨的证明（取决于分析中的一些标准事实）。因此，让我们给出定理的另一个证明，这也是其他线性代数教科书中的“标准”证明之一。

**一个“标准”的证明** 我们知道，见第六章定理 6.1.1，任何方阵都与一个上三角矩阵酉等价。由于对于任何多项式  $p$ ，我们有  $p(UAU^{-1}) = Up(A)U^{-1}$ ，并且酉等价矩阵的特征多项式是相同的，所以我们只需要证明该定理对于上三角矩阵成立。

因此，设  $A$  是一个上三角矩阵。我们知道三角矩阵的对角线元素与其特征值

<sup>52</sup> 这里我指的是**分析**，即连续性、收敛性等概念的严格处理，而不是微积分，微积分在其目前的教学方式下，仅仅是一系列技巧的集合。

相等, 所以设  $\lambda_1, \lambda_2, \dots, \lambda_n$  是  $A$  的特征值, 按其在对角线上的顺序排列, 即

$$A = \begin{pmatrix} \lambda_1 & & * \\ & \lambda_2 & \\ & & \ddots \\ \mathbf{0} & & & \lambda_n \end{pmatrix}.$$

$A$  的特征多项式  $p(z) = \det(A - zI)$  可以表示为

$$p(z) = (\lambda_1 - z)(\lambda_2 - z) \dots (\lambda_n - z) = (-1)^n (z - \lambda_1)(z - \lambda_2) \dots (z - \lambda_n),$$

所以

$$p(A) = (-1)^n (A - \lambda_1 I)(A - \lambda_2 I) \dots (A - \lambda_n I).$$

定义子空间  $E_k := \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k\}$ , 其中  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  是  $\mathbb{C}^n$  中的标准基。由于  $A$  的矩阵是上三角的, 子空间  $E_k$  是算子  $A$  的**不变子空间**(invariant subspace), 即  $AE_k \subset E_k$  (表示  $\mathbf{v} \in E_k$  的所有向量  $A\mathbf{v}$  都属于  $E_k$ )。此外, 由于对于任何  $\mathbf{v} \in E_k$  和任何  $\lambda$ ,

$$(A - \lambda I)\mathbf{v} = A\mathbf{v} - \lambda\mathbf{v} \in E_k,$$

因为  $A\mathbf{v}$  和  $\lambda\mathbf{v}$  都属于  $E_k$ , 因此  $(A - \lambda I)E_k \subset E_k$ , 即  $E_k$  是  $A - \lambda I$  的不变子空间。

我们可以进一步探讨子空间  $(A - \lambda_k I)E_k$  的性质。即,  $(A - \lambda_k I)\mathbf{e}_k \in \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{k-1}\}$ , 因为  $A - \lambda_k I$  矩阵的第  $k$  列的前  $k - 1$  个元素可能非零。另一方面, 对于  $j < k$ , 有  $(A - \lambda_k)\mathbf{e}_j \in E_j \subset E_k$  (因为  $E_j$  是  $A - \lambda_k I$  的不变子空间)。

取任意向量  $\mathbf{v} \in E_k$ . 根据  $E_k$  的定义, 它可以表示为向量  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$  的线性组合。由于所有向量  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$  都被  $A - \lambda_k I$  映射到  $E_{k-1}$  中的某个向量, 我们可以得出结论:

$$(1.1) \quad (A - \lambda_k I)\mathbf{v} \in E_{k-1} \quad \forall \mathbf{v} \in E_k.$$

取任意向量  $\mathbf{x} \in \mathbb{C}^n = E_n$ . 通过归纳地应用 (1.1), 令  $k = n, n - 1, \dots, 1$ , 我们得到

$$\mathbf{x}_1 := (A - \lambda_n I)\mathbf{x} \in E_{n-1},$$

$$\mathbf{x}_2 := (A - \lambda_{n-1} I)\mathbf{x}_1 = (A - \lambda_{n-1} I)(A - \lambda_n I)\mathbf{x} \in E_{n-2},$$

...

$$\mathbf{x}_n := (A - \lambda_2 I)\mathbf{x}_{n-1} = (A - \lambda_2 I) \dots (A - \lambda_{n-1} I)(A - \lambda_n I)\mathbf{x} \in E_1.$$

最后一个包含关系意味着  $\mathbf{x}_n = \alpha\mathbf{e}_1$ . 但是  $(A - \lambda_1 I)\mathbf{e}_1 = 0$ , 所以

$$\mathbf{0} = (A - \lambda_1 I)\mathbf{x}_n = (A - \lambda_1 I)(A - \lambda_2 I) \dots (A - \lambda_n I)\mathbf{x}.$$

因此,  $p(A)\mathbf{x} = \mathbf{0}$  对所有  $\mathbf{x} \in \mathbb{C}^n$  成立, 这意味着  $p(A) = \mathbf{0}$ . □

## 练习

1.1 (可对角化矩阵的凯莱-哈密顿定理)。如上节讨论, 凯莱-哈密顿定理说明, 如果  $A$  是一个方阵, 且其特征多项式为

$$p(\lambda) = \det(A - \lambda I) = \sum_{k=0}^n c_k \lambda^k,$$

则  $p(A) := \sum_{k=0}^n c_k A^k = \mathbf{0}$  (我们假设  $A^0 = I$ )。

证明该定理在  $A$  与一个对角矩阵相似的特殊情况下的表述, 即当  $A = SDS^{-1}$  时。

提示: 如果  $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  且  $p$  是任意多项式, 你能计算  $p(D)$  吗?  $p(A)$  呢?

## 2. 谱映射定理

## 2.1. 算子的多项式

同样需要回忆一下, 对于一个方阵 (算子)  $A$  和一个多项式  $p(z) = \sum_{k=0}^N a_k z^k$ , 算子  $p(A)$  是通过将独立变量替换为  $A$  来定义的,

$$p(A) := \sum_{k=0}^N a_k A^k = a_0 I + a_1 A + a_2 A^2 + \cdots + a_N A^N,$$

这里我们约定  $A^0 = I$ 。

我们知道, 一般而言, 矩阵乘法不是可交换的, 即通常  $AB \neq BA$ , 所以顺序很重要。然而,

$$A^k A^j = A^j A^k = A^{k+j},$$

由此很容易证明对于任意多项式  $p$  和  $q$ ,

$$p(A)q(A) = q(A)p(A) = R(A),$$

其中  $R(z) = p(z)q(z)$ 。

这意味着, 当只处理一个算子  $A$  的多项式时, 不必担心交换性, 可以将  $A$  看作是一个独立的 (标量) 变量。特别地, 如果一个多项式  $p(z)$  可以表示为单项式的乘积

$$p(z) = a(z - z_1)(z - z_2) \cdots (z - z_N),$$

其中  $z_1, z_2, \dots, z_N$  是  $p$  的根, 那么  $p(A)$  可以表示为

$$p(A) = a(A - z_1 I)(A - z_2 I) \cdots (A - z_N I).$$

## 2.2. 谱映射定理

回顾一下, 一个方阵 (算子)  $A$  的谱  $\sigma(A)$  是  $A$  的所有特征值 (不计重数) 的集合。

**定理 2.1.** (谱映射定理) 对于一个方阵  $A$  和任意多项式  $p$ ,

$$\sigma(p(A)) = p(\sigma(A)).$$

换句话说,  $\mu$  是  $p(A)$  的一个特征值, 当且仅当  $\mu = p(\lambda)$  对某个  $A$  的特征值  $\lambda$  成立。

注意, 如表述所示, 这个定理没有说明特征值的重数。

**注记** 需要注意的是, 一个包含关系是平凡的。即, 如果  $\lambda$  是  $A$  的一个特征值,  $A\mathbf{x} = \lambda\mathbf{x}$  对某个  $\mathbf{x} \neq \mathbf{0}$  成立, 那么  $A^k\mathbf{x} = \lambda^k\mathbf{x}$ , 并且  $p(A)\mathbf{x} = p(\lambda)\mathbf{x}$ , 所以  $p(\lambda)$  是  $p(A)$  的一个特征值。这意味着包含关系  $p(\sigma(A)) \subset \sigma(p(A))$  是平凡的。

如果我们考虑上述定理中的  $\mu = 0$  的特殊情况, 我们得到以下推论。

**推论 2.2** 设  $A$  是一个方阵, 其特征值为  $\lambda_1, \lambda_2, \dots, \lambda_n$ , 且  $p$  是一个多项式。那么,  $p(A)$  是可逆的, 当且仅当

$$p(\lambda_k) \neq 0 \quad \forall k = 1, 2, \dots, n.$$

**定理 2.1 的证明** 如上所述, 包含关系

$$p(\sigma(A)) \subset \sigma(p(A))$$

是平凡的。

为了证明另一个包含关系  $\sigma(p(A)) \subset p(\sigma(A))$ , 取一个值  $\mu \in \sigma(p(A))$ . 令  $q(z) = p(z) - \mu$ , 则  $q(A) = p(A) - \mu I$ . 由于  $\mu \in \sigma(p(A))$ , 算子  $q(A) = p(A) - \mu I$  是不可逆的。

我们将多项式  $q(z)$  表示为单项式的乘积:

$$q(z) = a(z - z_1)(z - z_2) \dots (z - z_N).$$

那么, 如第 2.1 节所讨论的, 我们可以将  $q(A)$  表示为

$$q(A) = a(A - z_1 I)(A - z_2 I) \dots (A - z_N I).$$

算子  $q(A)$  是不可逆的, 因此其中一个因子  $A - z_k I$  必须是不可逆的 (因为可逆变换的乘积总是可逆的)。这意味着  $z_k \in \sigma(A)$ 。

另一方面,  $z_k$  是  $q$  的一个根, 所以

$$0 = q(z_k) = p(z_k) - \mu,$$

因此  $\mu = p(z_k)$ . 所以我们证明了包含关系  $\sigma(p(A)) \subset p(\sigma(A))$ . □

### 练习

- 2.1. 一个算子  $A$  被称为**幂零**的, 如果  $\exists k \in \mathbb{N}_+$ , 使得  $A^k = \mathbf{0}$  成立。证明如果  $A$  是幂零的, 那么  $\sigma(A) = \{0\}$  (即 0 是  $A$  的唯一特征值)。  
你能不使用谱映射定理来做到这一点吗?

### 3. 广义特征子空间 代数重数的几何意义

#### 3.1. 不变子空间

**定义** 设  $A: V \rightarrow V$  是向量空间  $V$  上的一个算子 (线性变换)。子空间  $E$  被称为算子  $A$  的**不变子空间** (invariant subspace) (或简称,  $A$ -不变) 如果  $AE \subset E$ , 即如果  $\mathbf{v} \in E$  的所有向量  $A\mathbf{v}$  都属于  $E$ .

如果  $E$  是  $A$ -不变的, 那么

$$A^2E = A(AE) \subset AE \subset E,$$

即  $E$  是  $A^2$ -不变的。

类似地, 我们可以证明 (例如, 通过归纳法), 如果  $AE \subset E$ , 那么

$$A^k E \subset E \quad \forall k \geq 1.$$

这意味着  $P(A)E \subset E$  对于任何多项式  $p$ , 即:

任何  $A$ -不变子空间  $E$  都是  $p(A)$  的不变子空间。

如果  $E$  是一个  $A$ -不变子空间, 那么对于所有  $\mathbf{v} \in E$ , 结果  $A\mathbf{v}$  也属于  $E$ . 因此, 我们可以把  $A$  作为作用在  $E$  上的算子来处理, 而不是作用在整个空间  $V$  上。形式上, 对于一个  $A$ -不变子空间  $E$ , 我们定义  $A$  到  $E$  上的**限制** (restriction)  $A|_E: E \rightarrow E$  为

$$(A|_E)\mathbf{v} = A\mathbf{v} \quad \forall \mathbf{v} \in E.$$

这里我们改变了算子的定义域和目标空间, 但将值赋给自变量的规则保持不变。

我们将需要以下简单引理:

**引理 3.1.** 设  $p$  是一个多项式, 且  $E$  是一个  $A$ -不变子空间。则  $p(A|_E) = p(A)|_E$ .

**证明:** 证明是平凡的。 □

如果  $E_1, E_2, \dots, E_r$  是  $A$ -不变子空间的基, 并且  $A_k := A|_{E_k}$  是相应的限制, 那么由于  $AE_k = A_k E_k \subset E_k$ , 算子  $A_k$  独立地相互作用 (不交互), 为了分析  $A$  的作用, 我们可以分别分析算子  $A_k$ .

特别地, 如果我们选择每个子空间  $E_k$  中的一组基, 并将它们连接起来得到  $V$  中的一组基 (见第四章定理 2.6), 那么在基下算子  $A$  将具有以下块对角形式:

$$A = \begin{pmatrix} A_1 & & & \mathbf{0} \\ & A_2 & & \\ & & \ddots & \\ \mathbf{0} & & & A_r \end{pmatrix}.$$



(当然, 这里我们对  $V$  中的基进行了正确的排序, 首先在一组基  $E_1$  中, 然后是另一组基  $E_2$  等等)。

我们现在的目标是选择不变子空间  $E_1, E_2, \dots, E_r$  的基, 使得限制  $A_k$  具有简单的结构。在这种情况下, 我们将得到一组基, 其中  $A$  的矩阵具有简单的结构。

特征子空间  $\text{Ker}(A - \lambda_k I)$  将是好的候选者, 因为  $A$  在特征子空间  $\text{Ker}(A - \lambda_k I)$  上的限制仅仅是  $\lambda_k I$ . 不幸的是, 如我们所知, 特征子空间并不总是构成一组基 (当且仅当  $A$  可对角化时, 它们才构成一组基, 参见第四章定理 2.1)。

然而, 所谓的广义特征子空间将起作用。

### 3.2. 广义特征子空间

#### 定义 3.2

向量  $\mathbf{v}$  被称为**广义特征向量**(generalized eigenvector) (对应于特征值  $\lambda$ ), 如果  $(A - \lambda I)^k \mathbf{v} = 0$  对某个整数  $k \geq 1$  成立。

所有广义特征向量与 0 的集合被称为**广义特征子空间**(generalized eigenspace) (对应于特征值  $\lambda$ )。

换句话说, 广义特征子空间  $E_\lambda$  可以表示为

$$(3.1) \quad E_\lambda = \bigcup_{k \geq 1} \text{Ker}(A - \lambda I)^k.$$

子空间序列  $\text{Ker}(A - \lambda I)^k, k = 1, 2, 3, \dots$  是一个递增的子空间序列, 即

$$\text{Ker}(A - \lambda I)^k \subset \text{Ker}(A - \lambda I)^{k+1} \quad \forall k \geq 1.$$

(3.1) 这个表示看起来并不太简单, 因为它涉及无限并集。然而, 子空间  $\text{Ker}(A - \lambda I)^k$  的序列会**稳定**, 即

$$\text{Ker}(A - \lambda I)^k = \text{Ker}(A - \lambda I)^{k+1} \quad \forall k \geq k_\lambda,$$

所以实际上可以取有限并集。

为了证明核序列的稳定性, 让我们注意到, 如果对于有限维子空间  $E$  和  $F$ , 我们有  $E \subsetneq F$  (真子集符号  $E \subsetneq F$  表示  $E \subset F$  但  $E \neq F$ ), 那么  $\dim E < \dim F$ .

由于  $\dim \text{Ker}(A - \lambda I)^k \leq \dim V < \infty$ , 它不能无限增长, 所以某处

$$\text{Ker}(A - \lambda I)^k = \text{Ker}(A - \lambda I)^{k+1}.$$

其余部分遵循以下引理。

**引理 3.3.** 如果对某个  $k$ ,

$$\text{Ker}(A - \lambda I)^k = \text{Ker}(A - \lambda I)^{k+1}.$$

则

$$\text{Ker}(A - \lambda I)^{k+r} = \text{Ker}(A - \lambda I)^{k+r+1} \quad \forall r \geq 0.$$

**证明：** 设  $\mathbf{v} \in \text{Ker}(A - \lambda I)^{k+r+1}$ , 即  $(A - \lambda I)^{k+r+1}\mathbf{v} = \mathbf{0}$ . 那么

$$\mathbf{w} := (A - \lambda I)^r \mathbf{v} \in \text{Ker}(A - \lambda I)^{k+1}.$$

但我们知道  $\text{Ker}(A - \lambda I)^k = \text{Ker}(A - \lambda I)^{k+1}$ , 所以  $\mathbf{w} \in \text{Ker}(A - \lambda I)^k$ , 这意味着  $(A - \lambda I)^k \mathbf{w} = \mathbf{0}$ . 回忆  $\mathbf{w}$  的定义, 我们得到

$$(A - \lambda I)^{k+r} \mathbf{v} = (A - \lambda I)^k \mathbf{w} = \mathbf{0},$$

所以  $\mathbf{v} \in \text{Ker}(A - \lambda I)^{k+r}$ . 我们证明了  $\text{Ker}(A - \lambda I)^{k+r+1} \subset \text{Ker}(A - \lambda I)^{k+r}$ . 而反向包含关系是平凡的.  $\square$

**定义** 序列  $\text{Ker}(A - \lambda I)^k$  稳定的数字  $d = d(\lambda)$ , 即满足

$$\text{Ker}(A - \lambda I)^{d-1} \subsetneq \text{Ker}(A - \lambda I)^d = \text{Ker}(A - \lambda I)^{d+1}$$

的数字  $d$ , 被称为特征值  $\lambda$  的**深度 (depth)**。

从深度的定义可以得出, 对于广义特征子空间  $E_\lambda$ ,

$$(3.2) \quad (A - \lambda I)^{d(\lambda)} \mathbf{v} = \mathbf{0} \quad \forall \mathbf{v} \in E_\lambda.$$

现在总结一下我们对广义特征子空间的了解。

a)  $E_\lambda$  是  $A$  的一个不变子空间,  $AE_\lambda \subset E_\lambda$ .

b) 如果  $d(\lambda)$  是特征值  $\lambda$  的深度, 那么

$$((A - \lambda I)|_{E_\lambda})^{d(\lambda)} = (A|_{E_\lambda} - \lambda I_{E_\lambda})^{d(\lambda)} = \mathbf{0}.$$

(这只是 (3.2) 的另一种写法)

c)  $\sigma(A|_{E_\lambda}) = \{\lambda\}$ , 因为算子  $A|_{E_\lambda} - \lambda I_{E_\lambda}$  是幂零的, 见 2, 而幂零算子的谱只包含一个点 0, 见问题 2.1。

现在我们准备陈述本节的主要结果。设  $A: V \rightarrow V$ .

**定理 3.4.** 设  $\sigma(A)$  包含  $r$  个点  $\lambda_1, \lambda_2, \dots, \lambda_r$ , 设  $E_k := E_{\lambda_k}$  为相应的广义特征子空间。那么子空间系统  $E_1, E_2, \dots, E_r$  是  $V$  的一组基 (的子空间)。

**注记 3.5** 如果我们连接所有广义特征子空间  $E_k$  的基, 那么根据第四章定理 2.6, 我们将得到空间  $V$  的一组基。在该基下, 算子  $A$  的矩阵具有块对角形式:  $A = \text{diag}\{A_1, A_2, \dots, A_r\}$ , 其中  $A_k := A|_{E_k}$ ,  $E_k = E_{\lambda_k}$ . 也很容易看出 (见 (3.2)) 算子  $N_k := A_k - \lambda_k I_{E_k}$  是幂零的,  $N_k^{d_k} = 0$ .

**定理 3.4 的证明** 设  $m_k$  是特征值  $\lambda_k$  的重数, 所以  $p(z) = \prod_{k=1}^r (z - \lambda_k)^{m_k}$  是  $A$  的特征多项式。定义

$$p_k(z) = \frac{p(z)}{(z - \lambda_k)^{m_k}} = \prod_{j \neq k} (z - \lambda_j)^{m_j}.$$

□

**引理 3.6.**

$$(3.3) \quad (A - \lambda_k I)^{m_k}|_{E_k} = \mathbf{0},$$

**证明:** 有两种简单的证明方法。第一个是注意到  $m_k \geq d_k$ , 其中  $d_k$  是特征值  $\lambda_k$  的深度, 并利用事实

$$(A - \lambda_k I)^{d_k}|_{E_k} = (A|_{E_k} - \lambda_k I_{E_k})^{m_k} = \mathbf{0},$$

其中  $A_k := A|_{E_k}$  (广义特征子空间的性质 2)。

第二种可能性是注意到根据谱映射定理, 见推论 2.2, 算子  $p_k(A)|_{E_k} = p_k(A_k)$  是可逆的。根据凯莱-哈密顿定理 (定理 1.1),

$$\mathbf{0} = p(A) = (A - \lambda_k I)^{m_k} p_k(A),$$

将所有算子限制到  $E_k$  上我们得到

$$\mathbf{0} = p(A_k) = (A_k - \lambda_k I_{E_k})^{m_k} p_k(A_k),$$

所以

$$(A_k - \lambda_k I_{E_k})^{m_k} = p(A_k) p_k(A_k)^{-1} = \mathbf{0} p_k(A_k)^{-1} = \mathbf{0}.$$

□

为了证明定理, 定义

$$q(z) = \sum_{k=1}^r p_k(z).$$

由于  $p_k(\lambda_j) = 0$  对  $j \neq k$  且  $p_k(\lambda_k) \neq 0$ , 我们可以得出  $q(\lambda_k) \neq 0$  对所有  $k$  成立。因此, 根据谱映射定理, 见推论 2.2, 算子

$$B = q(A)$$

是可逆的。

注意,  $BE_k \subset E_k$  (任何  $A$ -不变子空间也必然是  $p(A)$ -不变的)。由于  $B$  是一个可逆算子,  $\dim(BE_k) = \dim E_k$ , 这与  $BE_k \subset E_k$  一起意味着  $BE_k = E_k$ 。将最后一个恒等式乘以  $B^{-1}$  得到  $B^{-1}E_k = E_k$ , 即  $E_k$  是  $B^{-1}$  的不变子空间。

还需注意, 从 (3.3) 可知,

$$p_k(A)|_{E_j} = \mathbf{0} \quad \forall j \neq k,$$

因为  $p_k(A)|_{E_j} = p_k(A_j)$  并且  $p_k(A_j)$  包含因子  $(A_j - \lambda_j I_{E_j})^{m_j} = \mathbf{0}$ .

定义算子  $\mathcal{P}_k$  为

$$\mathcal{P}_k := B^{-1}p_k(A).$$

**引理 3.7.** 对于上述定义的算子  $\mathcal{P}_k$ :

- a)  $\mathcal{P}_1 + \mathcal{P}_2 + \cdots + \mathcal{P}_r = I$ ;
- b)  $\mathcal{P}_k|_{E_j} = \mathbf{0}$  对  $j \neq k$ ;
- c)  $\text{Ran } \mathcal{P}_k \subset E_k$ ;
- d) 更进一步,  $\mathcal{P}_k \mathbf{v} = \mathbf{v} \quad \forall \mathbf{v} \in E_k$ , 所以实际上  $\text{Ran } \mathcal{P}_k = E_k$ .

**证明:** 第一个性质是平凡的:

$$\sum_{k=1}^r \mathcal{P}_k = B^{-1} \sum_{k=1}^r p_k(A) = B^{-1}B = I.$$

第二个性质来自 (3.3)。实际上,  $p_k(A)$  包含因子  $(A - \lambda_j)^{m_j}$ , 将其限制到  $E_j$  上为零。因此,  $p_k(A)|_{E_j} = \mathbf{0}$ , 从而  $\mathcal{P}_k|_{E_j} = B^{-1}p_k(A)|_{E_j} = \mathbf{0}$ .

为了证明第三个性质, 根据凯莱-哈密顿定理,  $p(A) = \mathbf{0}$ . 由于  $p(z) = (z - \lambda_k)^{m_k} p_k(z)$ , 我们得到对于  $\mathbf{w} = p_k(A)\mathbf{v}$ ,

$$(A - \lambda_k I)^{m_k} \mathbf{w} = (A - \lambda_k I)^{m_k} p_k(A)\mathbf{v} = p(A)\mathbf{v} = \mathbf{0}.$$

这意味着,  $\text{Ran } p_k(A)$  中的任何向量  $\mathbf{w}$  都被  $(A - \lambda_k I)$  的某个幂化为了零, 根据定义, 这意味着  $\text{Ran } p_k(A) \subset E_k$ .

为了证明最后个性质, 让我们注意到从 (3.3) 可以得出, 对于  $\mathbf{v} \in E_k$ ,

$$p_k(A)\mathbf{v} = \sum_{j=1}^r p_j(A)\mathbf{v} = B\mathbf{v},$$

这使得  $\mathcal{P}_k \mathbf{v} = B^{-1}B\mathbf{v} = \mathbf{v}$ . □

现在我们准备完成定理的证明。取  $\mathbf{v} \in V$ , 定义  $\mathbf{v}_k = \mathcal{P}_k \mathbf{v}$ . 那么根据引理 3.7 的陈述 c),  $\mathbf{v}_k \in E_k$ , 并且根据陈述 a),

$$\mathbf{v} = \sum_{k=1}^r \mathbf{v}_k,$$

所以  $\mathbf{v}$  允许表示为线性组合。

为了证明这个表示是唯一的, 我们可以注意到, 如果  $\mathbf{v}$  被表示为  $\mathbf{v} = \sum_{k=1}^r \mathbf{v}_k$ , 其中  $\mathbf{v}_k \in E_k$ , 那么根据引理 3.7 的陈述 b) 和 d),

$$\mathcal{P}_k \mathbf{v} = \mathcal{P}_k (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r) = \mathcal{P}_k \mathbf{v}_k = \mathbf{v}_k.$$

□

### 3.3. 代数重数的几何意义

**命题 3.8** 一个特征值的代数重数等于对应广义特征子空间的维数。

**证明：**根据注记 3.5，如果我们连接广义特征子空间  $E_k = E_{\lambda_k}$  的基得到整个空间的一组基，那么在该基下算子  $A$  的矩阵具有块对角形式  $\text{diag}\{A_1, A_2, \dots, A_r\}$ ，其中  $A_k := A|_{E_k}$ 。算子  $N_k = A_k - \lambda_k I_{E_k}$  是幂零的，所以  $\sigma(N_k) = \{0\}$ 。因此，算子  $A_k$  (回忆  $A_k = N_k + \lambda_k I$ ) 的谱只包含一个特征值  $\lambda_k$ ，其代数重数为  $n_k = \dim E_k$ 。重数等于  $n_k$ ，因为一个有限维空间  $V$  中的算子有恰好  $\dim V$  个特征值 (计入重数)，而  $A_k$  只有一个特征值。

注意，我们可以自由选择  $E_k$  中的基，所以我们选择它们使得相应的块  $A_k$  是上三角的。那么

$$\det(A - \lambda I) = \prod_{k=1}^r \det(A_k - \lambda I_{E_k}) = \prod_{k=1}^r (\lambda_k - \lambda)^{n_k}.$$

但这表示特征值  $\lambda_k$  的代数重数是  $n_k = \dim E_{\lambda_k}$ 。  $\square$

### 3.4. 一个重要的应用

以下推论对于有关微分方程的理论非常重要。

**推论 3.9** 任何算子  $A$  在  $V$  中都可以表示为  $A = D + N$ ，其中  $D$  是对角化的 (即在某组基下是对角形的) 且  $N$  是幂零的 ( $N^m = \mathbf{0}$  对某个  $m$ )，并且  $DN = ND$ 。

**证明：**如上所述，见注记 3.5，如果我们连接广义特征子空间  $E_k$  的基得到整个空间的一组基，那么在该基下  $A$  具有块对角形式  $A = \text{diag}\{A_1, A_2, \dots, A_r\}$ ，其中  $A_k := A|_{E_k}$ 。算子  $N_k = A_k - \lambda_k I_{E_k}$  是幂零的，并且算子  $D = \text{diag}\{\lambda_1 I_{E_1}, \lambda_2 I_{E_2}, \dots, \lambda_r I_{E_r}\}$  是对角的 (在该基下)。还需注意， $\lambda_k I_{E_k} N_k = N_k \lambda_k I_{E_k}$  (恒等 (单位) 算子与任何算子可交换)，所以块对角算子  $N = \text{diag}\{N_1, N_2, \dots, N_r\}$  与  $D$  可交换， $DN = ND$ 。因此，定义  $N$  为块对角算子  $N = \text{diag}\{N_1, N_2, \dots, N_r\}$ ，我们得到所需的分解。  $\square$

这个推论允许我们计算算子的函数。让我们回顾一下，如果  $p$  是一个  $d$  次多项式，那么  $p(a+x)$  可以用泰勒公式计算：

$$p(a+x) = \sum_{k=0}^d \frac{p^{(k)}(a)}{k!} x^k.$$

这是一个代数恒等式，意味着对于每个多项式  $p$ ，我们可以通过对  $a$  和  $x$  进行形式代数运算而不关心它们的特殊性质是否会改变该公式的正确性。

由于算子  $D$  和  $N$  可交换， $DN = ND$ ，因此它们遵循与普通 (标量) 变量相

同的规则，故我们可以这样写（通过用  $D$  替换  $a$  并用  $N$  替换  $x$ ）：

$$p(A) = p(D + N) = \sum_{k=0}^d \frac{p^{(k)}(D)}{k!} N^k.$$

这里，为了计算导数  $p^{(k)}(D)$ ，我们首先通过（普通的微积分）规则计算多项式  $p(x)$  的  $k$  阶导数，然后将  $x$  替换为  $D$ 。

但是由于  $N$  是幂零的， $N^m = \mathbf{0}$  对某个  $m$  成立，只有前  $m$  项可能非零，所以

$$p(A) = p(D + N) = \sum_{k=0}^{m-1} \frac{p^{(k)}(D)}{k!} N^k.$$

如果  $m$  远小于  $d$ ，这个公式将使  $p(A)$  的计算容易得多。

同样的方法也适用于  $p$  不是多项式，而是无穷幂级数的情况。对于一般的幂级数，我们必须注意所有级数的收敛性，所以我们不能说这个公式对任意幂级数  $p(x)$  都成立。然而，如果幂级数的收敛半径是  $\infty$ ，那么上文中一切都是正确的。特别地，如果  $p(x) = e^x$ ，那么使用  $(e^x)' = e^x$  的事实，我们得到：

$$e^A = e^{D+N} = \sum_{k=0}^{m-1} \frac{e^{(k)}(D)}{k!} N^k = \sum_{k=0}^{m-1} \frac{e^D}{k!} N^k = e^D \sum_{k=0}^{m-1} \frac{1}{k!} N^k.$$

这个公式在微分方程中具有重要的应用。

请注意， $ND = DN$  的事实在这里是非常重要的！

#### 4. 幂零算子的结构

回想一下，向量空间  $V$  中的一个算子  $A$  被称为**幂零**的，如果  $\exists k \in \mathbb{N}_+$ ，使得  $A^k = \mathbf{0}$  成立。

在上一节中，我们证明了（见注记 3.5），如果我们连接所有广义特征子空间  $E_k = E_{\lambda_k}$  的基得到空间  $V$  的一组基，那么算子  $A$  在该基下的矩阵具有块对角形式  $\text{diag}\{A_1, A_2, \dots, A_r\}$ ，并且算子  $A_k$  可以表示为  $A_k = \lambda_k I + N_k$ ，其中  $N_k$  是幂零算子。

在每个广义特征子空间  $E_k$  中，我们想选择一组基，使得  $A_k$  在该基下的矩阵具有最简单的形式。由于恒等算子在任何基下的矩阵都是单位矩阵，我们需要找到一组基，使得幂零算子  $N_k$  具有简单的形式。

由于我们可以分别处理每个  $N_k$ ，我们将需要考虑以下问题：

对于一个幂零算子  $A$ ，能否找到一组基，使得该算子在该基下的矩阵是简单的？

让我们看看，一个矩阵具有简单形式意味着什么。很容易看出以下矩阵

$$(4.1) \quad \begin{pmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & 0 & 1 \\ 0 & & & & 0 \end{pmatrix}$$

是幂零的。

这些矩阵（以及  $1 \times 1$  的零矩阵）将是我们的“构建模块”。即，我们将证明对于任何幂零算子，都可以找到一组基，使得算子在该基下的矩阵具有块对角形式  $\text{diag}\{A_1, A_2, \dots, A_r\}$ ，其中每个  $A_k$  要么是形式 (4.1) 的块，要么是  $1 \times 1$  的零块。

让我们看看我们应该寻找什么。假设一个算子  $A$  在基  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  下的矩阵是 (4.1) 的形式。那么

$$(4.2) \quad A\mathbf{v}_1 = \mathbf{0}$$

并且

$$(4.3) \quad A\mathbf{v}_{k+1} = \mathbf{v}_k, \quad k = 1, 2, \dots, p-1.$$

因此，我们必须寻找满足上述关系 (4.2)、(4.3) 的向量链  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ 。

#### 4.1. 广义特征向量循环

**定义** 设  $A$  是一个幂零算子。满足关系 (4.2)、(4.3) 的非零向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  的链被称为  $A$  的**广义特征向量循环** (cycle of generalized eigenvectors)。向量  $\mathbf{v}_1$  被称为循环的**初始向量** (initial vector)，向量  $\mathbf{v}_p$  被称为循环的**末端向量** (end vector)，并且数字  $p$  被称为循环的**长度** (length)。

**注记** 对于任意算子，也可以做出类似的定义。那么所有向量  $\mathbf{v}_k$  必须属于同一个广义特征子空间  $E_\lambda$ ，并且它们必须满足恒等式

$$(A - \lambda I)\mathbf{v}_1 = \mathbf{0}, \quad (A - \lambda I)\mathbf{v}_{k+1} = \mathbf{v}_k, \quad k = 1, 2, \dots, p-1.$$

**定理 4.1.** 设  $A$  是一个幂零算子，设  $C_1, C_2, \dots, C_r$  是其广义特征向量循环， $C_k = \{\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_{p_k}^k\}$ ，其中  $p_k$  是循环  $C_k$  的长度。假设初始向量  $\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^r$  是线性无关的。那么没有向量属于两个循环，并且所有向量组成的集合是线性无关的。

**证明：** 设  $n = p_1 + p_2 + \dots + p_r$  是所有循环中向量的总数<sup>53</sup>。我们将使用  $n$  的归纳法。如果  $n = 1$ ，则定理是平凡的。

<sup>53</sup>我们只是数每个循环中的向量，并将所有数字相加。我们不关心是否有些循环有共同的向量，我们将其计入它所属的每个循环（当然，根据定理，这是不可能的，但一开始我们不能假设这一点）

现在假设定理对于所有算子和所有循环集合成立, 只要所有循环中向量的总数严格小于  $n$ .

不失一般性, 我们可以假设向量  $\mathbf{v}_j^k$  构成整个空间  $V$  的基, 否则我们可以考虑算子  $A$  限制在不变子空间  $\text{span}\{\mathbf{v}_j^k : k = 1, 2, \dots, r, 1 \leq j \leq p_k\}$  上。

考虑子空间  $\text{Ran } A$ . 从关系 (4.2)、(4.3) 可知, 向量  $\mathbf{v}_j^k : k = 1, 2, \dots, r, 1 \leq j \leq p_k - 1$  构成  $\text{Ran } A$  的基. 注意, 如果  $p_k > 1$ , 那么系统  $\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_{p_k-1}^k$  是一个循环, 并且  $A$  使长度为 1 的任何循环的向量零化。

因此, 我们有有限数量的循环, 并且这些循环的初始向量是线性无关的, 所以归纳假设适用, 并且向量  $\mathbf{v}_j^k : k = 1, 2, \dots, r, 1 \leq j \leq p_k - 1$  是线性无关的。

由于这些向量也张成  $\text{Ran } A$ , 我们在那里有了一组基. 因此,

$$\text{rank } A = \dim \text{Ran } A = n - r.$$

(我们有  $n$  个向量, 并从每个循环  $\mathcal{C}_k$  中移除了一个向量  $\mathbf{v}_{p_k}^k$ , 其中  $k = 1, 2, \dots, r$ . 因此, 我们在基  $\mathbf{v}_j^k : k = 1, 2, \dots, r, 1 \leq j \leq p_k - 1$  中有  $n - r$  个向量)。另一方面,  $A\mathbf{v}_1^k = 0$  对  $k = 1, 2, \dots, r$  成立, 并且由于这些向量是线性无关的,  $\dim \text{Ker } A \geq r$ . 根据秩定理 (第二章定理 7.1),

$$\dim V = \text{rank } A + \dim \text{Ker } A = (n - r) + \dim \text{Ker } A \geq (n - r) + r = n,$$

所以  $\dim V \geq n$ .

另一方面,  $V$  由  $n$  个向量张成, 因此向量  $\mathbf{v}_j^k : k = 1, 2, \dots, r, 1 \leq j \leq p_k$  构成一组基, 所以它们是线性无关的.  $\square$

## 4.2. 幂零算子的若尔当标准型

**定理 4.2.** 设  $A : V \rightarrow V$  是一个幂零算子. 那么  $V$  有一组由  $A$  的广义特征向量循环组成的集合构成的基。

**证明:** 我们将使用  $n = \dim V$  的归纳法. 对于  $n = 1$ , 定理是平凡的。

假设定理对于任何作用在维数小于  $n$  的空间中的算子都成立. 考虑子空间  $X = \text{Ran } A$ .

$X$  是算子  $A$  的一个不变子空间, 所以我们可以考虑限制  $A|_X$ .

由于  $A$  是不可逆的,  $\dim \text{Ran } A < \dim V$ , 所以根据归纳假设存在广义特征向量循环  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_r$ , 使得它们的并集是  $X$  中的一组基. 设  $\mathcal{C}_k = \{\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_{p_k}^k\}$ , 其中  $\mathbf{v}_1^k$  是循环的初始向量。

由于末端向量  $\mathbf{v}_{p_k}^k$  属于  $\text{Ran } A$ , 可以找到一个向量  $\mathbf{v}_{p_k+1}^k$  使得  $A\mathbf{v}_{p_k+1}^k = \mathbf{v}_{p_k}^k$ . 因此, 我们可以将每个循环  $\mathcal{C}_k$  扩展成一个更大的循环  $\tilde{\mathcal{C}}_k = \{\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_{p_k}^k, \mathbf{v}_{p_k+1}^k\}$ . 由



于循环  $\tilde{C}_k, k = 1, 2, \dots, r$  的初始向量  $\mathbf{v}_1^k$  是线性无关的, 上述定理 4.1 暗示了这些循环的并集是一个线性无系统。

根据循环的定义, 我们有  $\mathbf{v}_1^k \in \text{Ker } A$ , 并且我们假设初始向量  $\mathbf{v}_1^k, k = 1, 2, \dots, r$  是线性无关的。让这个系统扩展成  $\text{Ker } A$  中的一组基, 即找到向量  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q$ , 使得系统  $\{\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^r, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q\}$  是  $\text{Ker } A$  中的一组基 (可能发生的情况是系统  $\mathbf{v}_1^k, k = 1, 2, \dots, r$  已经是  $\text{Ker } A$  中的一组基, 在这种情况下, 我们让  $q = 0$ , 这并没有添加任何内容)。

向量  $\mathbf{u}_j$  可以被视为长度为 1 的循环, 因此我们有一个循环集合  $\{\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_r, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q\}$ , 其初始向量是线性无关的。所以, 我们可以应用定理 4.1 得到所有这些循环的并集是一个线性无系统。

为了证明它是一组基, 让我们计算维数。我们知道循环  $C_1, C_2, \dots, C_r$  总共有  $\dim \text{Ran } A = \text{rank } A$  个向量。每个循环  $\tilde{C}_k$  是从  $C_k$  通过添加 1 个向量得到的, 所以所有循环  $\tilde{C}_k$  中的向量总数是  $\text{rank } A + r$ 。

我们知道  $\dim \text{Ker } A = r + q$  (因为  $\{\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^r, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q\}$  是那里的一组基)。我们将循环  $\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_r$  添加了额外的  $q$  个向量, 所以我们得到了

$$\text{rank } A + r + q = \text{rank } A + \dim \text{Ker } A = \dim V$$

个线性无关的向量。但是  $\dim V$  个线性无关的向量构成一组基。□

**定义** 由幂零算子  $A$  的广义特征向量循环的并集构成的基 (其存在由定理 4.2 保证) 称为  $A$  的**若尔当标准基**(Jordan canonical basis)。

注意, 这样的基不是唯一的。

**推论 4.3** 设  $A$  是一个幂零算子。存在一组基 (若尔当标准基), 使得该算子在该基下的矩阵是块对角的  $\text{diag}\{A_1, A_2, \dots, A_r\}$ , 其中所有  $A_k$  (可能有一个例外) 是形式 (4.1) 的块, 并且其中一个块  $A_k$  可以是零。

算子在若尔当标准基下的矩阵称为该算子的**若尔当标准型**。我们稍后会看到, 如果我们要约定块的顺序 (即约定基中的向量顺序), 那么若尔当标准型是唯一的。

**定理 4.3 的证明** 根据定理 4.2, 可以找到一组由广义特征向量循环的并集构成的基。大小为  $p$  的循环产生一个  $p \times p$  的对角块, 形式为 (4.1), 而长度为 1 的循环对应于一个  $1 \times 1$  的零块。我们可以将这些  $1 \times 1$  的零块连接成一个大的零块 (因为非对角线元素是 0)。□

### 4.3. 点图 若尔当标准型的唯一性

有一种很好的方法来可视化定理 4.2 和推论 4.3，即所谓的**点图**。这种方法还可以让我们回答许多自然问题，例如“由推论 4.3 给出的块对角表示是否是唯一的？”

当然，如果我们字面上处理这个问题，答案是“否”，因为我们可以随时更改块的顺序。但是，如果我们排除这种微不足道的情况，例如约定某种块的顺序（比如，如果我们把所有非零块按降序排列，然后放零块），那么这个表示是不是唯一的？

为了更好地理解第 4.1 节中描述的幂零算子的结构，让我们绘制所谓的点图。即，假设我们有一组基，它是广义特征向量循环的并集。让我们用一个元素为点的数组来表示基，这样每一列代表一个循环。第一行由循环的初始向量组成，我们按照长度的降序排列这些列（循环），将最长的放在左边。

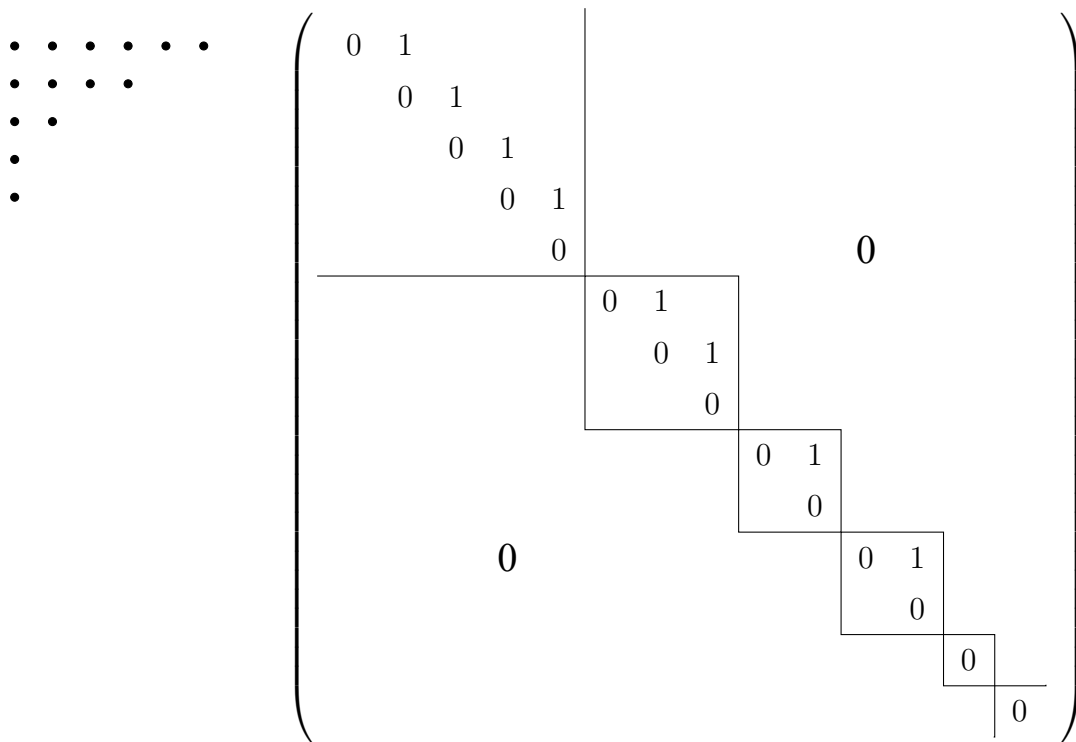


图 0.6 幂零算子的点图和相应的若尔当标准型

在图0.6中，我们有一个幂零算子的点图，以及它的若尔当标准型。这个点图表明，基由一个长度为5的循环，一个长度为3的循环，两个长度为2的循环，以及两个长度为1的循环组成。长度为5的循环对应于矩阵的 $5 \times 5$ 块，长度为3的循环对应于一个 $3 \times 3$ 非零块，两个长度为2的循环对应于两个 $2 \times 2$ 块。三个长度为1的循环对应于对角线上的两个零项。这里，在每个块中，我们只给出主对角线和它上面的对角线；矩阵的所有其他项都是零。

如果我们约定块的顺序, 那么点图与若尔当标准型 (对于幂零算子) 之间存在一一对应关系。因此, 关于若尔当标准型唯一性的问题等同于关于点图唯一性的问题。

为了回答这个问题, 让我们分析算子  $A$  如何变换点图。由于算子  $A$  使循环的初始向量化零, 并将循环中的向量  $\mathbf{v}_{k+1}$  移到向量  $\mathbf{v}_k$ , 我们可以看到算子  $A$  通过删除图的第一行 (顶部) 来作用于其点图。

新的点图对应于  $\text{Ran } A$  中的若尔当标准基, 并且允许我们写出限制  $A|_{\text{Ran } A}$  的若尔当标准型。

类似地, 不难看出算子  $A^k$  删除图的前  $k$  行。因此, 如果对于所有  $k$ , 我们都知维数  $\dim \text{Ker}(A^k)$ , 我们就知道了算子  $A$  的点图。即, 第一行的点的数量是

$$\dim \text{Ker } A,$$

第二行的点的数量是

$$\dim \text{Ker}(A^2) - \dim \text{Ker } A,$$

而第  $k$  行的点的数量是

$$\dim \text{Ker}(A^k) - \dim \text{Ker}(A^{k+1}).$$

但这意味着, 最初使用若尔当标准基定义的点图, 并不取决于该标准基的特定选择。因此, 点图是唯一的! 这意味着如果我们约定块的顺序, 那么若尔当标准型是唯一的。

#### 4.4. 计算若尔当标准基

让我们简单谈谈如何为幂零算子计算若尔当标准基。设  $p_1$  是最大的整数, 使得  $A^{p_1} \neq \mathbf{0}$  (因此  $A^{p_1+1} = \mathbf{0}$ )。从上面对点图的分析可以看出,  $p_1$  是最长循环的长度。

计算算子  $A^k, k = 1, 2, \dots, p_1$ , 并计数  $\dim \text{Ker}(A^k)$ , 我们可以构造  $A$  的点图。现在我们想用向量替换点, 并找到一个构成循环并集的基。

我们从找到最长的循环开始 (因为我们知道点图, 我们知道有多少个循环, 以及每个循环的长度)。考虑  $\text{Ran}(A^{p_1})$  中的一组基。将该基中的向量命名为  $\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^{r_1}$ , 这些将是循环的初始向量。然后通过求解方程

$$A^{p_1} \mathbf{v}_{p_1}^k = \mathbf{v}_1^k, k = 1, 2, \dots, r_1$$

来找到循环的末端向量  $\mathbf{v}_{p_1}^1, \mathbf{v}_{p_1}^2, \dots, \mathbf{v}_{p_1}^{r_1}$ 。通过连续地将算子  $A$  应用于末端向量  $\mathbf{v}_{p_1}^k$ , 我们得到循环中的所有向量  $\mathbf{v}_j^k$ 。因此, 我们构造了所有最大长度的循环。

设  $p_2$  是剩余循环中最大循环的长度。考虑  $\text{Ran}(A^{p_2})$  子空间, 并设  $\dim \text{Ran}(A^{p_2}) = r_2$ . 由于  $\text{Ran}(A^{p_1}) \subset \text{Ran}(A^{p_2})$ , 我们可以将基  $\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^{r_1}$  扩展成  $\text{Ran}(A^{p_2})$  中的一组基  $\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^{r_1}, \mathbf{v}_1^{r_1+1}, \dots, \mathbf{v}_1^{r_2}$ . 然后通过求解方程

$$A^{p_1} \mathbf{v}_{p_2}^k = \mathbf{v}_1^k, \quad k = r_1 + 1, r_1 + 2, \dots, r_2$$

来找到循环  $\mathcal{C}_{r_1+1}, \dots, \mathcal{C}_{r_2}$  的末端向量, 从而构造长度为  $p_2$  的循环。

设  $p_3$  表示剩余循环中最大循环的长度。然后, 通过将基  $\mathbf{v}_1^1, \mathbf{v}_1^2, \dots, \mathbf{v}_1^{r_2}$  在  $\text{Ker}(A^{p_2})$  中扩展成  $\text{Ker}(A^{p_3})$  中的一组基, 我们构造长度为  $p_3$  的循环, 依此类推。

最后还有一个说明: 如上所述, 如果我们知道点图, 我们就知道标准型, 所以一旦我们找到了一组若尔当标准基, 我们就不需要计算该基下算子  $A$  的矩阵: 我们已经知道了!

## 5. 若尔当分解定理

**定理 5.1.** 给定一个算子  $A$ , 存在一组基 (若尔当标准基), 使得  $A$  在这组基下的矩阵具有分块对角形式, 其对角块形如

$$(5.1) \quad \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix}$$

其中  $\lambda$  是  $A$  的一个特征值。这里我们假定大小为 1 的块就是  $\lambda$ 。

定理 5.1 中得到的分块对角形式称为算子  $A$  的**若尔当标准型**。相应的基称为算子  $A$  的**若尔当标准基**。

**定理 5.1 的证明** 根据定理 3.4 和注记 3.5, 如果我们将广义特征空间  $E_k = E_{\lambda_k}$  中的基合并起来, 得到全空间的一组基, 那么算子  $A$  在这组基下的矩阵具有分块对角形式  $\text{diag}\{A_1, A_2, \dots, A_r\}$ , 其中  $A_k = A|_{E_k}$ . 算子  $N_k = A_k - \lambda_k I_{E_k}$  是幂零的, 因此根据定理 4.2 (更确切地说, 根据推论 4.3), 我们可以在  $E_k$  中找到一组基, 使得  $N_k$  在这组基下的矩阵是  $N_k$  的若尔当标准型。要得到  $A_k$  在这组基下的矩阵, 只需将主对角线上的 0 替换为  $\lambda_k$  即可。□

### 5.1. 关于计算若尔当标准基的注记

首先, 我们回顾一下, 计算特征值是最困难的部分, 但我们在这里不讨论这一部分, 并假设特征值已经计算出来了。

对于每个特征值  $\lambda$ , 我们计算子空间  $\text{Ker}(A - \lambda I)^k, k = 1, 2, \dots$ , 直到子空间序列稳定为止。事实上, 由于我们有一个递增的子空间序列 ( $\text{Ker}(A - \lambda I)^k \subset \text{Ker}(A - \lambda I)^{k+1}$ ), 因此只需追踪它们的维数 (或算子  $(A - \lambda I)^k$  的秩) 就足够了。对于一个特征值  $\lambda$ , 令  $m = m_\lambda$  是使得序列  $\text{Ker}(A - \lambda I)^k$  稳定的那个数, 即  $m$  满足

$$\dim \text{Ker}(A - \lambda I)^{m-1} < \dim \text{Ker}(A - \lambda I)^m = \dim \text{Ker}(A - \lambda I)^{m+1}.$$

那么  $E_\lambda = \text{Ker}(A - \lambda I)^m$  就是对应于特征值  $\lambda$  的广义特征空间。

在我们计算出所有广义特征空间后, 有两种可能的方法。第一种方法是在每个广义特征空间中找到一组基, 这样算子  $A$  在这组基下的矩阵就具有分块对角形式  $\text{diag}\{A_1, A_2, \dots, A_r\}$ , 其中  $A_k = A|_{E_{\lambda_k}}$ . 然后我们可以分别处理每个矩阵  $A_k$ . 算子  $N_k = A_k - \lambda_k I$  是幂零的, 因此应用 4.4 节中描述的算法, 我们得到  $N_k$  的若尔当标准表示, 然后将主对角线上的 0 替换为  $\lambda_k$ , 就得到分块  $A_k$  的若尔当标准表示。这种方法的优点是我们处理的是更小的块。但我们需要找到算子在新基下的矩阵, 这涉及到矩阵求逆和矩阵乘法。

另一种方法是直接处理算子  $A$ , 而不是先将其分解为块, 在每个广义特征空间  $E_{\lambda_k}$  中找到若尔当标准基。同样, 我们在 4.4 节中概述的算法稍作修改后仍然适用。也就是说, 在为广义特征空间  $E_{\lambda_k}$  计算若尔当标准基时, 我们不需要像单独处理分块  $A_k$  时那样考虑子空间  $\text{Ran}(A_k - \lambda_k I)^j$ , 而是需要考虑子空间  $(A - \lambda_k I)^j E_{\lambda_k}$ .



## 附录 课后习题解答

译者最后还是借助 AI (Gemini 3.0 Pro) 生成了本书所有课后习题的解答, 然后把了一道关: 压缩、优化了答案的叙述结构, 删除了冗余表达, 并将原始 Markdown 格式整理成书中这样排版良好、可用于打印的 L<sup>A</sup>T<sub>E</sub>X 格式, 以供读者学习参考。因为我发现如果一本书的习题没有配答案, 那对于读者来说真的很劝退, 而且学起来也很不方便, 缺乏实时的反馈和纠偏。

对于这一部分内容, 我虽然对照着之前做过的作业对其进行了仔细的审校, 但难免会出错; AI 也有可能给出不够优秀甚至错误的答案, 受时间和译者水平所限, 也未能及时发现并加以改正, 请读者仔细甄别。总之, 种种问题请读者谅解, 我也将在后续版本中持续优化。如果读者能在开源项目中贡献自己更好的解法, 那就再好不过了。

当然, 如果读者只是照抄上面的答案来糊弄老师和自己, 那这就违背了我制作答案的初衷。

### 第一章习题解答

1.1. 解:

$$\begin{aligned}2\mathbf{x} &= 2(1, 2, 3)^T = (2, 4, 6)^T \\3\mathbf{y} &= 3(y_1, y_2, y_3)^T = (3y_1, 3y_2, 3y_3)^T \\ \mathbf{x} + 2\mathbf{y} - 3\mathbf{z} &= (1, 2, 3)^T + (2y_1, 2y_2, 2y_3)^T - (12, 6, 3)^T \\ &= (1 + 2y_1 - 12, 2 + 2y_2 - 6, 3 + 2y_3 - 3)^T \\ &= (2y_1 - 11, 2y_2 - 4, 2y_3)^T\end{aligned}$$

1.2. 解:

- a) 是。连续函数的和仍是连续函数, 连续函数的标量倍数仍是连续函数, 且包含零函数 (也是连续的)。
- b) 不是。对于非零函数  $f$ , 其加法逆元  $-f$  是非正的 (通常是负的), 因此不在该集合中。
- c) 不是。该集合不包含零多项式 (零多项式的次数通常定义为  $-\infty$  或未定义)。此外, 两个  $n$  次多项式相加可能会消去最高次项, 导致次数降低。
- d) 是。若  $A, B$  是对称的, 则  $(A + B)^T = A^T + B^T = A + B$ , 故和是对称的。同理  $(\alpha A)^T = \alpha A^T = \alpha A$ , 标量乘法封闭。零矩阵满足  $O^T = O$ , 也在集合中。

1.3. 解: a) 正确。根据向量空间公理 3。

b) 错误。零向量是唯一的 (见习题 1.4)。

c) 正确。根据定义。

d) 错误。例如  $f(t) = t^2 + t$  和  $g(t) = -t^2 + 1$  都是 2 次多项式, 但  $(f + g)(t) = t + 1$  是 1 次多项式。

e) 正确。两个多项式相加, 其次数不可能超过原来两个多项式中次数较高的那个。

1.4. 证明: 假设  $\mathbf{0}$  和  $\mathbf{0}'$  都是零向量。

根据  $\mathbf{0}'$  作为零向量的性质, 有  $\mathbf{0} + \mathbf{0}' = \mathbf{0}$ 。

根据  $\mathbf{0}$  作为零向量的性质, 有  $\mathbf{0} + \mathbf{0}' = \mathbf{0}'$  (利用交换律)。

因此  $\mathbf{0} = \mathbf{0}'$ 。

1.5. 解: 这是一个所有元素均为 0 的  $2 \times 3$  矩阵:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

1.6. 证明: 设向量  $\mathbf{v} \in V$  有两个加法逆元  $\mathbf{w}$  和  $\mathbf{w}'$ , 即  $\mathbf{v} + \mathbf{w} = \mathbf{0}$  且  $\mathbf{v} + \mathbf{w}' = \mathbf{0}$ 。

考虑  $\mathbf{w} + \mathbf{v} + \mathbf{w}'$ :

$$\mathbf{w} = \mathbf{w} + \mathbf{0} = \mathbf{w} + (\mathbf{v} + \mathbf{w}') = (\mathbf{w} + \mathbf{v}) + \mathbf{w}' = \mathbf{0} + \mathbf{w}' = \mathbf{w}'$$

因此  $\mathbf{w} = \mathbf{w}'$ , 加法逆元是唯一的。

1.7. 证明:

$$0\mathbf{v} = (0 + 0)\mathbf{v} = 0\mathbf{v} + 0\mathbf{v}$$

在等式两边加上  $0\mathbf{v}$  的加法逆元  $-(0\mathbf{v})$ :

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

即  $\mathbf{0} = 0\mathbf{v}$ 。

1.8. 证明: 我们需要证明  $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$ 。

$$\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1 + (-1))\mathbf{v} = 0\mathbf{v}$$

利用习题 1.7 的结论  $0\mathbf{v} = \mathbf{0}$ , 我们得到  $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$ 。

由习题 1.6 可知加法逆元唯一, 因此  $(-1)\mathbf{v}$  必定是  $\mathbf{v}$  的加法逆元  $-\mathbf{v}$ 。

2.1. 解: 可以取以下 6 个矩阵作为一组基 (标准基):

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

2.2. 解: **a) 正确。** 如果集合包含  $\mathbf{0}$ , 则  $1 \cdot \mathbf{0} = \mathbf{0}$  是一个非平凡的线性组合 (系数 1 非零), 故线性相关。

**b) 错误。** 由 (a) 可知, 若基包含  $\mathbf{0}$  则线性相关, 但这与基必须线性无关矛盾。

**c) 错误。** 例如  $\{\mathbf{v}, 2\mathbf{v}\}$  是线性相关的, 但其子集  $\{\mathbf{v}\}$  (若  $\mathbf{v} \neq \mathbf{0}$ ) 是线性无关的。

**d) 正确。** 如果子集有非平凡线性组合为  $\mathbf{0}$ , 则原集合也有 (其余系数取 0), 这与原集合线性无关矛盾。

**e) 错误。** 这仅在向量系是线性无关时才成立。对于线性相关的系统, 存在不全为零的  $\alpha_k$  满足该等式。

2.3. 解: 任意  $2 \times 2$  对称矩阵的形式为  $\begin{pmatrix} a & b \\ b & c \end{pmatrix} = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ 。

基可以选为:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

基中有 3 个元素。

2.4. 解: 令  $E_{j,k}$  为第  $j$  行第  $k$  列元素为 1, 其余元素为 0 的  $n \times n$  矩阵。

**a)** 基由 6 个矩阵组成: 对角线元素:  $E_{1,1}, E_{2,2}, E_{3,3}$ ; 非对角线元素 (对称对):  $E_{1,2} + E_{2,1}, E_{1,3} + E_{3,1}, E_{2,3} + E_{3,2}$ 。

**b)** 基由  $E_{k,k}$  ( $1 \leq k \leq n$ ) 和  $E_{j,k} + E_{k,j}$  ( $1 \leq j < k \leq n$ ) 组成。元素个数为  $n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$ 。

**c)** 反对称矩阵对角线必须为 0。基由  $E_{j,k} - E_{k,j}$  ( $1 \leq j < k \leq n$ ) 组成。元素个数为  $\frac{n(n-1)}{2}$ 。



- 2.5. 证明：由于系统不是生成的（完备的），存在向量  $\mathbf{v}_{r+1}$  不能写成  $\mathbf{v}_1, \dots, \mathbf{v}_r$  的线性组合。考虑方程

$$\alpha_1 \mathbf{v}_1 + \dots + \alpha_r \mathbf{v}_r + \alpha_{r+1} \mathbf{v}_{r+1} = \mathbf{0}$$

我们需要证明所有系数必为 0。

如果  $\alpha_{r+1} \neq 0$ ，则可以写出  $\mathbf{v}_{r+1} = -\frac{1}{\alpha_{r+1}} \sum_{k=1}^r \alpha_k \mathbf{v}_k$ 。这意味着  $\mathbf{v}_{r+1}$  是前  $r$  个向量的线性组合，与我们的选择矛盾。

因此必须有  $\alpha_{r+1} = 0$ 。方程变为  $\sum_{k=1}^r \alpha_k \mathbf{v}_k = \mathbf{0}$ 。

又因为  $\mathbf{v}_1, \dots, \mathbf{v}_r$  是线性无关的，所以  $\alpha_1 = \dots = \alpha_r = 0$ 。

综上，所有系数均为 0，新系统是线性无关的。

- 2.6. 解：不可能。我们可以将  $\mathbf{v}_k$  表示为  $\mathbf{w}_k$  的线性组合。解方程组可得：

$$\mathbf{v}_1 = \frac{1}{2}(\mathbf{w}_1 - \mathbf{w}_2 + \mathbf{w}_3), \quad \mathbf{v}_2 = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2 - \mathbf{w}_3), \quad \mathbf{v}_3 = \frac{1}{2}(-\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3)$$

假设  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  是线性无关的。

考虑  $\mathbf{v}_k$  的线性组合： $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 = \mathbf{0}$ 。

将  $\mathbf{v}_k$  的表达式代入，整理  $\mathbf{w}_k$  的系数。由于  $\mathbf{w}_k$  线性无关，其系数必须为 0。

例如  $\mathbf{w}_1$  的系数为  $\frac{1}{2}(c_1 + c_2 - c_3) = 0$ 。同理得到  $\frac{1}{2}(-c_1 + c_2 + c_3) = 0$  和  $\frac{1}{2}(c_1 - c_2 + c_3) = 0$ 。

这三个方程构成的方程组只有零解  $c_1 = c_2 = c_3 = 0$ 。

这意味着：如果  $\mathbf{w}_k$  线性无关，则  $\mathbf{v}_k$  必须线性无关。

根据逆否命题：如果  $\mathbf{v}_k$  线性相关，那么  $\mathbf{w}_k$  也必须线性相关。

- 3.1. 解：a)

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 3 + 3 \cdot 2 \\ 4 \cdot 1 + 5 \cdot 3 + 6 \cdot 2 \end{pmatrix} = \begin{pmatrix} 13 \\ 31 \end{pmatrix}$$

b)

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 3 \\ 0 \cdot 1 + 1 \cdot 3 \\ 2 \cdot 1 + 0 \cdot 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \\ 2 \end{pmatrix}$$

c)

$$\begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1+4 \\ 2+6 \\ 3+8 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ 11 \\ 4 \end{pmatrix}$$

d) 未定义。左侧矩阵大小为  $4 \times 3$ ，右侧向量大小为  $4 \times 1$ 。矩阵的列数 (3) 与向量的行数 (4) 不匹配，因此无法进行乘法运算。

- 3.2. 解：直线方向向量为  $\mathbf{v} = (3, 1)^T$ ，法向量为  $\mathbf{b} = (1, -3)^T$ 。反射矩阵由  $R = I - 2P_{\mathbf{b}}$  给出，其中  $P_{\mathbf{b}}$  是在该法向量上的投影。

$$P_{\mathbf{b}} \mathbf{x} = \frac{\mathbf{b} \mathbf{b}^T}{\|\mathbf{b}\|^2} \mathbf{x} = \frac{1}{1^2 + (-3)^2} \begin{pmatrix} 1 \\ -3 \end{pmatrix} (1, -3) \mathbf{x} = \frac{1}{10} \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix} \mathbf{x}$$

因此

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2}{10} \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.2 & -0.6 \\ -0.6 & 1.8 \end{pmatrix} = \begin{pmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{pmatrix}$$

也可以写成分数形式：

$$\frac{1}{5} \begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix}$$

- 3.3. 解：a) 将  $T$  作用于基向量  $\mathbf{e}_1, \mathbf{e}_2$ ：

$$T(\mathbf{e}_1) = (1, 2, 0)^T, \quad T(\mathbf{e}_2) = (2, -5, 7)^T$$

矩阵为:

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

b) 将  $T$  作用于标准基, 分别提取  $x_1, x_2, x_3, x_4$  的系数作为列:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 3 & 0 & 6 \end{pmatrix}$$

c) 这是一个  $(n+1) \times (n+1)$  矩阵.  $T(1) = 0, T(t) = 1, T(t^2) = 2t, \dots, T(t^k) = kt^{k-1}$ . 第  $k$  列对应  $t^{k-1}$  的导数 (注意: 列索引从 1 开始, 对应基向量  $1, t, \dots, t^n$ ). 矩阵在主对角线上方的一条对角线上元素为  $1, 2, \dots, n$ , 其余为 0:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

d) 设  $D$  为 (c) 中的求导矩阵, 所求矩阵为  $M = 2I + 3D - 4D^2$ . 这是一个上三角矩阵: 对角线元素均为 2. 第一条超对角线元素为  $3 \times (1, 2, \dots, n)$ . 第二条超对角线元素为  $-4 \times (1 \cdot 2, 2 \cdot 3, \dots, (n-1)n)$ . 前几项看起来像:

$$\begin{pmatrix} 2 & 3 & -8 & 0 & \cdots \\ 0 & 2 & 6 & -24 & \cdots \\ 0 & 0 & 2 & 9 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

3.4. 解: a) 映射为  $(x, y, z)^T \mapsto (x, y, 0)^T$ .

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

b) 映射为  $(x, y, z)^T \mapsto (x, y, -z)^T$ .

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

c)  $z$  轴不变,  $x, y$  进行旋转.  $\cos 30^\circ = \sqrt{3}/2, \sin 30^\circ = 1/2$ .

$$\begin{pmatrix} \sqrt{3}/2 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3.5. 证明:  $\mathbf{z}$  是线段  $[\mathbf{x}, \mathbf{y}]$  的中点意味着  $\mathbf{z} = \frac{1}{2}(\mathbf{x} + \mathbf{y})$ . 利用  $A$  的线性性质:

$$A\mathbf{z} = A\left(\frac{1}{2}(\mathbf{x} + \mathbf{y})\right) = \frac{1}{2}(A\mathbf{x} + A\mathbf{y})$$

这正是线段  $[A\mathbf{x}, A\mathbf{y}]$  中点的定义.

3.6. 解: a)  $\mathbb{C}$  作为复向量空间是一维的, 基向量为  $\{1\}$ . 对于任何  $z \in \mathbb{C}$ , 变换  $T(z) = \alpha z$  满足线性性. 由于空间是 1 维的, 其矩阵是  $1 \times 1$  矩阵:  $(\alpha)$ .

b) 对应到  $\mathbb{R}^2$ , 基为  $1 \sim (1, 0)^T$  和  $i \sim (0, 1)^T$ .

$$\alpha \cdot 1 = a + ib \sim (a, b)^T$$

$$\alpha \cdot i = (a + ib)i = -b + ia \sim (-b, a)^T$$

因此矩阵为:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

c) 检查复线性:  $T(i) = T(0 + i \cdot 1) = -1 - 3i$ . 另一方面, 若  $T$  是复线性的, 应有  $T(i) = iT(1) = i(2 + i) = 2i - 1 = -1 + 2i$ . 两者不相等, 故不是复线性的。

视作实线性变换:  $T(x, y)^T = (2x - y, x - 3y)^T$ . 这是线性的, 其矩阵为:

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

3.7. 证明: 复向量空间  $\mathbb{C}$  的维度为 1, 取  $\{1\}$  为基. 设  $T: \mathbb{C} \rightarrow \mathbb{C}$  是线性变换. 令  $\alpha = T(1)$ . 对于任意  $z \in \mathbb{C}$ , 由于  $z$  是复数标量, 我们可以写  $z = z \cdot 1$ . 利用  $T$  的复线性 (标量可以提出来):

$$T(z) = T(z \cdot 1) = zT(1) = z\alpha = \alpha z$$

得证。

5.1. 解: a) 各矩阵维数:  $A: 2 \times 2, B: 2 \times 3, C: 2 \times 3, D: 3 \times 1$ .

$AB$ : 有定义,  $(2 \times 2) \cdot (2 \times 3) \rightarrow 2 \times 3$ .

$BA$ : 无定义 ( $3 \neq 2$ ).

$ABC$ : 无定义 ( $AB$  为  $2 \times 3$ , 不能乘以  $C$ ).

$ABD$ : 有定义,  $(2 \times 3) \cdot (3 \times 1) \rightarrow 2 \times 1$ .

$BC$ : 无定义.

$BC^T$ : 有定义,  $(2 \times 3) \cdot (3 \times 2) \rightarrow 2 \times 2$ .

$B^T C$ : 有定义,  $(3 \times 2) \cdot (2 \times 3) \rightarrow 3 \times 3$ .

$DC$ : 无定义 ( $1 \neq 2$ ).

$D^T C^T$ : 有定义,  $(1 \times 3) \cdot (3 \times 2) \rightarrow 1 \times 2$ .

b)

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 1+6 & 0+2 & 2-4 \\ 3+3 & 0+1 & 6-2 \end{pmatrix} = \begin{pmatrix} 7 & 2 & -2 \\ 6 & 1 & 4 \end{pmatrix}$$

$$\begin{aligned} A(3B + C) &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \left( \begin{pmatrix} 3 & 0 & 6 \\ 9 & 3 & -6 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 3 \\ -2 & 1 & -1 \end{pmatrix} \right) \\ &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 4 & -2 & 9 \\ 7 & 4 & -7 \end{pmatrix} \\ &= \begin{pmatrix} 4+14 & -2+8 & 9-14 \\ 12+7 & -6+4 & 27-7 \end{pmatrix} = \begin{pmatrix} 18 & 6 & -5 \\ 19 & -2 & 20 \end{pmatrix} \end{aligned}$$

$$B^T A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1+9 & 2+3 \\ 0+3 & 0+1 \\ 2-6 & 4-2 \end{pmatrix} = \begin{pmatrix} 10 & 5 \\ 3 & 1 \\ -4 & 2 \end{pmatrix}$$

$$A(BD) = A \left( \begin{pmatrix} 1 & 0 & 2 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -6 \end{pmatrix} = \begin{pmatrix} -12 \\ -6 \end{pmatrix}$$

$$(AB)D = A(BD) = \begin{pmatrix} -12 \\ -6 \end{pmatrix} \quad (\text{利用结合律})$$

5.2. 解: 旋转矩阵为  $T_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$ . 注意到  $\cos(-\gamma) = \cos \gamma$  且  $\sin(-\gamma) = -\sin \gamma$ , 故

$$T_{-\gamma} = \begin{pmatrix} \cos \gamma & \sin \gamma \\ -\sin \gamma & \cos \gamma \end{pmatrix}.$$

$$\begin{aligned} T_{\gamma} T_{-\gamma} &= \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} \cos \gamma & \sin \gamma \\ -\sin \gamma & \cos \gamma \end{pmatrix} \\ &= \begin{pmatrix} \cos^2 \gamma + \sin^2 \gamma & \cos \gamma \sin \gamma - \sin \gamma \cos \gamma \\ \sin \gamma \cos \gamma - \cos \gamma \sin \gamma & \sin^2 \gamma + \cos^2 \gamma \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \end{aligned}$$

同理可证  $T_{-\gamma} T_{\gamma} = I$ .

5.3. 解: 我们计算  $T_{\alpha} T_{\beta}$ :

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

另一方面, 旋转  $\beta$  角再旋转  $\alpha$  角等同于旋转  $\alpha + \beta$  角, 即  $T_{\alpha+\beta}$ :

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

比较对应元素可得:

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

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

5.4. 解: 该直线的方向向量可取为  $\mathbf{v} = (-2, 1)^T$ . 使用正交投影矩阵公式  $P = \frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2}$ :

$$\mathbf{v}\mathbf{v}^T = \begin{pmatrix} -2 \\ 1 \end{pmatrix} (-2, 1) = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}$$

$$\|\mathbf{v}\|^2 = (-2)^2 + 1^2 = 5$$

$$P = \frac{1}{5} \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 0.2 \end{pmatrix}$$

(注: 若利用提示, 可将直线旋转至  $x_1$  轴, 投影后再转回.)

5.5. 解: 取  $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ ,  $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ .

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq 0$$

5.6. 证明: 设  $A$  为  $m \times n$  矩阵,  $B$  为  $n \times m$  矩阵.  $AB$  是  $m \times m$  矩阵, 其第  $j$  个对角元为  $(AB)_{jj} = \sum_{k=1}^n A_{jk} B_{kj}$ .

$$\text{trace}(AB) = \sum_{j=1}^m (AB)_{jj} = \sum_{j=1}^m \sum_{k=1}^n A_{jk} B_{kj}$$

$BA$  是  $n \times n$  矩阵, 其第  $k$  个对角元为  $(BA)_{kk} = \sum_{j=1}^m B_{kj} A_{jk}$ .

$$\text{trace}(BA) = \sum_{k=1}^n (BA)_{kk} = \sum_{k=1}^n \sum_{j=1}^m B_{kj} A_{jk}$$

由于标量乘法满足交换律, 且有限求和次序可交换, 上述两个和式相等。

5.7. 解:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

计算验证:

$$A^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

5.8. 解: 我们将直线  $y = -2x/3$  (即斜率为  $-2/3$ ) 旋转到  $x$  轴, 进行反射, 然后转回. 设直线与  $x$  轴夹角为  $\theta$ , 则  $\tan \theta = -2/3$ . 这意味着  $\sin \theta = -2/\sqrt{13}$ ,  $\cos \theta = 3/\sqrt{13}$ . 旋转矩阵  $R_{-\theta}$  (将直线转到  $x$  轴) 为:

$$R_{-\theta} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \frac{1}{\sqrt{13}} \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix}$$

关于  $x$  轴的反射矩阵为  $S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . 逆旋转矩阵  $R_{\theta} = \frac{1}{\sqrt{13}} \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}$ . 所求反射矩阵  $M = R_{\theta} S R_{-\theta}$ :

$$\begin{aligned} M &= \frac{1}{13} \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix} \\ &= \frac{1}{13} \begin{pmatrix} 3 & -2 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 3 \end{pmatrix} \\ &= \frac{1}{13} \begin{pmatrix} 9-4 & -6-6 \\ -6-6 & 4-9 \end{pmatrix} = \frac{1}{13} \begin{pmatrix} 5 & -12 \\ -12 & -5 \end{pmatrix} \end{aligned}$$

6.1. 证明: 由于  $A$  是同构, 它既是单射也是满射.

1. **线性无关性:** 考虑方程  $\sum_{k=1}^n c_k (A\mathbf{v}_k) = \mathbf{0}$ . 利用  $A$  的线性性质, 得  $A(\sum_{k=1}^n c_k \mathbf{v}_k) = \mathbf{0}$ . 因为  $A$  是可逆的 (核为  $\{\mathbf{0}\}$ ), 所以  $\sum_{k=1}^n c_k \mathbf{v}_k = \mathbf{0}$ . 又因为  $\mathbf{v}_k$  是基 (线性无关), 所以所有  $c_k = 0$ . 因此  $A\mathbf{v}_1, \dots, A\mathbf{v}_n$  线性无关.

2. **生成性:** 对任意  $\mathbf{w} \in W$ , 由于  $A$  是可逆的, 存在  $\mathbf{v} \in V$  使得  $A\mathbf{v} = \mathbf{w}$ . 因为  $\mathbf{v}_k$  生成  $V$ , 我们可以写  $\mathbf{v} = \sum_{k=1}^n \alpha_k \mathbf{v}_k$ . 则  $\mathbf{w} = A(\sum_{k=1}^n \alpha_k \mathbf{v}_k) = \sum_{k=1}^n \alpha_k (A\mathbf{v}_k)$ . 因此  $A\mathbf{v}_1, \dots, A\mathbf{v}_n$  生成  $W$ .

6.2. 解:  $A$  是  $1 \times 2$  矩阵. 右逆  $B$  必须是  $2 \times 1$  矩阵, 即  $B = (x, y)^T$ , 满足  $AB = I_1 = (1)$ .

$$(1, 1) \begin{pmatrix} x \\ y \end{pmatrix} = x + y = 1$$

解为  $y = 1 - x$ . 所有右逆的形式为  $B = (x, 1-x)^T$ , 其中  $x \in \mathbb{R}$ .

如果  $A$  也是左可逆的, 那么它是可逆的 (既有左逆又有右逆). 这意味着  $A$  必须是方阵 ( $n \times n$ ). 但  $A$  是  $1 \times 2$ , 不是方阵, 所以  $A$  不可逆. 既然它有右逆, 那么它一定不能有左逆 (否则它就是可逆的了). 或者直接验证: 设  $C = (c_1, c_2)^T$  是左逆, 则  $CA = \begin{pmatrix} c_1 & c_1 \\ c_2 & c_2 \end{pmatrix}$ ,

这永远不可能等于  $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ .

6.3. 解: 设  $A = (1, 2, 3)^T$ . 左逆  $B$  是  $1 \times 3$  矩阵, 即  $B = (x, y, z)$ , 满足  $BA = I_1 = (1)$ .

$$(x, y, z) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = x + 2y + 3z = 1$$

这是一个有无穷多解的平面方程. 所有左逆由向量  $(x, y, z)$  组成, 满足  $x = 1 - 2y - 3z$ , 其中  $y, z \in \mathbb{R}$  为任意实数.

6.4. 解: 不是。设  $A = (1, 2, 3)^T$ . 若存在右逆  $C$  (必须是  $1 \times 3$ ), 则  $AC = I_3$ .

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} (c_1, c_2, c_3) = \begin{pmatrix} c_1 & c_2 & c_3 \\ 2c_1 & 2c_2 & 2c_3 \\ 3c_1 & 3c_2 & 3c_3 \end{pmatrix}$$

要使该矩阵等于  $I_3$ , 对角线元素必须为 1, 非对角线元素必须为 0. 比较元素 (1, 2):  $c_2$  必须为 0. 比较元素 (2, 2):  $2c_2$  必须为 1. 矛盾 ( $2 \cdot 0 \neq 1$ ). 因此不存在右逆。

6.5. 解: 令  $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} (2 \times 3)$ ,  $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} (3 \times 2)$ .  $A, B$  均非方阵, 故定义上不可逆。

$$AB = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

$AB$  是  $2 \times 2$  的单位矩阵, 是可逆的。

6.6. 证明: 设  $C = (AB)^{-1}$ , 即  $(AB)C = I$  且  $C(AB) = I$ .

对于  $A$ : 我们有  $A(BC) = (AB)C = I$ . 因此  $BC$  是  $A$  的一个右逆。所以  $A$  是右可逆的。

对于  $B$ : 我们有  $(CA)B = C(AB) = I$ . 因此  $CA$  是  $B$  的一个左逆。所以  $B$  是左可逆的。

6.7. 证明: 由于  $A$  是可逆的, 存在逆矩阵  $A^{-1}$ . 我们可以将  $B$  写为:

$$B = IB = (A^{-1}A)B = A^{-1}(AB)$$

因为  $A$  可逆, 所以  $A^{-1}$  可逆。已知  $AB$  可逆。因为两个可逆矩阵的乘积是可逆的 (定理 6.3), 所以  $B = A^{-1}(AB)$  是可逆的。

6.8. 证明: 假设  $A$  是可逆的, 则存在  $A^{-1}$ . 从  $A^2 = \mathbf{0}$  出发, 我们在等式两边同时左乘  $A^{-1}$ :

$$A^{-1}(AA) = A^{-1}\mathbf{0}$$

$$(A^{-1}A)A = \mathbf{0}$$

$$IA = \mathbf{0}$$

$$A = \mathbf{0}$$

但是零矩阵 (在  $n \geq 1$  时) 不可逆 (因为它没有逆矩阵能满足  $\mathbf{0}B = I$ )。这与假设  $A$  可逆矛盾。因此  $A$  不可逆。

6.9. 解: 不能。假如  $A$  是可逆的, 左乘  $A^{-1}$ :

$$A^{-1}(AB) = A^{-1}\mathbf{0} \implies (A^{-1}A)B = \mathbf{0} \implies IB = \mathbf{0} \implies B = \mathbf{0}$$

这与题目条件  $B$  是非零矩阵矛盾。因此  $A$  不可能是可逆的。

6.10. 解: 矩阵表示:  $T_1$  将标准基向量  $\mathbf{e}_2$  与  $\mathbf{e}_4$  互换, 保持  $\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_5$  不变。

$$T_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$T_2$  中,  $T_2(\mathbf{e}_4)$  的第 2 分量变成了  $a$ , 其余不变。即  $T_2(\mathbf{e}_4) = \mathbf{e}_4 + a\mathbf{e}_2$ .

$$T_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & a & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

**逆变换与逆矩阵:** 对于  $T_1$ : 其逆操作是“再次交换  $x_2$  和  $x_4$ ”以恢复原状。这意味着  $T_1^{-1} = T_1$ . 验证:  $T_1 T_1 = I$ .

$$T_1^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

对于  $T_2$ : 其逆操作是将  $x_2$  减去  $ax_4$ , 即加  $(-a)x_4$ . 这恢复了原来的  $x_2$ . 因此  $T_2^{-1}$  的矩阵与  $T_2$  形式相同, 只是  $a$  变成了  $-a$ .

$$T_2^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -a & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

6.11. 解: 我们将变换分解为: 旋转坐标系使向量  $(1, 2, 3)^T$  与  $z$  轴重合, 进行  $z$  轴旋转, 然后逆向旋转回原坐标系。

1. 将  $(1, 2, 3)^T$  的投影  $(1, 2, 0)^T$  旋转到  $x$  轴上.  $\mathbf{v}_{xy} = (1, 2)^T$ , 长度  $\sqrt{5}$ .  $\cos \phi = 1/\sqrt{5}$ ,  $\sin \phi = 2/\sqrt{5}$ . 我们需要顺时针旋转  $\phi$ , 即旋转  $-\phi$ . 矩阵  $M_1$  (绕  $z$  轴旋转  $-\phi$ ):

$$M_1 = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

变换后向量变为  $(\sqrt{5}, 0, 3)^T$ .

2. 将  $(\sqrt{5}, 0, 3)^T$  旋转到  $z$  轴上 (绕  $y$  轴旋转). 当前向量在  $xz$  平面. 长度  $\sqrt{5+9} = \sqrt{14}$ . 与  $z$  轴夹角  $\theta$  满足  $\cos \theta = 3/\sqrt{14}$ ,  $\sin \theta = \sqrt{5}/\sqrt{14}$ . 我们需要逆时针旋转  $\theta$  (将  $x$  轴偏向的向量转到  $z$  轴, 注意方向, 或者说是绕  $y$  轴负方向). 或者直接写出绕  $y$  轴旋转  $-\theta$  的矩阵 (将向量从  $x$  转到  $z$ ):

$$M_2 = \begin{pmatrix} 3/\sqrt{14} & 0 & -\sqrt{5}/\sqrt{14} \\ 0 & 1 & 0 \\ \sqrt{5}/\sqrt{14} & 0 & 3/\sqrt{14} \end{pmatrix}$$

此时向量变为  $(0, 0, \sqrt{14})^T$ , 即  $z$  轴方向。

3. 绕  $z$  轴旋转  $\alpha$ .

$$R_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4. 逆操作恢复原坐标系. 最终矩阵为:

$$M = M_1^{-1} M_2^{-1} R_\alpha M_2 M_1$$

其中  $M_1^{-1} = M_1^T$ ,  $M_2^{-1} = M_2^T$  (因为旋转矩阵是正交的)。

6.12. 解:

a) 取  $A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $B = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ .  $A, B$  可逆, 但  $A + B = \mathbf{0}$  不可逆。

b) 取  $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ ,  $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ .  $A, B$  均不可逆 (有零行), 但  $A + B = I$  可逆。

c) 取  $A = I, B = I$ .  $A, B$  可逆,  $A + B = 2I = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$  也是可逆的。

6.13. 解: 是的,  $A^{-1}$  是对称的. 利用定理 6.5 ( $(A^T)^{-1} = (A^{-1})^T$ ) 和  $A$  的对称性 ( $A^T = A$ ):

$$(A^{-1})^T = (A^T)^{-1} = A^{-1}$$

由于  $(A^{-1})^T = A^{-1}$ , 所以  $A^{-1}$  是对称矩阵。

- 7.1. 证明: 我们需要验证子空间的三个条件: 1. **零向量**: 因为  $X$  和  $Y$  是子空间, 所以  $\mathbf{0} \in X$  且  $\mathbf{0} \in Y$ , 因此  $\mathbf{0} \in X \cap Y$ .  $X \cap Y$  非空.  
 2. **加法封闭**: 设  $\mathbf{u}, \mathbf{v} \in X \cap Y$ . 因为  $\mathbf{u}, \mathbf{v} \in X$  且  $X$  是子空间, 所以  $\mathbf{u} + \mathbf{v} \in X$ . 同理, 因为  $\mathbf{u}, \mathbf{v} \in Y$  且  $Y$  是子空间, 所以  $\mathbf{u} + \mathbf{v} \in Y$ . 因此  $\mathbf{u} + \mathbf{v} \in X \cap Y$ .  
 3. **标量乘法封闭**: 设  $\mathbf{v} \in X \cap Y$ ,  $\alpha$  为标量. 因为  $\mathbf{v} \in X$ , 所以  $\alpha\mathbf{v} \in X$ . 因为  $\mathbf{v} \in Y$ , 所以  $\alpha\mathbf{v} \in Y$ . 因此  $\alpha\mathbf{v} \in X \cap Y$ .  
 综上,  $X \cap Y$  是  $V$  的子空间。

- 7.2. 证明: 1. **零向量**:  $\mathbf{0} = \mathbf{0} + \mathbf{0}$ . 因为  $\mathbf{0} \in X, \mathbf{0} \in Y$ , 所以  $\mathbf{0} \in X + Y$ .  
 2. **加法封闭**: 设  $\mathbf{w}_1, \mathbf{w}_2 \in X + Y$ . 存在  $\mathbf{x}_1, \mathbf{x}_2 \in X$  和  $\mathbf{y}_1, \mathbf{y}_2 \in Y$  使得  $\mathbf{w}_1 = \mathbf{x}_1 + \mathbf{y}_1$ ,  $\mathbf{w}_2 = \mathbf{x}_2 + \mathbf{y}_2$ .

$$\mathbf{w}_1 + \mathbf{w}_2 = (\mathbf{x}_1 + \mathbf{y}_1) + (\mathbf{x}_2 + \mathbf{y}_2) = (\mathbf{x}_1 + \mathbf{x}_2) + (\mathbf{y}_1 + \mathbf{y}_2)$$

因为  $X, Y$  是子空间, 所以  $\mathbf{x}_1 + \mathbf{x}_2 \in X$ ,  $\mathbf{y}_1 + \mathbf{y}_2 \in Y$ . 因此  $\mathbf{w}_1 + \mathbf{w}_2 \in X + Y$ .

3. **标量乘法封闭**: 设  $\mathbf{w} \in X + Y$  (即  $\mathbf{w} = \mathbf{x} + \mathbf{y}$ ),  $\alpha$  为标量。

$$\alpha\mathbf{w} = \alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$$

因为  $X, Y$  是子空间, 所以  $\alpha\mathbf{x} \in X$ ,  $\alpha\mathbf{y} \in Y$ . 因此  $\alpha\mathbf{w} \in X + Y$ .

- 7.3. 证明 (反证法): 假设  $\mathbf{x} + \mathbf{v} \in X$ . 因为  $X$  是子空间且  $\mathbf{x} \in X$ , 所以其加法逆元  $-\mathbf{x} \in X$ . 由子空间的加法封闭性, 它们的和也应该在  $X$  中:

$$(\mathbf{x} + \mathbf{v}) + (-\mathbf{x}) \in X$$

左边化简为  $\mathbf{v}$ , 即这意味着  $\mathbf{v} \in X$ . 这与已知条件  $\mathbf{v} \notin X$  矛盾. 因此假设不成立, 必有  $\mathbf{x} + \mathbf{v} \notin X$ .

- 7.4. 证明: ( $\Leftarrow$ ) 如果  $X \subset Y$ , 则  $X \cup Y = Y$ , 这是已知的子空间. 反之亦然.  
 ( $\Rightarrow$ ) 我们证明如果  $X \not\subset Y$  且  $Y \not\subset X$ , 则  $X \cup Y$  不是子空间. 假设  $X \not\subset Y$  且  $Y \not\subset X$ . 这意味着存在向量  $\mathbf{x} \in X$  但  $\mathbf{x} \notin Y$ ; 同时存在向量  $\mathbf{y} \in Y$  但  $\mathbf{y} \notin X$ . 显然  $\mathbf{x}, \mathbf{y} \in X \cup Y$ . 考虑它们的和  $\mathbf{z} = \mathbf{x} + \mathbf{y}$ . 如果  $X \cup Y$  是子空间, 那么必须有  $\mathbf{z} \in X \cup Y$ , 即  $\mathbf{z} \in X$  或  $\mathbf{z} \in Y$ .  
 假设  $\mathbf{z} \in X$ : 即  $\mathbf{x} + \mathbf{y} \in X$ . 已知  $\mathbf{x} \in X$ , 且  $\mathbf{y} \notin X$ . 根据习题 7.3 的结论,  $\mathbf{x} + \mathbf{y}$  不可能属于  $X$ . 矛盾.  
 假设  $\mathbf{z} \in Y$ : 即  $\mathbf{y} + \mathbf{x} \in Y$ . 已知  $\mathbf{y} \in Y$ , 且  $\mathbf{x} \notin Y$ . 根据习题 7.3 的结论 (交换角色),  $\mathbf{y} + \mathbf{x}$  不可能属于  $Y$ . 矛盾.  
 因此  $\mathbf{x} + \mathbf{y} \notin X \cup Y$ . 这违反了加法封闭性. 所以  $X \cup Y$  不是子空间。

- 7.5. 解: 1. **包含两者的最小子空间**: 即它们的和空间  $U + S$ . 对于任意  $4 \times 4$  矩阵  $M$ , 我们总可以将其写为一个上三角矩阵和一个对称矩阵之和吗? 设  $M$  为任意矩阵. 构造对称矩阵  $S$  如下: 当  $j < k$  时, 令  $S_{jk} = 0$  (或任意值); 当  $j > k$  时, 必须有  $S_{jk} = M_{jk}$  (为了抵消  $M$  的下三角部分, 稍后解释); 当  $j = k$  时, 令  $S_{jj} = 0$ . 更简单的构造: 我们需要  $M = U + S$ . 即  $M - S = U$ . 这意味着  $M - S$  的下三角部分 ( $j > k$ ) 必须为 0. 即  $M_{jk} - S_{jk} = 0 \Rightarrow S_{jk} = M_{jk}$  对所有  $j > k$  成立. 由于  $S$  是对称的, 这确定了  $S$  的上三角部分:  $S_{kj} = S_{jk} = M_{jk}$  对所有  $j > k$  (即  $k < j$ ) 成立. 我们可以自由选择  $S$  的对角线元素 (例如全为 0). 这样构造出的  $S$  是对称的. 然后令  $U = M - S$ , 则  $U$  必然是上三角的. 结论: 任意矩阵都可以分解, 所以包含两者的最小子空间是**所有  $4 \times 4$  矩阵的空间  $M_{4 \times 4}$** .

2. **包含在两者中的最大子空间**: 即它们的交空间  $U \cap S$ . 矩阵  $A$  必须既是上三角的又是对称的. 上三角意味着: 当  $j > k$  时,  $A_{jk} = 0$  (下三角部分为 0). 对称意味着:  $A_{kj} = A_{jk}$ . 结合两者: 对于  $k < j$  (上三角部分),  $A_{kj} = A_{jk} = 0$ . 所以,  $A$  的下三角部分是 0, 上三角部分也是 0. 只有对角线元素可以非零. 结论: 这是**所有对角矩阵的空间** (记作  $D_4$  或对角矩阵集合)。

- 8.1. 解: 从齐次坐标  $(x, y, z, w)^T$  转换回  $\mathbb{R}^3$  中的笛卡尔坐标, 需要将前三个分量除以第四个分



量  $w$  (假设  $w \neq 0$ )。这里  $w = 5$ 。

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 10/5 \\ 20/5 \\ 30/5 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$$

所求向量为  $(2, 4, 6)^T$ 。

- 8.2. 证明: 我们计算两个矩阵的乘积, 检查哪种顺序能得到旋转矩阵  $R_\gamma = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$ 。  
尝试  $T_2 T_1$ :

$$\begin{aligned} T_2 T_1 &= \begin{pmatrix} \sec \gamma & -\tan \gamma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin \gamma & \cos \gamma \end{pmatrix} \\ &= \begin{pmatrix} \sec \gamma - \tan \gamma \sin \gamma & -\tan \gamma \cos \gamma \\ 0 \cdot 1 + 1 \cdot \sin \gamma & 0 \cdot 0 + 1 \cdot \cos \gamma \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\cos \gamma} - \frac{\sin^2 \gamma}{\cos \gamma} & -\frac{\sin \gamma}{\cos \gamma} \cos \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \\ &= \begin{pmatrix} \frac{1 - \sin^2 \gamma}{\cos \gamma} & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \\ &= \begin{pmatrix} \frac{\cos^2 \gamma}{\cos \gamma} & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \end{aligned}$$

这正是旋转矩阵  $R_\gamma$ 。

(注: 若计算  $T_1 T_2$ , 结果为  $\begin{pmatrix} \sec \gamma & -\tan \gamma \\ \tan \gamma & \frac{\cos(2\gamma)}{\cos \gamma} \end{pmatrix}$ , 不符合)。

因此, 正确的变换顺序是先应用  $T_1$ , 再应用  $T_2$ , 即矩阵乘积为  $T_2 T_1$ 。

### 8.3. 解: 情况 1: $(AB)D$

1. 计算  $C = AB$ : 这是两个  $2 \times 2$  矩阵相乘。结果有 4 个元素, 每个元素需要 2 次乘法。共  $4 \times 2 = 8$  次乘法。
2. 计算  $CD$ :  $C$  是  $2 \times 2$ ,  $D$  是  $2 \times 1000$ 。对于  $D$  中的每一列 (共 1000 列), 我们需要用  $C$  乘以它。每次向量乘法需 4 次乘法。共  $1000 \times 4 = 4000$  次乘法。
3. 总计:  $4000 + 8 = 4008$  次乘法。

**情况 2:  $A(BD)$**  1. 计算  $E = BD$ :  $B$  是  $2 \times 2$ ,  $D$  是  $2 \times 1000$ 。这需要  $1000 \times 4 = 4000$  次乘法。结果  $E$  是  $2 \times 1000$  矩阵。

2. 计算  $AE$ :  $A$  是  $2 \times 2$ ,  $E$  是  $2 \times 1000$ 。这同样需要  $1000 \times 4 = 4000$  次乘法。

3. 总计:  $4000 + 4000 = 8000$  次乘法。

**结论:** 先计算矩阵乘积  $(AB)D$  效率更高, 大约快一倍。

- 8.4. 解: 正如书中正文 (第 32 页) 提示的, 我们可以通过以下三个步骤构建此变换: 1. **平移**  $T_{in}$ : 将空间移动, 使得投影中心  $(d_1, d_2, d_3)$  移至  $(0, 0, d_3)$ 。这需要平移向量  $(-d_1, -d_2, 0)$ 。

$$T_{in} = \begin{pmatrix} 1 & 0 & 0 & -d_1 \\ 0 & 1 & 0 & -d_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. **投影  $P$** : 应用标准的中心在  $(0, 0, d_3)$  的投影矩阵。根据书中的公式, 投影到  $z = 0$  平面的矩阵为:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d_3 & 1 \end{pmatrix}$$

3. **平移回**  $T_{out}$ : 将坐标系移回, 使得原点恢复。这需要平移向量  $(d_1, d_2, 0)$ 。

$$T_{out} = \begin{pmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

最终矩阵  $M = T_{out}PT_{in}$ . 计算乘积:

$$PT_{in} = \begin{pmatrix} 1 & 0 & 0 & -d_1 \\ 0 & 1 & 0 & -d_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d_3 & 1 \end{pmatrix}$$

$$M = \begin{pmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -d_1 \\ 0 & 1 & 0 & -d_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d_3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -d_1/d_3 & 0 \\ 0 & 1 & -d_2/d_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d_3 & 1 \end{pmatrix}$$

8.5. 解: 为了绕一条任意直线旋转, 我们需要移动坐标系, 使该直线与标准轴 (如  $x$  轴) 重合, 进行标准旋转, 然后逆向操作复原。直线  $L: y = 2x + 3$  在  $z = 0$  平面上。步骤如下:

1. **平移**  $M_1$ : 将直线上的点  $(0, 3, 0)$  移至原点。平移向量为  $(0, -3, 0)$ 。

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

变换后直线方程为  $y = 2x$ 。

2. **旋转对齐**  $M_2$ : 将直线  $y = 2x$  (方向向量  $(1, 2, 0)$ ) 旋转到  $x$  轴。直线斜率为 2, 设倾角为  $\theta$ , 则  $\tan \theta = 2$ ,  $\sin \theta = 2/\sqrt{5}$ ,  $\cos \theta = 1/\sqrt{5}$ 。我们需要绕  $z$  轴顺时针旋转  $\theta$  (即旋转  $-\theta$ )。

$$M_2 = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} & 0 & 0 \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

此时直线与  $x$  轴重合。

3. **执行旋转**  $R_\gamma$ : 绕  $x$  轴旋转  $\gamma$  角。

$$R_\gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma & 0 \\ 0 & \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4. **逆操作**: 逆旋转  $M_2^{-1}$  (即  $M_2^T$ , 因为旋转矩阵是正交的)。逆平移  $M_1^{-1}$  (将  $-3$  改为  $3$ )。最终矩阵  $M$  为:

$$M = M_1^{-1}M_2^{-1}R_\gamma M_2 M_1$$

## 第二章习题解答

2.1. 解: a) 矩阵形式:

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 2 & 1 \\ 3 & 5 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$$

向量方程形式:

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$$

求解: 对增广矩阵进行行约简:

$$\begin{pmatrix} 1 & 2 & -1 & -1 \\ 2 & 2 & 1 & 1 \\ 3 & 5 & -2 & -1 \end{pmatrix} \xrightarrow{R_2-2R_1, R_3-3R_1} \begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & -2 & 3 & 3 \\ 0 & -1 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3, R_2 \times -1} \begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & 1 & -1 & -2 \\ 0 & -2 & 3 & 3 \end{pmatrix} \xrightarrow{R_3+2R_2} \begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

后向代入:  $x_3 = -1$ .  $x_2 - (-1) = -2 \Rightarrow x_2 = -3$ .  $x_1 + 2(-3) - (-1) = -1 \Rightarrow x_1 = 4$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} 4 \\ -3 \\ -1 \end{pmatrix}$$

b) 矩阵形式:

$$\begin{pmatrix} 1 & -2 & -1 \\ 2 & -3 & 1 \\ 3 & -5 & 0 \\ 1 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 7 \\ 9 \end{pmatrix}$$

向量方程形式:

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ -3 \\ -5 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 7 \\ 9 \end{pmatrix}$$

求解:

$$\begin{pmatrix} 1 & -2 & -1 & 1 \\ 2 & -3 & 1 & 6 \\ 3 & -5 & 0 & 7 \\ 1 & 0 & 5 & 9 \end{pmatrix} \xrightarrow{\text{行约简}} \begin{pmatrix} 1 & -2 & -1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$x_3$  为自由变量. 设  $x_3$ .  $x_2 + 3x_3 = 4 \Rightarrow x_2 = 4 - 3x_3$ .  $x_1 - 2(4 - 3x_3) - x_3 = 1 \Rightarrow x_1 = 9 - 5x_3$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} 9 \\ 4 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -5 \\ -3 \\ 1 \end{pmatrix}$$

c) 以下矩阵形式和 向量方程形式略. 求解: 化简增广矩阵:

$$\begin{pmatrix} 1 & 2 & 0 & 2 & 6 \\ 0 & -1 & -1 & 0 & -1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$x_4 = -1$ .  $x_3 - 2(-1) = 0 \Rightarrow x_3 = -2$ .  $-x_2 - (-2) = -1 \Rightarrow x_2 = 3$ .  $x_1 + 2(3) + 2(-1) = 6 \Rightarrow x_1 = 2$ .

解 (向量形式):

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

d) 求解:

$$\left( \begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & -3 & 6 & -3 \end{array} \right) \rightarrow \left( \begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

自由变量:  $x_2, x_4$ .  $x_3 = 1 + 2x_4$ .  $x_1 = 3 + 4x_2 + (1 + 2x_4) - x_4 = 4 + 4x_2 + x_4$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} 4 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

e) 求解: 交换行并化简:

$$\left( \begin{array}{cccc|c} 1 & 2 & -1 & 3 & 2 \\ 0 & 1 & 0 & 2 & 3 \\ 0 & 0 & 1 & 0 & 1 \end{array} \right)$$

自由变量:  $x_4$ .  $x_3 = 1$ .  $x_2 = 3 - 2x_4$ .  $x_1 = 2 - 2(3 - 2x_4) + 1 - 3x_4 = -3 + x_4$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} -3 \\ 3 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix}$$

f) 求解:

$$\left( \begin{array}{ccccc|c} 1 & -1 & 1 & 2 & -1 & 2 \\ 0 & 0 & -3 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & -9 & 9 \end{array} \right)$$

主元列为 1, 3, 4。自由变量:  $x_2, x_5$ .  $x_4 = 9 + 9x_5$ .  $-3x_3 + 2(9 + 9x_5) = -3 \implies -3x_3 = -21 - 18x_5 \implies x_3 = 7 + 6x_5$ .  $x_1 = 2 + x_2 - (7 + 6x_5) - 2(9 + 9x_5) + x_5 = -23 + x_2 - 23x_5$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} -23 \\ 0 \\ 7 \\ 9 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -23 \\ 0 \\ 6 \\ 9 \\ 1 \end{pmatrix}$$

g) 求解: 化简阶梯形:

$$\left( \begin{array}{ccccc|c} 1 & -1 & -1 & -2 & -1 & 2 \\ 0 & 1 & 2 & 2 & 3 & 1 \\ 0 & 0 & 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

主元列: 1, 2, 4。自由变量:  $x_3, x_5$ .  $x_4 = -3 + x_5$ .  $x_2 = 1 - 2x_3 - 2(-3 + x_5) - 3x_5 = 7 - 2x_3 - 5x_5$ .  $x_1 = 2 + (7 - 2x_3 - 5x_5) + x_3 + 2(-3 + x_5) + x_5 = 3 - x_3 - 2x_5$ .

解 (向量形式):

$$\mathbf{x} = \begin{pmatrix} 3 \\ 7 \\ 0 \\ -3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -2 \\ -5 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

2.2. 解: 该方程对应的线性系统为:

$$\begin{cases} x_1 + x_3 = 0 \\ x_1 + x_2 = 0 \\ x_2 + x_3 = 0 \end{cases}$$

增广矩阵化简:

$$\left( \begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array} \right) \xrightarrow{R_2-R_1} \left( \begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{array} \right) \xrightarrow{R_3-R_2} \left( \begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 2 & 0 \end{array} \right)$$

从最后一行得  $2x_3 = 0 \Rightarrow x_3 = 0$ . 进而  $x_2 = 0, x_1 = 0$ . 因此, 唯一的解是平凡解  $\mathbf{x} = \mathbf{0}$  (即  $x_1 = x_2 = x_3 = 0$ ).

**结论:** 因为齐次方程只有平凡解, 所以向量  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  是线性无关的。

3.1. 解: 我们对增广矩阵进行行约简:

$$\left( \begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 2 & 4 & 6 & 4 \\ 1 & 2 & 3 & b \end{array} \right)$$

执行行运算  $R_2 - 2R_1$  和  $R_3 - R_1$ :

$$\left( \begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 1 & b-1 \end{array} \right)$$

进一步对第二行除以 2, 得到  $R'_2$ :

$$\left( \begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & b-1 \end{array} \right)$$

执行  $R_3 - R'_2$ :

$$\left( \begin{array}{ccc|c} 1 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & b-2 \end{array} \right)$$

系统有解 (一致) 当且仅当增广列中没有主元, 即最后一行对应的方程  $0 = b - 2$  成立. 因此, 必须有  $b = 2$ .

当  $b = 2$  时, 矩阵变为简化阶梯形 (进一步化简  $R_1 - 2R'_2$ ):

$$\left( \begin{array}{ccc|c} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

主元在第 1 列和第 3 列.  $x_2$  是自由变量.  $x_3 = 1, x_1 + 2x_2 = -1 \Rightarrow x_1 = -1 - 2x_2$ . 通解为:

$$\mathbf{x} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$$

3.2. 解: 我们将这些向量作为列构成矩阵  $A$ , 并进行行约简以寻找主元位置.

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_2-R_1} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\xrightarrow{R_3+R_2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{R_4-R_3} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

结果表明矩阵只有 3 个主元。

1. **线性无关性**：因为只有 3 个主元，而有 4 个向量（列），必然存在自由变量。因此向量是**线性相关**的。

2. **张成**  $\mathbb{R}^4$ ：因为只有 3 个主元，最后一行全为 0，说明这些向量不能生成整个 4 维空间。它们只能张成  $\mathbb{R}^4$  中的一个 3 维子空间。

3. **对于**  $\mathbb{C}^4$ ：同样不能张成。因为维数不够（只有 3 个线性无关的向量）。

3.3. 解：  $\mathbb{R}^3$  中的 3 个向量构成基，当且仅当它们线性无关（即对应矩阵的行列式非零或有 3 个主元）。

a) 计算行列式：

$$\det \begin{pmatrix} 1 & 1 & 2 \\ 2 & 0 & 1 \\ -1 & 2 & 1 \end{pmatrix} = 1(0-2) - 1(2+1) + 2(4-0) = -2 - 3 + 8 = 3 \neq 0$$

是基。

b) 观察或计算。

$$\begin{pmatrix} -1 & -3 & 2 \\ 3 & 1 & 10 \\ 2 & 3 & 2 \end{pmatrix} \xrightarrow{R_2+3R_1, R_3+2R_1} \begin{pmatrix} -1 & -3 & 2 \\ 0 & -8 & 16 \\ 0 & -3 & 6 \end{pmatrix}$$

第二行对应  $-8x_2 + 16x_3 = 0 \implies x_2 = 2x_3$ 。第三行对应  $-3x_2 + 6x_3 = 0 \implies x_2 = 2x_3$ 。这两行成比例，存在自由变量。线性相关，不是基。

c) 将向量作为行或列排列，容易看出的三角结构（如果重新排序向量：第三个，第二个，第一个）：

$$\det \begin{pmatrix} 3 & \pi & 67 \\ 0 & -7.84 & 13 \\ 0 & 0 & -47 \end{pmatrix} = 3 \times (-7.84) \times (-47) \neq 0$$

线性无关，是基。

关于  $\mathbb{C}^3$ ：基的定义不依赖于标量域是  $\mathbb{R}$  还是  $\mathbb{C}$ （只要向量本身的元素属于该域）。因为系统 (a) 和 (c) 在代数上是线性无关的，它们也是  $\mathbb{C}^3$  的基。

3.4. 解：不能。空间  $\mathbb{P}_3$ （次数不超过 3 的多项式）的维数是 4（标准基为  $1, t, t^2, t^3$ ）。题目中只给出了 3 个向量（多项式）。根据线性代数的基本理论，维数为  $n$  的空间不可能由少于  $n$  个向量生成。

3.5. 解：不可能。 $\mathbb{F}^4$  的维数是 4。根据命题 3.2（或基本定理）， $\mathbb{F}^n$  中任何线性无关的向量组最多包含  $n$  个向量。5 个向量必然线性相关。

3.6. 证明：这是真的。如果  $n \times n$  矩阵  $A$  的列线性无关，则  $A$  是可逆的（命题 3.6）。两个可逆矩阵的乘积也是可逆的。因此  $A^2 = A \cdot A$  是可逆的。可逆矩阵的列必然是线性无关的。

3.7. 证明：这是真的。理由同上： $A$  的列线性无关  $\implies A$  可逆。 $\implies A^3$  可逆。 $\implies \det(A^3) \neq 0$ 。 $\implies \det((A^3)^T) = \det(A^3) \neq 0$ 。 $\implies (A^3)^T$  是可逆的。 $\implies (A^3)^T$  的列是线性无关的。注意  $(A^3)^T$  的列就是  $A^3$  的行。因此  $A^3$  的行是线性无关的。

3.8. 证明：设  $A$  是  $m \times n$  矩阵。因为  $A$  的每一列都有主元，所以主元的数量为  $n$ 。这意味着  $A$  的简化阶梯形  $R$  的前  $n$  行构成一个  $n \times n$  的单位矩阵  $I_n$ ，而（如果  $m > n$ ）下方的  $m-n$  行全为零。即

$$R = \begin{pmatrix} I_n \\ \mathbf{0} \end{pmatrix}$$

行约简的过程等价于左乘一系列初等矩阵。设这些初等矩阵的乘积为  $E$ （ $E$  是  $m \times m$  的可逆矩阵）。则有：

$$EA = R = \begin{pmatrix} I_n \\ \mathbf{0} \end{pmatrix}$$

我们将  $E$  分块为  $E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}$ , 其中  $E_1$  是  $n \times m$  矩阵,  $E_2$  是  $(m-n) \times m$  矩阵。执行矩阵乘法:

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} A = \begin{pmatrix} E_1 A \\ E_2 A \end{pmatrix} = \begin{pmatrix} I_n \\ \mathbf{0} \end{pmatrix}$$

比较上部分块, 得  $E_1 A = I_n$ . 因此,  $E_1$  就是  $A$  的一个左逆。

3.9. 解: 是的, 简化阶梯形是唯一的。

**理由:** 主元列的位置由原矩阵列向量之间的线性依赖关系唯一确定, 这与行运算的过程无关。具体来说, 第  $k$  列是主元列, 当且仅当第  $k$  列向量不能写成前  $k-1$  列向量的线性组合。这完全由向量本身的性质决定。

对于非主元列, 在简化阶梯形中, 该列的数值表示该列向量如何被其左侧的主元列向量线性表示的 (唯一的) 系数。因为在一组确定的基 (这里是主元列) 下, 坐标是唯一的, 所以简化阶梯形中的数值也是唯一的。

因此, 无论采用何种行运算顺序, 只要最终满足简化阶梯形的定义, 得到的矩阵必然相同。

4.1. 解: 对于第一个矩阵: 设  $A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 7 & 3 \\ 2 & 3 & 4 \end{pmatrix}$ . 构造增广矩阵  $(A|I)$ :

$$\left( \begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 3 & 7 & 3 & 0 & 1 & 0 \\ 2 & 3 & 4 & 0 & 0 & 1 \end{array} \right)$$

执行行运算  $R_2 \leftarrow R_2 - 3R_1$  和  $R_3 \leftarrow R_3 - 2R_1$  以消去第一列下方的元素:

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

执行行运算  $R_3 \leftarrow R_3 + R_2$  以消去第二列下方的元素:

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

此时矩阵已通过高斯消元变为上三角形式。现在我们需要进行回代或进一步行约简以得到单位矩阵。将第三行除以 2 ( $R_3 \leftarrow \frac{1}{2}R_3$ ):

$$\left( \begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & -5/2 & 1/2 & 1/2 \end{array} \right)$$

利用第三行消去第一行的第三个元素 ( $R_1 \leftarrow R_1 - R_3$ ):

$$\left( \begin{array}{ccc|ccc} 1 & 2 & 0 & 7/2 & -1/2 & -1/2 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & -5/2 & 1/2 & 1/2 \end{array} \right)$$

利用第二行消去第一行的第二个元素 ( $R_1 \leftarrow R_1 - 2R_2$ )。第一行右侧变为:

$$(7/2, -1/2, -1/2) - 2(-3, 1, 0) = (7/2 + 6, -1/2 - 2, -1/2) = (19/2, -5/2, -1/2).$$

$$\left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 19/2 & -5/2 & -1/2 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & -5/2 & 1/2 & 1/2 \end{array} \right)$$

因此, 逆矩阵为:

$$A^{-1} = \begin{pmatrix} 19/2 & -5/2 & -1/2 \\ -3 & 1 & 0 \\ -5/2 & 1/2 & 1/2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 19 & -5 & -1 \\ -6 & 2 & 0 \\ -5 & 1 & 1 \end{pmatrix}$$

对于第二个矩阵: 设  $B = \begin{pmatrix} 1 & -1 & 2 \\ 1 & 1 & -2 \\ 1 & 1 & 4 \end{pmatrix}$ . 构造增广矩阵  $(B|I)$ :

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

执行  $R_2 \leftarrow R_2 - R_1$  和  $R_3 \leftarrow R_3 - R_1$ :

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

执行  $R_3 \leftarrow R_3 - R_2$ :

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

将对角线元素归一化:  $R_2 \leftarrow \frac{1}{2}R_2$ ,  $R_3 \leftarrow \frac{1}{6}R_3$ :

$$\left( \begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & -2 & -1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & -1/6 & 1/6 \end{array} \right)$$

消去上方元素。先用  $R_3$  消去  $R_2$  和  $R_1$  中的第三列元素。 $R_2 \leftarrow R_2 + 2R_3$ : 右侧:  $(-1/2, 1/2, 0) + (0, -2/6, 2/6) = (-1/2, 3/6 - 2/6, 2/6) = (-1/2, 1/6, 1/3)$ .  $R_1 \leftarrow R_1 - 2R_3$ : 右侧:  $(1, 0, 0) - (0, -2/6, 2/6) = (1, 1/3, -1/3)$ . 矩阵变为:

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

最后, 用  $R_2$  消去  $R_1$  中的第二列元素 ( $R_1 \leftarrow R_1 + R_2$ ): 右侧:  $(1, 1/3, -1/3) + (-1/2, 1/6, 1/3) = (1/2, 3/6, 0) = (1/2, 1/2, 0)$ .

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

因此, 逆矩阵为:

$$B^{-1} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ -1/2 & 1/6 & 1/3 \\ 0 & -1/6 & 1/6 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 3 & 3 & 0 \\ -3 & 1 & 2 \\ 0 & -1 & 1 \end{pmatrix}$$

- 5.1. **a) 正确。**这是第1章命题2.8或本节引用的结论: 任何有限生成集都包含一组基。  
**b) 错误。**只有有限维向量空间才有有限基。无限维向量空间(如所有多项式的空间 $\mathbb{P}$ )没有有限基。  
**c) 错误。**一个向量空间(除非是零空间)有无穷多个不同的基。例如在 $\mathbb{R}^2$ 中,  $\{(1,0)^T, (0,1)^T\}$ 是基,  $\{(1,1)^T, (1,-1)^T\}$ 也是基。  
**d) 正确。**这是维数的定义基础(命题3.3), 即基中向量的数量是不变量。  
**e) 错误。** $\mathbb{P}_n$ (次数不超过 $n$ 的多项式)的基是 $\{1, t, t^2, \dots, t^n\}$ , 共有 $n+1$ 个向量。维数是 $n+1$ 。  
**f) 错误。** $M_{m \times n}$ 的标准基包含 $m \times n$ 个矩阵(每个位置一个1, 其余为0)。维数是 $mn$ 。  
**g) 错误。**只有当向量组是基(即既生成空间又线性无关)时, 表示才是唯一的。如果仅是生成(张成), 可能存在多余的向量, 导致表示不唯一。  
**h) 正确。**这是定理5.5。  
**i) 正确。**0维子空间只能是零空间 $\{0\}$ ,  $n$ 维子空间必须是 $V$ 本身(定理5.5)。



- 5.2. 证明: 设系统  $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  包含  $n$  个向量, 且  $\dim V = n$ .  
 $(\Rightarrow)$  假设  $\mathcal{S}$  是线性无关的. 根据命题 5.4 (补全为基), 我们可以将任何线性无关集补全为基. 但基的大小必须等于  $\dim V = n$ . 由于  $\mathcal{S}$  已经包含  $n$  个向量, 无法再添加向量. 因此  $\mathcal{S}$  本身就是一组基, 这意味着它张成  $V$ .  
 $(\Leftarrow)$  假设  $\mathcal{S}$  张成  $V$ . 根据命题 5.3 (或 5.1), 任何生成集都包含一组基. 基的大小必须是  $n$ . 由于  $\mathcal{S}$  恰好包含  $n$  个向量, 如果删去任何一个向量, 其大小将小于  $n$ , 无法构成基. 因此  $\mathcal{S}$  本身必须是基, 这意味着它是线性无关的.
- 5.3. 证明: 设  $\mathcal{S} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  是线性无关的.  
 $(\Rightarrow)$  如果  $\mathcal{S}$  是基, 根据维数的定义,  $\dim V$  等于基中向量的个数, 即  $\dim V = n$ .  
 $(\Leftarrow)$  如果  $n = \dim V$ . 我们已知  $\mathcal{S}$  是线性无关的. 根据习题 5.2 的结论, 在  $n$  维空间中,  $n$  个线性无关的向量必然生成该空间, 因此它们构成一组基.
- 5.4. 解: 不可能. 设  $V' = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ . 因为  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  线性相关, 所以它们不能构成  $V'$  的基. 这意味着  $V'$  的维数必然小于 3 (即  $\dim V' \leq 2$ ). 向量  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  是  $\mathbf{v}_i$  的线性组合, 因此  $\mathbf{w}_i \in V'$ . 如果在  $V'$  中存在 3 个线性无关的向量  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ , 根据命题 5.2, 这将意味着  $\dim V' \geq 3$ . 这与  $\dim V' \leq 2$  矛盾. 因此  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  必然是线性相关的.
- 5.5. 证明: 因为  $\dim V = 3$  (原基有 3 个向量), 我们要证明这 3 个新向量构成基, 只需证明它们线性无关即可 (根据习题 5.2). 考察线性组合为零的情况:

$$x_1(\mathbf{u} + \mathbf{v} + \mathbf{w}) + x_2(\mathbf{v} + \mathbf{w}) + x_3(\mathbf{w}) = \mathbf{0}$$

整理各项:

$$x_1\mathbf{u} + (x_1 + x_2)\mathbf{v} + (x_1 + x_2 + x_3)\mathbf{w} = \mathbf{0}$$

由于  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  是基, 它们线性无关, 因此对应系数必须全为 0:

$$\begin{cases} x_1 = 0 \\ x_1 + x_2 = 0 \\ x_1 + x_2 + x_3 = 0 \end{cases}$$

由第一个方程得  $x_1 = 0$ . 代入第二个得  $x_2 = 0$ . 代入第三个得  $x_3 = 0$ . 只有平凡解, 故向量线性无关. 因为向量个数 (3) 等于空间维数, 所以它们构成一组基.

- 5.6. 解: **a) 和 b)** 我们可以同时解决. 观察向量的“最后一个非零分量”的位置 (反向阶梯形):  $\mathbf{v}_1$  的第 5 个分量是  $-3 \neq 0$ , 而  $\mathbf{v}_2, \mathbf{v}_3$  的第 5 分量均为 0.  $\mathbf{v}_3$  的第 4 个分量是  $-921 \neq 0$ , 而  $\mathbf{v}_2$  的第 4 分量为 0 ( $\mathbf{v}_1$  无所谓, 已被区分).  $\mathbf{v}_2$  的第 2 个分量是  $-2 \neq 0$  (且 3, 4, 5 分量为 0).

这种阶梯状结构表明它们是线性无关的. 我们可以通过补全“缺失的台阶”来构造基. 目前, “非零结尾”占据了坐标索引 5 ( $\mathbf{v}_1$ ), 4 ( $\mathbf{v}_3$ ) 和 2 ( $\mathbf{v}_2$ ). 我们需要涵盖坐标索引 1 和 3. 我们可以添加标准基向量:  $\mathbf{e}_1 = (1, 0, 0, 0, 0)^T$  (在索引 1 处非零, 后续为 0).  $\mathbf{e}_3 = (0, 0, 1, 0, 0)^T$  (在索引 3 处非零, 后续为 0).

为了验证, 我们将这 5 个向量按  $\mathbf{e}_1, \mathbf{v}_2, \mathbf{e}_3, \mathbf{v}_3, \mathbf{v}_1$  的顺序作为列构成矩阵:

$$M = \begin{pmatrix} 1 & 3 & 0 & 1 & 2 \\ 0 & -2 & 0 & 1 & -1 \\ 0 & 0 & 1 & 50 & 1 \\ 0 & 0 & 0 & -921 & 5 \\ 0 & 0 & 0 & 0 & -3 \end{pmatrix}$$

这是一个上三角矩阵, 且对角线元素  $(1, -2, 1, -921, -3)$  均不为零. 因此行列式不为零, 这 5 个向量线性无关.

**结论:** a) 原向量线性无关. b) 补全后的基为  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_1, \mathbf{e}_3$ .

- 6.1. **a) 错误.** 线性方程组可能是不一致的 (无解), 例如  $0x_1 + 0x_2 = 1$ .  
**b) 错误.** 如果有自由变量, 系统可能有无穷多解.  
**c) 正确.** 齐次系统  $A\mathbf{x} = \mathbf{0}$  总是至少有一个解, 即平凡解  $\mathbf{x} = \mathbf{0}$ .

**d) 错误。**即使方程个数等于未知数个数，如果方程之间线性相关且常数项不匹配（如平行线），系统也可能无解。

**e) 错误。**如果矩阵不可逆（行列式为 0），可能存在无穷多解。

**f) 错误。**齐次方程组**总是**有解的（零解）。这个前提条件对任何线性系统都成立，但这并不能保证非齐次系统  $A\mathbf{x} = \mathbf{b}$  有解（取决于  $\mathbf{b}$  是否在列空间中）。

**g) 正确。**如果系数矩阵可逆，则  $A\mathbf{x} = \mathbf{0} \implies \mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$ 。只有零解。

**h) 错误。**只有当方程组是齐次的时候（右侧为  $\mathbf{0}$ ），解集才包含零向量并对加法/数乘封闭。非齐次方程组的解集不包含零向量，因此不是子空间（它是仿射子空间）。

**i) 正确。**这是核（Kernel）或零空间（Null space）的定义，它是  $\mathbb{R}^n$  的子空间。

- 6.2. 解：通解的形式为  $\mathbf{x} = \mathbf{x}_p + s\mathbf{x}_h$ ，其中  $\mathbf{x}_p = (1, 1, 0)^T$  是特解， $\mathbf{x}_h = (1, 2, 1)^T$  是齐次方程  $A\mathbf{x} = \mathbf{0}$  的解。我们需要构造一个  $2 \times 3$  矩阵  $A$  和向量  $\mathbf{b}$ 。

**第一步：寻找矩阵  $A$**  由于齐次解包含  $\mathbf{x}_h = (1, 2, 1)^T$ ，这意味着矩阵  $A$  的每一行必须与向量  $(1, 2, 1)^T$  正交（点积为 0）。我们需要找到两个线性无关的行向量  $\mathbf{r}_1, \mathbf{r}_2$  满足这一条件。设  $\mathbf{r} = (a, b, c)$ ，则  $a + 2b + c = 0$ 。我们可以选取简单的整数解：

1. 取  $b = 0, a = 1, c = -1 \implies \mathbf{r}_1 = (1, 0, -1)$ 。验证： $1(1) + 0(2) + (-1)(1) = 0$ 。

2. 取  $a = 0, b = 1, c = -2 \implies \mathbf{r}_2 = (0, 1, -2)$ 。验证： $0(1) + 1(2) + (-2)(1) = 0$ 。

这两行显然线性无关。因此：

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \end{pmatrix}$$

**第二步：寻找向量  $\mathbf{b}$**  由于  $\mathbf{x}_p = (1, 1, 0)^T$  是解，代入  $A\mathbf{x} = \mathbf{b}$ ：

$$\mathbf{b} = A \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1(1) + 0(1) + (-1)(0) \\ 0(1) + 1(1) + (-2)(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

结果所求的系统为：

$$\begin{cases} x_1 - x_3 = 1 \\ x_2 - 2x_3 = 1 \end{cases}$$

或者写成矩阵形式：

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- 7.1. **a) 错误。**秩等于**线性无关**的非零列的数量（即主元列的数量），而不仅仅是非零列的数量。如果一个非零列是另一个的倍数，它们只贡献 1 个秩。

**b) 正确。**只有零矩阵的所有子式都为 0，秩为 0。

**c) 正确。**初等行运算将行向量变为其线性组合，行空间保持不变，因此行秩不变。由秩定理，列秩也不变。

**d) 错误。**初等列运算改变了列空间，但**保持**列之间的线性相关性关系，因此它**保持秩**。如果题目意思是“列运算不改变列空间”，那是错的；但说它不保持秩，则命题本身是错误的。

**e) 正确。**这是列秩的定义。

**f) 正确。**这是行秩的定义，由秩定理知它等于秩。

**g) 正确。** $n \times n$  矩阵的秩最大为  $n$ （满秩）。

**h) 正确。**秩为  $n$  意味着  $A$  行等价于  $I_n$ ，或者是可逆的。

- 7.2. 解：设矩阵  $A$  大小为  $m \times n = 54 \times 37$ ，秩  $r = 31$ 。

1.  $\dim \text{Ran } A = r = 31$ 。

2.  $\dim \text{Ran } A^T = r = 31$ 。

3.  $\dim \text{Ker } A = n - r = 37 - 31 = 6$ 。

4.  $\dim \text{Ker } A^T = m - r = 54 - 31 = 23$ 。

7.3. 解: 对于第一个矩阵  $A$ :

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \xrightarrow{R_3 - R_1} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

秩为 2。

**Ran  $A$  的基** (原矩阵的主元列):  $(1, 0, 1)^T, (1, 1, 1)^T$ .

**Ran  $A^T$  的基** (阶梯形的非零行):  $(1, 1, 0)^T, (0, 1, 1)^T$ .

**Ker  $A$  的基**: 解方程  $Ax = 0$ .  $x_2 + x_3 = 0 \Rightarrow x_2 = -x_3$ .  $x_1 + x_2 = 0 \Rightarrow x_1 = -x_2 = x_3$ . 取  $x_3 = 1$ , 得基向量  $(1, -1, 1)^T$ .

**Ker  $A^T$  的基**: 寻找行之间的线性关系。观察原矩阵, 容易发现  $R_3 = R_1$ , 即  $R_1 - R_3 = 0$ . 或者计算  $A^T$  的核。系数向量为  $(1, 0, -1)^T$ .

对于第二个矩阵  $B$ :

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

观察  $R_4$ , 第一列是主元列。  $R_1$  和  $R_2$  的第一列可以被消去。注意到  $R_1 = (1, 2, 3, 1, 1)$ .  $R_2 = (1, 4, 0, 1, 2)$ .  $R_2 - R_1 = (0, 2, -3, 0, 1)$ , 这恰好是  $R_3$ . 所以  $R_1 - R_2 + R_3 = 0$ . 这一行是多余的。由于  $R_4$  显然与其他行无关 (在第 2-5 列为 0), 且  $R_1, R_3$  显然无关。所以秩为 3。

**Ran  $B$  的基**: 原矩阵的第 1, 2, 3 列 (或其他任何 3 个线性无关列)。  $(1, 1, 0, 1)^T, (2, 4, 2, 0)^T, (3, 0, -3, 0)^T$ .

**Ran  $B^T$  的基**: 行空间。取  $R_4, R_3, R_1$ .  $(1, 0, 0, 0, 0)^T, (0, 2, -3, 0, 1)^T, (1, 2, 3, 1, 1)^T$ .

**Ker  $B^T$  的基**: 行的线性依赖关系。由  $R_1 - R_2 + R_3 = 0$ , 系数为  $(1, -1, 1, 0)^T$ .

**Ker  $B$  的基**: 使用行约简后的方程: 1)  $x_1 = 0$  (由  $R_4$ ). 2)  $2x_2 - 3x_3 + x_5 = 0$ . 3)  $2x_2 + 3x_3 + x_4 + x_5 = 0$  (来自  $R_1$ , 代入  $x_1 = 0$ ). 两式相减:  $(2x_2 + 3x_3 + x_4 + x_5) - (2x_2 - 3x_3 + x_5) = 6x_3 + x_4 = 0 \Rightarrow x_4 = -6x_3$ . 代回 2):  $2x_2 = 3x_3 - x_5 \Rightarrow x_2 = \frac{3}{2}x_3 - \frac{1}{2}x_5$ . 自由变量  $x_3, x_5$ . 令  $x_3 = 2, x_5 = 0 \Rightarrow x_2 = 3, x_4 = -12$ . 向量  $v_1 = (0, 3, 2, -12, 0)^T$ . 令  $x_3 = 0, x_5 = 2 \Rightarrow x_2 = -1, x_4 = 0$ . 向量  $v_2 = (0, -1, 0, 0, 2)^T$ .

7.4. 证明:  $AV$  定义为  $\{Av : v \in V\}$ . 因为  $V \subset X$ , 显然  $AV \subset \text{Ran } A$ . 因为  $AV$  是  $\text{Ran } A$  的子空间, 所以其维数不能超过  $\text{Ran } A$  的维数。即  $\dim AV \leq \dim \text{Ran } A = \text{rank } A$ .

对于  $AB$ , 其像空间为  $\text{Ran}(AB) = A(\text{Ran } B)$ . 令  $V = \text{Ran } B$ , 则  $V$  是  $A$  定义域的子空间。应用上述结论:  $\text{rank}(AB) = \dim(A(\text{Ran } B)) \leq \text{rank } A$ .

7.5. 证明: 设  $v_1, \dots, v_k$  是  $V$  的一组基, 则  $\dim V = k$ .  $AV$  中的任意向量  $y$  可写为  $y = Av = A(\sum c_i v_i) = \sum c_i (Av_i)$ . 这表明  $Av_1, \dots, Av_k$  生成 (张成)  $AV$ . 由第 5 节知, 生成的子空间维数  $\leq$  生成元的数量。所以  $\dim AV \leq k = \dim V$ .

对于  $AB$ ,  $\text{rank}(AB) = \dim A(\text{Ran } B)$ . 令  $V = \text{Ran } B$ , 由上述结论:  $\text{rank}(AB) = \dim AV \leq \dim V = \dim \text{Ran } B = \text{rank } B$ .

7.6. 证明: 如果  $AB$  可逆, 则  $\text{rank}(AB) = n$ . 由 7.4 和 7.5 的结论:  $n = \text{rank}(AB) \leq \text{rank } A \leq n \Rightarrow \text{rank } A = n \Rightarrow A$  可逆。  $n = \text{rank}(AB) \leq \text{rank } B \leq n \Rightarrow \text{rank } B = n \Rightarrow B$  可逆。

7.7. 证明: 设  $A$  是  $m \times n$  矩阵。  $Ax = 0$  只有唯一解 (平凡解)  $\iff \dim \text{Ker } A = 0$ . 由秩定理 (7.2),  $\dim \text{Ran } A = n - \dim \text{Ker } A = n - 0 = n$ . 所以  $\text{rank } A = n$ . 由定理 7.1,  $A^T$  的秩 (即  $\dim \text{Ran } A^T$ ) 也等于  $n$ . 注意  $A^T$  是  $n \times m$  矩阵, 它将向量从  $\mathbb{R}^m$  映射到  $\mathbb{R}^n$ . 因为  $\text{Ran } A^T$  的维数是  $n$ , 且它是  $\mathbb{R}^n$  的子空间, 所以  $\text{Ran } A^T$  必须是整个  $\mathbb{R}^n$ . 这意味着对于任何  $b \in \mathbb{R}^n$ , 都存在  $x$  使得  $A^T x = b$ .

7.8. a) 列空间包含  $e_1, e_3$  (在  $\mathbb{R}^3$  中), 行空间包含  $(1, 1)^T, (1, 2)^T$  (在  $\mathbb{R}^2$  中)。

解: 这意味着矩阵是  $3 \times 2$  的。秩至少是 2 (因为  $e_1, e_3$  线性无关)。  $3 \times 2$  矩阵的最大秩也是

2。我们可以简单地让列空间就是  $\text{span}(e_1, e_3)$ , 行空间就是  $\mathbb{R}^2$ . 例如:  $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$ . 列

是  $e_1, e_3$ . 行是  $(1, 0), (0, 0), (0, 1)$ , 生成  $\mathbb{R}^2$ , 显然包含  $(1, 1), (1, 2)$ .

b) 列空间由  $(1, 1, 1)^T$  张成, 零空间由  $(1, 2, 3)^T$  张成。

解: 不存在。矩阵是  $3 \times 3$ . 列空间维数为 1  $\Rightarrow$  秩  $r = 1$ . 零空间由 1 个向量生成  $\Rightarrow$  零化度  $k = 1$  (除非该向量为 0, 这里不是)。根据定理,  $r + k = n = 3$ . 但这里  $1 + 1 = 2 \neq 3$ .

c) 列空间是  $\mathbb{R}^4$ , 行空间是  $\mathbb{R}^3$ .

解: 不存在。列空间是  $\mathbb{R}^4 \Rightarrow$  秩为 4. 行空间是  $\mathbb{R}^3 \Rightarrow$  秩为 3. 由定理 7.1, 行秩必须等于列秩。

7.9. 解: 不成立。反例: 取  $A = I$  (单位矩阵),  $B = 2I$ . 它们的列空间、行空间都是全空间, 零空间、左零空间都是零空间。但  $A \neq B$ .

7.10. 解: 观察矩阵的阶梯结构 (忽略具体的数值, 只看非零模式):  $R_4 = (0, \dots, 0, 1)$ , 主元在第 7 列。  $R_3 = (0, \dots, 0, 3, -3, 2)$ , 主元在第 5 列。  $R_2 = (0, 0, 2, \dots)$ , 主元在第 3 列。  $R_1 = (e^3, 3, \dots)$ , 主元在第 1 列。 现有的主元位置: 1, 3, 5, 7. 我们需要增加主元位置在 2, 4, 6 的行。 最简单的方法是添加标准基向量  $\mathbf{e}_2, \mathbf{e}_4, \mathbf{e}_6$ . 这些行将填补阶梯形的空缺, 使整体构成一个上三角矩阵且对角线非零, 从而线性无关。

7.11. 解: 进行行约简:  $R_2 - 2R_1 \rightarrow (0, -2, 3, 1, -1)$ .  $R_3 - 3R_1 \rightarrow (0, 0, 0, -6, 15)$ .  $R_4 + R_1 \rightarrow (0, -2, 3, -5, 14)$ . 接着  $R_4 - R_2 \rightarrow (0, 0, 0, -6, 15)$ . 显然  $R_4$  现在与  $R_3$  相同, 可以消去。 阶梯形主元在第 1, 2, 4 列。

列空间基 (原矩阵对应列):  $\mathbf{v}_1 = (1, 2, 3, -1)^T, \mathbf{v}_2 = (2, 2, 6, -4)^T, \mathbf{v}_4 = (2, 5, 0, -7)^T$ .

行空间基 (阶梯形的非零行):  $\mathbf{r}_1 = (1, 2, -1, 2, 3)^T, \mathbf{r}_2 = (0, -2, 3, 1, -1)^T, \mathbf{r}_3 = (0, 0, 0, -6, 15)^T$  (或化简为  $(0, 0, 0, 2, -5)^T$ )。

7.12. 解: 行空间的基的主元位置在 1, 2, 4. 缺失的主元位置是 3, 5. 添加标准基向量  $\mathbf{e}_3 = (0, 0, 1, 0, 0)^T$  和  $\mathbf{e}_5 = (0, 0, 0, 0, 1)^T$ . 这 5 个向量构成  $\mathbb{R}^5$  的基。

7.13. 解: 行约简:  $R_2 - iR_1 \Rightarrow i - i(1) = 0, -1 - i(i) = -1 - (-1) = 0$ . 得到  $\begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}$ .

Ran  $A$ : 由第一列生成,  $\text{span}((1, i)^T)$ .

Ker  $A$ :  $x_1 + ix_2 = 0 \Rightarrow x_1 = -ix_2$ .

取  $x_2 = i \Rightarrow x_1 = 1$ . 或者取  $x_2 = 1, x_1 = -i$ . 基向量为  $(-i, 1)^T$ . 注意到  $(-i, 1)^T = -i(1, i)^T$ . 所以  $\text{Ran } A = \text{Ker } A$ . 关系: 它们是相同的子空间。

7.14. 解: **实数矩阵: 不可能** (除非  $A$  是零矩阵)。原因: 对于实矩阵, 我们有基本定理  $\text{Ran } A \perp \text{Ker } A^T$  (即  $\text{Ran } A$  中的向量与  $\text{Ker } A^T$  中的向量正交)。如果两个子空间相等, 设为  $V$ , 则  $V \perp V$ . 这意味着对于任意  $\mathbf{v} \in V, \mathbf{v} \cdot \mathbf{v} = 0 \Rightarrow \|\mathbf{v}\|^2 = 0 \Rightarrow \mathbf{v} = \mathbf{0}$ . 所以只有当  $V = \{\mathbf{0}\}$  时才可能。

**复数矩阵: 可能**. 习题 7.13 就是一个例子。在复数空间中, 正交性定义为  $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \overline{\mathbf{w}} = 0$  (或  $\mathbf{w}^* \mathbf{v} = 0$ )。而基本定理陈述的是  $\text{Ran } A \perp \text{Ker } A^*$  (共轭转置)。这里题目问的是  $\text{Ker } A^T$  (普通转置)。在 7.13 中,  $A$  是对称的 ( $A^T = A$ ), 所以  $\text{Ker } A^T = \text{Ker } A$ . 我们发现  $\text{Ran } A = \text{Ker } A$ . 向量  $(1, i)^T$  满足自点积  $(1)^2 + (i)^2 = 1 - 1 = 0$ , 这是“各向同性”向量。

7.15. 解: 将这三个向量作为矩阵的行, 检查独立性并寻找主元。

$$\begin{pmatrix} 1 & 2 & -1 & 2 & 3 \\ 2 & 2 & 1 & 5 & 5 \\ -1 & -4 & 4 & 7 & -11 \end{pmatrix}$$

$R_2 - 2R_1 \rightarrow (0, -2, 3, 1, -1)$ .  $R_3 + R_1 \rightarrow (0, -2, 3, 9, -8)$ .  $R'_3 - R'_2 \rightarrow (0, 0, 0, 8, -7)$ . 阶梯形显示非零行有 3 个, 且主元位置在 1, 2, 4. 这三个向量线性无关。为了补全为基, 我们需要在缺失主元的位置 (3 和 5) 添加向量。最简单的是添加  $\mathbf{e}_3 = (0, 0, 1, 0, 0)^T$  和  $\mathbf{e}_5 = (0, 0, 0, 0, 1)^T$ .

8.1. a) **正确**. 坐标变换矩阵  $[I]_{B,A}$  的大小是  $n \times n$ , 其中  $n$  是向量空间的维数 (基中向量的个数)。

b) **正确**. 坐标变换矩阵将一组基映射到另一组基, 它是可逆的, 其逆矩阵是反向的坐标变换矩阵 ( $[I]_{B,A}^{-1} = [I]_{A,B}$ )。

c) 错误。这是合同 (congruence) 的定义, 通常与二次型有关。相似性的定义涉及逆矩阵  $Q^{-1}$ 。

d) 正确。这是相似矩阵的定义。

e) 错误。相似性是针对线性算子  $T: V \rightarrow V$  定义的, 其矩阵表示必须是方阵。

8.2. 解: a) 将这四个向量作为列构成矩阵  $M$ :

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 3 & 1 \\ 1 & 3 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

我们要证明  $\det M \neq 0$ . 利用第一行展开行列式 (第一行只有第一个元素为 1, 其余为 0):

$$\det M = 1 \cdot \det \begin{pmatrix} 1 & 3 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

对于剩下的  $3 \times 3$  矩阵, 利用第三列展开 (只有第一个元素为 1, 其余为 0):

$$\det \begin{pmatrix} 1 & 3 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 1 \cdot \det \begin{pmatrix} 3 & 2 \\ 1 & 0 \end{pmatrix} = 1(0 - 2) = -2$$

因此  $\det M = -2 \neq 0$ , 向量线性无关, 构成  $\mathbb{F}^4$  的基。

b) 设题中给出的基为  $\mathcal{B}$ , 标准基为  $\mathcal{S}$ . 我们需要找到矩阵  $[I]_{\mathcal{S}\mathcal{B}}$ . 根据定义, 该矩阵的第  $k$  列就是基  $\mathcal{B}$  中第  $k$  个向量在标准基下的坐标。因为给出的向量本身就是标准坐标形式, 所以坐标变换矩阵就是上面的矩阵  $M$ :

$$[I]_{\mathcal{S}\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 3 & 1 \\ 1 & 3 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

8.3. 解: 设旧基  $\mathcal{A} = \{1, 1+t\}$ , 新基  $\mathcal{B} = \{1-t, 2t\}$ , 标准基  $\mathcal{S} = \{1, t\}$ . 我们需要计算  $[I]_{\mathcal{B}\mathcal{A}}$ . 利用公式  $[I]_{\mathcal{B}\mathcal{A}} = [I]_{\mathcal{B}\mathcal{S}}[I]_{\mathcal{S}\mathcal{A}} = ([I]_{\mathcal{S}\mathcal{B}})^{-1}[I]_{\mathcal{S}\mathcal{A}}$ . 先写出相对于标准基的矩阵:

$$[I]_{\mathcal{S}\mathcal{A}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad [I]_{\mathcal{S}\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$$

计算  $[I]_{\mathcal{S}\mathcal{B}}$  的逆:

$$([I]_{\mathcal{S}\mathcal{B}})^{-1} = \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$$

相乘得到结果:

$$[I]_{\mathcal{B}\mathcal{A}} = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1/2 & 1 \end{pmatrix}$$

检查: 对于  $\mathcal{A}$  中第一个向量 1:  $1(1-t) + \frac{1}{2}(2t) = 1 - t + t = 1$ . 正确。对于  $\mathcal{A}$  中第二个向量  $1+t$ :  $1(1-t) + 1(2t) = 1 - t + 2t = 1 + t$ . 正确。

8.4. 解: 1. 在标准基  $\mathcal{S}$  下的矩阵  $A$ :

$$T(\mathbf{e}_1) = T(1, 0)^T = (3, 1)^T$$

$$T(\mathbf{e}_2) = T(0, 1)^T = (1, -2)^T$$

$$A = [T]_{\mathcal{S}} = \begin{pmatrix} 3 & 1 \\ 1 & -2 \end{pmatrix}$$

2. 在新基  $\mathcal{B}$  下的矩阵  $B$ : 设新基矩阵为  $Q = [I]_{\mathcal{S}\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ . 我们需要计算  $B = Q^{-1}AQ$ . 首先求  $Q^{-1}$ :

$$Q^{-1} = \frac{1}{2-1} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

计算  $AQ$ :

$$AQ = \begin{pmatrix} 3 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3+1 & 3+2 \\ 1-2 & 1-4 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ -1 & -3 \end{pmatrix}$$

最后计算  $Q^{-1}(AQ)$ :

$$B = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 5 \\ -1 & -3 \end{pmatrix} = \begin{pmatrix} 8+1 & 10+3 \\ -4-1 & -5-3 \end{pmatrix} = \begin{pmatrix} 9 & 13 \\ -5 & -8 \end{pmatrix}$$

8.5. 证明: 如果  $A$  和  $B$  相似, 则存在可逆矩阵  $Q$  使得  $B = Q^{-1}AQ$ . 计算  $B$  的迹:

$$\text{trace } B = \text{trace}(Q^{-1}AQ)$$

利用迹的循环性质  $\text{trace}(XY) = \text{trace}(YX)$ . 令  $X = Q^{-1}$ ,  $Y = AQ$ .

$$\text{trace}(Q^{-1}(AQ)) = \text{trace}((AQ)Q^{-1})$$

结合律:

$$\text{trace}(A(QQ^{-1})) = \text{trace}(AI) = \text{trace } A$$

证毕。

8.6. 解: **不相似**。根据习题 8.5 的结论, 相似矩阵必须具有相同的迹 (trace)。第一个矩阵的迹为  $1+2=3$ 。第二个矩阵的迹为  $0+2=2$ 。因为  $3 \neq 2$ , 所以这两个矩阵不相似。

### 第三章习题解答

3.1. 解: 根据行列式的多重线性性质 (每一行都有线性性质), 如果我们把矩阵的一行乘以标量  $k$ , 行列式也乘以  $k$ 。矩阵  $3A$  意味着  $A$  的每一行都乘以了 3。由于  $A$  有  $n$  行, 我们需要提出  $n$  个 3。因此:

$$\det(3A) = 3^n \det A$$

3.2. 解: **a)** 矩阵  $B$  的第一列是  $A$  的第一列的 2 倍; 第二列是  $A$  的 3 倍; 第三列是  $A$  的 5 倍。根据行列式对每一列的线性性质:

$$\det B = 2 \cdot 3 \cdot 5 \det A = 30 \det A$$

$$\text{b) } A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, \quad B = \begin{pmatrix} 3a_1 & 4a_2 + 5a_1 & 5a_3 \\ 3b_1 & 4b_2 + 5b_1 & 5b_3 \\ 3c_1 & 4c_2 + 5c_1 & 5c_3 \end{pmatrix}.$$

解: 首先, 将  $B$  分解。第二列是一个线性组合。行列式关于第二列是线性的, 所以可以拆分为两项之和。但其中一项包含  $5 \times$  (第一列对应元素), 这与第一列 ( $3a_1$  等) 成比例, 因此该部分行列式为 0。只剩下  $4a_2$  部分。具体步骤如下:

1. 从第 1 列提取因子 3。

2. 从第 3 列提取因子 5。

3. 此时第 2 列为  $(4a_2 + 5a_1, \dots)^T$ 。利用列运算 ( $C_2 - \frac{5}{3}C_1$ ), 消去  $5a_1$  部分, 不改变行列式的值 (或者利用线性性质拆分)。

4. 从第 2 列提取因子 4。

总系数为  $3 \times 4 \times 5 = 60$ 。

$$\det B = 60 \det A$$

3.3. 解: **1. 第一个矩阵:**

$$\begin{vmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{vmatrix} = -1 \begin{vmatrix} -1 & -3 \\ 2 & 0 \end{vmatrix} + 2 \begin{vmatrix} -1 & 0 \\ 2 & 3 \end{vmatrix} = -1(0 - (-6)) + 2(-3 - 0) = -6 - 6 = -12$$

**2. 第二个矩阵:** 观察行之间的关系:  $R_2 - R_1 = (3, 3, 3)$ , 且  $R_3 - R_2 = (3, 3, 3)$ . 这意味着  $R_1, R_2, R_3$  是等差数列关系, 即  $R_1 + R_3 = 2R_2$ . 行线性相关。

$$\det = 0$$

**3. 第三个矩阵 ( $4 \times 4$ ):**

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

执行行运算:  $R_2 \leftarrow R_2 + 3R_1$ ,  $R_4 \leftarrow R_4 - 2R_1$ .

$$D = \begin{vmatrix} 1 & 0 & -2 & 3 \\ 0 & 1 & -5 & 11 \\ 0 & 4 & -1 & 1 \\ 0 & 3 & 4 & -5 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & -5 & 11 \\ 4 & -1 & 1 \\ 3 & 4 & -5 \end{vmatrix}$$

接着计算  $3 \times 3$  子行列式:  $R_2 \leftarrow R_2 - 4R_1$ ,  $R_3 \leftarrow R_3 - 3R_1$ .

$$\begin{vmatrix} 1 & -5 & 11 \\ 0 & 19 & -43 \\ 0 & 19 & -38 \end{vmatrix}$$

最后  $R_3 \leftarrow R_3 - R_2$ :

$$\begin{vmatrix} 1 & -5 & 11 \\ 0 & 19 & -43 \\ 0 & 0 & 5 \end{vmatrix} = 1 \cdot 19 \cdot 5 = 95$$

**4. 第四个矩阵:**

$$\begin{vmatrix} 1 & x \\ 1 & y \end{vmatrix} = 1 \cdot y - 1 \cdot x = y - x$$

3.4. 证明: 利用转置性质  $\det A = \det(A^T)$  和标量乘法性质  $\det(kA) = k^n \det A$ .

$$\det A = \det(A^T) = \det(-A) = \det((-1)A) = (-1)^n \det A$$

如果  $n$  是奇数, 则  $(-1)^n = -1$ . 于是  $\det A = -\det A \implies 2\det A = 0 \implies \det A = 0$ .

对于偶数  $n$ , 这不一定成立。例如  $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$  是反对称的 ( $n = 2$ ), 但  $\det A = 0 - (-1) = 1 \neq 0$ .

3.5. 证明: 利用行列式的乘法性质  $\det(XY) = \det X \det Y$ .

$$(\det A)^k = \underbrace{\det A \cdot \det A \cdots \det A}_{k \text{ 次}} = \det(A \cdot A \cdots A) = \det(A^k)$$

已知  $A^k = 0$  (零矩阵), 且零矩阵的行列式为 0.

$$(\det A)^k = 0 \implies \det A = 0$$

3.6. 证明: 如果  $A$  和  $B$  相似, 则存在可逆矩阵  $Q$  使得  $A = Q^{-1}BQ$ .

$$\det A = \det(Q^{-1}BQ) = \det(Q^{-1}) \det(B) \det(Q)$$

标量乘法满足交换律, 且  $\det(Q^{-1}) = 1/\det Q$ :

$$\det A = \det B \cdot \det(Q^{-1}) \det Q = \det B \cdot \frac{1}{\det Q} \cdot \det Q = \det B \cdot 1 = \det B$$

3.7. 证明: 对等式两边取行列式:

$$\det(Q^T Q) = \det(I)$$

$$\det(Q^T) \det(Q) = 1$$

因为  $\det(Q^T) = \det Q$ , 所以:

$$(\det Q)^2 = 1$$

因此  $\det Q = 1$  或  $\det Q = -1$ .

3.8. 证明: 执行行运算:  $R_2 \leftarrow R_2 - R_1, R_3 \leftarrow R_3 - R_1$ .

$$\begin{vmatrix} 1 & x & x^2 \\ 0 & y-x & y^2-x^2 \\ 0 & z-x & z^2-x^2 \end{vmatrix}$$

将差平方公式展开  $a^2 - b^2 = (a-b)(a+b)$ , 并利用行列式关于行的线性性质, 从第 2 行提取  $(y-x)$ , 从第 3 行提取  $(z-x)$ :

$$= (y-x)(z-x) \begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 1 & z+x \end{vmatrix}$$

针对第一列展开, 只剩右下角的  $2 \times 2$  行列式:

$$= (y-x)(z-x) \cdot 1 \cdot \begin{vmatrix} 1 & y+x \\ 1 & z+x \end{vmatrix}$$

计算该子行列式:  $(z+x) - (y+x) = z-y$ . 所以原式  $= (y-x)(z-x)(z-y)$ . 整理顺序即得  $(z-x)(z-y)(y-x)$ .

3.9. 证明: 执行行运算:  $R_2 \leftarrow R_2 - R_1, R_3 \leftarrow R_3 - R_1$ . 这相当于将三角形平移, 使顶点  $A$  移至原点. 行列式的值不变.

$$\frac{1}{2} \det \begin{pmatrix} 1 & x_1 & y_1 \\ 0 & x_2 - x_1 & y_2 - y_1 \\ 0 & x_3 - x_1 & y_3 - y_1 \end{pmatrix}$$

沿第一列展开:

$$\frac{1}{2} \cdot 1 \cdot \det \begin{pmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{pmatrix}$$

这就是由向量  $\vec{AB} = (x_2 - x_1, y_2 - y_1)$  和  $\vec{AC} = (x_3 - x_1, y_3 - y_1)$  构成的  $2 \times 2$  行列式. 我们知道由两个向量  $\mathbf{u}, \mathbf{v}$  张成的平行四边形的面积等于其行列式的绝对值. 三角形面积是平行四边形面积的一半. 得证.

3.10. 证明: 以  $\begin{pmatrix} A & * \\ \mathbf{0} & I \end{pmatrix}$  为例. 设  $A$  为  $n \times n$ ,  $I$  为  $m \times m$ . 对后  $m$  列 (单位矩阵所在的列) 进行拉普拉斯展开. 每次展开都选取  $I$  中的对角元 1, 去除了对应的行和列, 且符号为正 (对角线位置). 经过  $m$  次展开后, 剩下的就是矩阵  $A$ . 所以  $\det = 1 \cdot 1 \cdots 1 \cdot \det A = \det A$ . 其他情况同理, 或者是对行进行展开.

3.11. 证明: 根据提示中的分解以及行列式的乘法性质:

$$\det \begin{pmatrix} A & B \\ \mathbf{0} & C \end{pmatrix} = \det \left( \begin{pmatrix} I & B \\ \mathbf{0} & C \end{pmatrix} \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \right) = \det \begin{pmatrix} I & B \\ \mathbf{0} & C \end{pmatrix} \cdot \det \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}$$

利用习题 3.10 的结论: 第一项:  $\det \begin{pmatrix} I & B \\ \mathbf{0} & C \end{pmatrix} = \det C$  (这里  $I$  在左上角, 对应 3.10 的第一种情况但  $A$  换成了  $C$ ). 第二项:  $\det \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} = \det A$  (对应 3.10 的第四种情况). 所以结果为  $(\det C)(\det A) = (\det A)(\det C)$ .

3.12. 证明: 根据提示, 我们将原分块矩阵右乘给定的矩阵:

$$\begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix} \begin{pmatrix} I & \mathbf{0} \\ B & I \end{pmatrix} = \begin{pmatrix} \mathbf{0} \cdot I + A \cdot B & \mathbf{0} \cdot \mathbf{0} + A \cdot I \\ -B \cdot I + I \cdot B & -B \cdot \mathbf{0} + I \cdot I \end{pmatrix} = \begin{pmatrix} AB & A \\ \mathbf{0} & I \end{pmatrix}$$



现在取两边的行列式。左边：

$$\det \left( \begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix} \begin{pmatrix} I & \mathbf{0} \\ B & I \end{pmatrix} \right) = \det \begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix} \cdot \det \begin{pmatrix} I & \mathbf{0} \\ B & I \end{pmatrix}$$

注意到  $\begin{pmatrix} I & \mathbf{0} \\ B & I \end{pmatrix}$  是一个下三角矩阵，其对角线上全是 1，所以它的行列式是 1。因此左边等于  $\det \begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix}$ 。

右边：

$$\det \begin{pmatrix} AB & A \\ \mathbf{0} & I \end{pmatrix}$$

这是一个分块上三角矩阵，根据习题 3.10 和 3.11 的结论，其行列式等于对角块行列式的乘积：

$$\det(AB) \cdot \det(I) = \det(AB)$$

比较左右两边，得证：

$$\det \begin{pmatrix} \mathbf{0} & A \\ -B & I \end{pmatrix} = \det(AB)$$

- 4.1. 解：a) 我们可以通过计算逆序对 (inversions) 的数量来确定符号。逆序对是指满足  $i < j$  但  $\sigma(i) > \sigma(j)$  的数对  $(i, j)$ 。排列结果为：5, 4, 1, 2, 3。5 在 4, 1, 2, 3 之前 (4 个逆序) - 4 在 1, 2, 3 之前 (3 个逆序) - 1 没有逆序 - 2 没有逆序总逆序数为  $4 + 3 = 7$ 。因为 7 是奇数，所以  $\text{sign } \sigma = -1$ 。或者通过循环分解： $1 \rightarrow 5 \rightarrow 3 \rightarrow 1$  (循环长度 3)， $2 \rightarrow 4 \rightarrow 2$  (循环长度 2)。偶长度循环贡献 -1，奇长度循环贡献 +1 (或者：总长度 - 循环个数 =  $5 - 2 = 3$  (奇数)，即 3 次对换)。所以符号为 odd (-1)。
- b)  $\sigma$  的作用是：位置  $1 \rightarrow 5$ ，位置  $2 \rightarrow 4$ ，位置  $3 \rightarrow 1$ ，位置  $4 \rightarrow 2$ ，位置  $5 \rightarrow 3$ 。应用两次  $\sigma$ ： $1 \xrightarrow{\sigma} 5 \xrightarrow{\sigma} 3 \xrightarrow{\sigma} 2 \xrightarrow{\sigma} 4 \xrightarrow{\sigma} 1 \xrightarrow{\sigma} 5 \xrightarrow{\sigma} 4 \xrightarrow{\sigma} 2 \xrightarrow{\sigma} 4 \xrightarrow{\sigma} 5 \xrightarrow{\sigma} 3 \xrightarrow{\sigma} 1$  结果为 (3, 2, 5, 4, 1)。
- c)  $\sigma^{-1}$  是反向映射：原排列中 5 在位置 1  $\Rightarrow \sigma^{-1}(5) = 1$ 。原排列中 4 在位置 2  $\Rightarrow \sigma^{-1}(4) = 2$ 。原排列中 1 在位置 3  $\Rightarrow \sigma^{-1}(1) = 3$ 。原排列中 2 在位置 4  $\Rightarrow \sigma^{-1}(2) = 4$ 。原排列中 3 在位置 5  $\Rightarrow \sigma^{-1}(3) = 5$ 。对 (1, 2, 3, 4, 5) 的作用结果为 (3, 4, 5, 2, 1)。
- d) 逆排列的符号与原排列相同。因为  $\sigma\sigma^{-1} = \text{id}$ ，且  $\text{sign}(\sigma\tau) = \text{sign}(\sigma)\text{sign}(\tau)$ 。1 =  $\text{sign}(\text{id}) = \text{sign}(\sigma)\text{sign}(\sigma^{-1})$ 。所以  $\text{sign } \sigma^{-1} = \text{sign } \sigma = -1$ 。
- 4.2. 解：a) 设  $P$  的第  $j$  列在第  $k_j$  行有一个 1。那么  $P\mathbf{e}_j = \mathbf{e}_{k_j}$ 。这意味着线性变换  $T(\mathbf{x}) = P\mathbf{x}$  将标准基向量  $\mathbf{e}_1, \dots, \mathbf{e}_n$  重新排列 (置换)。对于任意向量  $\mathbf{x} = (x_1, \dots, x_n)^T$ ，计算  $P\mathbf{x}$  实际上就是将  $\mathbf{x}$  的分量按照特定的规则重新排列。
- b)  $P$  的列是标准基向量的一个排列。由于标准基是正交归一的，因此  $P$  的列也是正交归一的。这意味着  $P$  是正交矩阵 ( $Q$ )，满足  $P^T P = I$ 。因此  $P$  是可逆的，且  $P^{-1} = P^T$ 。  $P^T$  也是一个排列矩阵，对应于原排列的逆排列。
- c)  $n \times n$  排列矩阵的集合是一一对应于  $n$  个元素的排列集合  $S_n$  的。这个集合的大小是有限的 ( $n!$ )。考虑序列  $P, P^2, P^3, \dots$ 。由于集合有限，这个序列中必然存在两个相同的项，即存在  $k > j$  使得  $P^k = P^j$ 。因为  $P$  是可逆的，我们可以两边同乘  $(P^{-1})^j$  (即  $(P^T)^j$ )。得到  $P^{k-j} = I$ 。令  $N = k - j$ ，则  $N > 0$  且  $P^N = I$ 。
- 4.3. 解：考虑所有  $9 \times 9$  排列矩阵的集合。排列  $\sigma$  的奇偶性定义为对应排列矩阵  $P_\sigma$  的行列式 (1 为偶，-1 为奇)。考虑行列式函数  $\det : \{\text{排列矩阵}\} \rightarrow \{1, -1\}$ 。我们需要证明映射到 1 和 -1 的矩阵数量相等。取任意一个特定的奇排列矩阵  $T$  (例如，交换前两行，其余不变的矩阵， $\det T = -1$ )。对于任何偶排列矩阵  $P_{\text{even}}$ ，乘积  $TP_{\text{even}}$  是一个奇排列矩阵 (因为  $\det(TP_{\text{even}}) = (-1)(1) = -1$ )。这个映射  $P \mapsto TP$  是一个双射 (因为它可逆，逆映射为左乘  $T^{-1}$ )。因此，偶排列集合和奇排列集合之间存在一一对应关系。所以总数是偶数，且奇偶各占一半。
- 4.4. 解：利用符号的乘法性质： $\text{sign}(\tau\rho) = \text{sign}(\tau)\text{sign}(\rho)$ 。如果  $\sigma$  是奇排列，则  $\text{sign } \sigma = -1$ 。对

于  $\sigma^2$ :

$$\text{sign}(\sigma^2) = \text{sign}(\sigma) \text{sign}(\sigma) = (-1)(-1) = 1$$

所以  $\sigma^2$  是偶排列。对于  $\sigma^{-1}$ :

$$\text{sign}(\sigma\sigma^{-1}) = \text{sign}(\text{id}) = 1$$

$$(-1) \cdot \text{sign}(\sigma^{-1}) = 1 \implies \text{sign}(\sigma^{-1}) = -1$$

所以  $\sigma^{-1}$  是奇排列。

4.5. 解: 公式为:

$$\det A = \sum_{\sigma \in \text{Perm}(n)} \text{sign}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

1. 求和项的数量: 排列的总数是  $n!$ 。

2. 乘法次数: 每一项是  $n$  个元素的乘积:  $a_{1,\sigma(1)} \times \cdots \times a_{n,\sigma(n)}$ . 这需要  $n-1$  次乘法。(注: 如果不考虑符号的乘法, 或者将符号视为简单的正负号翻转)。总乘法次数 =  $n!(n-1)$ 。

3. 加法次数: 我们需要将  $n!$  个数相加。这需要  $n!-1$  次加法。

总结: 需要  $n!(n-1)$  次乘法和  $n!-1$  次加法。

注: 随着  $n$  的增加, 这个计算量是阶乘级增长的, 非常不切实际。

5.1. 解: 1. 第一个矩阵 ( $3 \times 3$ ): 沿第一行展开:

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

计算  $2 \times 2$  子行列式:

$$-1(1(-3) - (-5)(6)) + 1(1(4) - 2(6)) = -1(-3+30) + 1(4-12) = -1(27) + (-8) = -27-8 = -35$$

2. 第二个矩阵 ( $4 \times 4$ ): 记该矩阵为  $A$ . 我们可以使用行运算来简化计算。观察到第 2, 3, 4 行的第一列元素可以通过第 1 行消去。  $R_2 \leftarrow R_2 + 5R_1$   $R_3 \leftarrow R_3 + 9R_1$   $R_4 \leftarrow R_4 + 4R_1$

$$\det A = \begin{vmatrix} 1 & -2 & 3 & -12 \\ 0 & 2 & 1 & -41 \\ 0 & 4 & 7 & -77 \\ 0 & 1 & -2 & -33 \end{vmatrix}$$

沿第一列展开, 转化为  $3 \times 3$  行列式:

$$\det A = 1 \cdot \begin{vmatrix} 2 & 1 & -41 \\ 4 & 7 & -77 \\ 1 & -2 & -33 \end{vmatrix}$$

交换第 1 行和第 3 行 (行列式变号), 使左上角为 1, 便于计算:

$$= - \begin{vmatrix} 1 & -2 & -33 \\ 4 & 7 & -77 \\ 2 & 1 & -41 \end{vmatrix}$$

执行行运算:  $R_2 \leftarrow R_2 - 4R_1$ ,  $R_3 \leftarrow R_3 - 2R_1$ :

$$= - \begin{vmatrix} 1 & -2 & -33 \\ 0 & 15 & 55 \\ 0 & 5 & 25 \end{vmatrix}$$

从第 2 行提取公因子 5, 从第 3 行提取公因子 5:

$$= -(5)(5) \begin{vmatrix} 1 & -2 & -33 \\ 0 & 3 & 11 \\ 0 & 1 & 5 \end{vmatrix} = -25(3(5) - 11(1)) = -25(15 - 11) = -25(4) = -100$$

5.2. 解: 1. 第一个矩阵: 第 3 列有两个零. 沿第 3 列展开:

$$\begin{aligned}\text{Det} &= 0 \cdot C_{13} + 5 \cdot C_{23} + 0 \cdot C_{33} = 5(-1)^{2+3} \begin{vmatrix} 1 & 2 \\ 1 & -3 \end{vmatrix} \\ &= -5(1(-3) - 2(1)) = -5(-3 - 2) = -5(-5) = 25\end{aligned}$$

2. 第二个矩阵: 第 2 行有两个零. 沿第 2 行展开:

$$\text{Det} = -2 \begin{vmatrix} -6 & -4 & 4 \\ -3 & 1 & 3 \\ 2 & -3 & -5 \end{vmatrix} + 1 \begin{vmatrix} 4 & -4 & 4 \\ 0 & 1 & 3 \\ -2 & -3 & -5 \end{vmatrix}$$

记左边行列式为  $D_1$ , 右边为  $D_2$ . 计算  $D_1$ :  $R_1 \leftarrow R_1 + 4R_2, R_3 \leftarrow R_3 + 3R_2$

$$D_1 = \begin{vmatrix} -18 & 0 & 16 \\ -3 & 1 & 3 \\ -7 & 0 & 4 \end{vmatrix} = 1 \cdot \begin{vmatrix} -18 & 16 \\ -7 & 4 \end{vmatrix} = -18(4) - 16(-7) = -72 + 112 = 40$$

计算  $D_2$  (沿第 1 列展开):

$$\begin{aligned}D_2 &= 4 \begin{vmatrix} 1 & 3 \\ -3 & -5 \end{vmatrix} - 0 + (-2) \begin{vmatrix} -4 & 4 \\ 1 & 3 \end{vmatrix} \\ &= 4(-5 - (-9)) - 2(-12 - 4) = 4(4) - 2(-16) = 16 + 32 = 48\end{aligned}$$

总行列式  $= -2(40) + 1(48) = -80 + 48 = -32$ .

5.3. 解: 矩阵  $A + tI$  的形式如下 (主对角线加上  $t$ ):

$$A + tI = \begin{pmatrix} t & 0 & 0 & \cdots & 0 & a_0 \\ -1 & t & 0 & \cdots & 0 & a_1 \\ 0 & -1 & t & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & t & a_{n-2} \\ 0 & 0 & 0 & \cdots & -1 & t + a_{n-1} \end{pmatrix}$$

沿第一行展开:

$$\det(A + tI) = t \cdot \det M_{1,1} + (-1)^{1+n} a_0 \det M_{1,n}$$

其中  $M_{1,1}$  是去除第一行第一列后的矩阵, 它具有与原矩阵相同的结构, 只是维数变为  $n-1$ , 且系数索引从  $a_1$  到  $a_{n-1}$ .  $M_{1,n}$  是去除第一行和最后一列的矩阵:

$$M_{1,n} = \begin{pmatrix} -1 & t & 0 & \cdots \\ 0 & -1 & t & \cdots \\ \vdots & & & \ddots \\ 0 & \cdots & 0 & -1 \end{pmatrix}$$

这是一个上三角矩阵(对角线为 -1, 上方有元素), 其行列式为对角线元素的乘积  $(-1)^{n-1}$ . 因此, 我们可以推测行列式形式为多项式  $P_n(t)$ . 令  $D_n(a_0, \dots, a_{n-1})$  表示该行列式.

$$D_n = tD_{n-1}(a_1, \dots, a_{n-1}) + (-1)^{n+1} a_0 (-1)^{n-1}$$

注意  $(-1)^{n+1}(-1)^{n-1} = (-1)^{2n} = 1$ .

$$D_n = tD_{n-1} + a_0$$

我们可以通过归纳法验证结论:

$$\det(A + tI) = t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0.$$

**基础步骤** ( $n=1$ ): 矩阵是  $(t + a_0)$ . 其行列式就是  $= t + a_0$ . 符合公式.

**归纳步骤**: 假设对  $n-1$  成立, 即  $D_{n-1}(a_1, \dots, a_{n-1}) = t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_1$ .

代入递推公式:

$$D_n = t(t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_1) + a_0 = t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0$$

结论: 行列式为多项式  $t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0$ .

5.4. 解: 公式为  $A^{-1} = \frac{1}{\det A} C^T$ , 其中  $C$  是代数余子式矩阵。

1. 矩阵  $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ :  $\det = 4 - 6 = -2$ . 代数余子式:  $C_{11} = 4, C_{12} = -3, C_{21} = -2, C_{22} =$

$$1. C^T = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}. A^{-1} = \frac{1}{-2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}.$$

2. 矩阵  $\begin{pmatrix} 19 & -17 \\ 3 & -2 \end{pmatrix}$ :  $\det = 19(-2) - (-17)(3) = -38 + 51 = 13$ .  $C^T = \begin{pmatrix} -2 & 17 \\ -3 & 19 \end{pmatrix}$  (交换对角元, 其余变号).  $A^{-1} = \frac{1}{13} \begin{pmatrix} -2 & 17 \\ -3 & 19 \end{pmatrix}$ .

3. 矩阵  $\begin{pmatrix} 1 & 0 \\ 3 & 5 \end{pmatrix}$ :  $\det = 5$ .  $C^T = \begin{pmatrix} 5 & 0 \\ -3 & 1 \end{pmatrix}$ .  $A^{-1} = \frac{1}{5} \begin{pmatrix} 5 & 0 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -0.6 & 0.2 \end{pmatrix}$ .

4. 矩阵  $A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$ : 计算行列式:  $1(1-2) - 1(2-0) + 0 = -1 - 2 = -3$ . 计算代数

余子式  $C_{ij}$ :  $C_{11} = +(1-2) = -1$ ,  $C_{12} = -(2-0) = -2$ ,  $C_{13} = +(2-0) = 2$ ,  $C_{21} = -(1-0) = -1$ ,  $C_{22} = +(1-0) = 1$ ,  $C_{23} = -(1-0) = -1$ ,  $C_{31} = +(2-0) = 2$ ,  $C_{32} =$

$-(2-0) = -2$ ,  $C_{33} = +(1-2) = -1$  代数余子式矩阵  $C = \begin{pmatrix} -1 & -2 & 2 \\ -1 & 1 & -1 \\ 2 & -2 & -1 \end{pmatrix}$ . 转置得到伴随矩阵  $C^T = \begin{pmatrix} -1 & -1 & 2 \\ -2 & 1 & -2 \\ 2 & -1 & -1 \end{pmatrix}$ . 逆矩阵  $A^{-1} = -\frac{1}{3} \begin{pmatrix} -1 & -1 & 2 \\ -2 & 1 & -2 \\ 2 & -1 & -1 \end{pmatrix}$ .

5.5. 证明: 沿第一行展开行列式:

$$D_n = 1 \cdot \det M_{1,1} - (-1) \cdot \det M_{1,2}$$

其中  $M_{1,1}$  是去掉第一行第一列后的子矩阵, 它显然是同类型的  $(n-1) \times (n-1)$  矩阵, 其行列式为  $D_{n-1}$ . 矩阵  $M_{1,2}$  去掉了第一行第二列, 形式为:

$$M_{1,2} = \begin{pmatrix} 1 & -1 & 0 & \cdots \\ 0 & 1 & -1 & \cdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \end{pmatrix}$$

注意第一列只有第一个元素为 1, 其余为 0. 沿这一列展开  $\det M_{1,2} = 1 \cdot \det(\text{其余部分})$ . 剩下的部分恰好是同类型的  $(n-2) \times (n-2)$  矩阵, 其行列式为  $D_{n-2}$ . 因此:

$$D_n = 1 \cdot D_{n-1} - (-1) \cdot D_{n-2} = D_{n-1} + D_{n-2}$$

计算前几项:  $D_1 = |1| = 1$ .  $D_2 = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 1 - (-1) = 2$ .  $D_3 = D_2 + D_1 = 3$ . 数列为 1, 2, 3, 5, ..., 确实是斐波那契数列。

5.6. 解: a) 当  $n = 1$  时 (矩阵大小  $2 \times 2$ ):

$$\det = \begin{vmatrix} 1 & c_0 \\ 1 & c_1 \end{vmatrix} = c_1 - c_0$$

公式  $\prod_{0 \leq j < k \leq 1} (c_k - c_j) = c_1 - c_0$ . 成立。当  $n = 2$  时 (矩阵大小  $3 \times 3$ ):

$$\begin{vmatrix} 1 & c_0 & c_0^2 \\ 1 & c_1 & c_1^2 \\ 1 & c_2 & c_2^2 \end{vmatrix}$$

根据习题 3.8 的结果, 这等于  $(c_2 - c_1)(c_2 - c_0)(c_1 - c_0)$ , 符合公式。

b) 设  $V_{n+1}$  为该行式, 将最后一行  $(1, c_n, c_n^2, \dots, c_n^n)$  替换为  $(1, x, x^2, \dots, x^n)$ . 沿最后一

行展开:

$$\det = 1 \cdot C_{n+1,1} + x \cdot C_{n+1,2} + \cdots + x^n \cdot C_{n+1,n+1}$$

由于  $C_{n+1,k}$  仅由前  $n$  行决定 (即只包含  $c_0, \dots, c_{n-1}$ ), 对于  $x$  来说是常数。显然这是一个关于  $x$  的多项式, 最高次项为  $x^n$ , 系数  $A_n = C_{n+1,n+1}$ 。

c) 如果令  $x = c_i$  (其中  $0 \leq i \leq n-1$ ), 那么最后一行与第  $i+1$  行完全相同。含有两行相同的行列式为 0。因此  $c_0, c_1, \dots, c_{n-1}$  都是该多项式的根。根据代数基本定理, 我们可以将多项式写为:

$$P(x) = A_n(x - c_0)(x - c_1) \cdots (x - c_{n-1})$$

d) 系数  $A_n$  是  $C_{n+1,n+1}$ , 即去掉最后一行和最后一列的子行列式。这恰好是基于  $c_0, \dots, c_{n-1}$  的  $n \times n$  范德蒙德行列式 (对应题目中的  $n-1$  情况)。根据归纳假设:

$$A_n = \prod_{0 \leq j < k \leq n-1} (c_k - c_j)$$

代回 c) 中的表达式, 并令  $x = c_n$ :

$$\det V_{n+1} = \left( \prod_{0 \leq j < k \leq n-1} (c_k - c_j) \right) (c_n - c_0)(c_n - c_1) \cdots (c_n - c_{n-1})$$

这正是将  $k = n$  的项乘进去, 从而得证公式对  $n$  也成立。

- 5.7. 解: 令  $M_n$  为计算  $n \times n$  行列式所需的乘法次数。公式为  $\det A = \sum_{k=1}^n a_{1k}(-1)^{1+k} \det A_{1k}$ 。这一步需要: 1. 计算  $n$  个  $(n-1) \times (n-1)$  子行列式, 即  $nM_{n-1}$  次乘法。2. 将每个子行列式与  $a_{1k}$  相乘 (忽略符号的乘法, 或将其视为 1 次), 共  $n$  次乘法。递推关系:

$$M_n = nM_{n-1} + n$$

且  $M_1 = 0$  (直接读取数值, 无需乘法)。  $M_2 = 2(0) + 2 = 2$ 。  $M_3 = 3(2) + 3 = 9$ 。  $M_4 = 4(9) + 4 = 40$ 。我们可以将递推式除以  $n!$ :

$$\frac{M_n}{n!} = \frac{M_{n-1}}{(n-1)!} + \frac{1}{(n-1)!}$$

展开求和:

$$\frac{M_n}{n!} = \frac{M_1}{1!} + \sum_{k=1}^{n-1} \frac{1}{k!} = \sum_{k=1}^{n-1} \frac{1}{k!}$$

所以:

$$M_n = n! \sum_{k=1}^{n-1} \frac{1}{k!}$$

当  $n$  很大时, 级数  $\sum_{k=1}^{\infty} \frac{1}{k!} = e - 1$ 。所以  $M_n \approx (e - 1)n!$ 。这证明了计算复杂度是阶乘级的。

- 7.1. 解: a) 正确。行列式仅针对方阵 ( $n \times n$ ) 定义。

b) 正确。如果两行 (列) 相同, 矩阵不可逆, 或者通过交换这两行行列式变号且保持不变, 这意味着  $\det A = -\det A \implies \det A = 0$ 。

c) 错误。交换两行 (列) 会改变行列式的符号。应该是  $\det B = -\det A$ 。

d) 错误。如果将某一行乘以  $\alpha$ , 行列式变为  $\alpha \det A$ 。只有当  $\alpha = 1$  时才相等。

e) 正确。这是行运算的性质 (剪切操作), 不改变行列式的值。

f) 正确。三角矩阵的特征值就是对角线元素, 行列式是特征值的乘积。

g) 错误。  $\det(A^T) = \det(A)$ 。

h) 正确。这是行列式的乘法性质。

i) 正确。这是可逆性的主要判别准则。

j) 正确。因为  $1 = \det(I) = \det(AA^{-1}) = \det(A) \det(A^{-1})$ , 所以  $\det(A^{-1}) = 1/\det(A)$ 。

7.2. 解: 利用行列式的多线性性质 (每一行提取一个标量):

1.  $\det(3A) = 3^n \det A$  (因为  $A$  有  $n$  行, 每行都乘以了 3)。
2.  $\det(-A) = \det((-1)A) = (-1)^n \det A$ 。
3.  $\det(A^2) = \det(AA) = \det(A) \det(A) = (\det A)^2$ 。

7.3. 解: **不可能**。原因如下: 行列式的定义涉及矩阵元素的加法和乘法。如果  $A$  的所有元素都是整数, 那么  $\det A$  必须是一个整数。同理, 如果  $A^{-1}$  的所有元素都是整数, 那么  $\det(A^{-1})$  也必须是一个整数。我们要用到性质:

$$\det(A) \det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1$$

我们需要两个整数  $x = \det A$  和  $y = \det(A^{-1})$  满足  $xy = 1$ 。整数环中只有两个可逆元: 1 和 -1。因此,  $\det A$  只能是 1 或 -1。它不可能是 3。

7.4. 证明: 设  $A = [\mathbf{v}_1, \mathbf{v}_2]$ 。

**情况 1:**  $\mathbf{v}_1$  位于  $x$  轴上, 即  $\mathbf{v}_1 = (x_1, 0)^T$ 。此时  $\mathbf{v}_2 = (x_2, y_2)^T$ 。矩阵为  $A = \begin{pmatrix} x_1 & x_2 \\ 0 & y_2 \end{pmatrix}$ 。这是一个上三角矩阵,  $\det A = x_1 y_2$ 。由  $\mathbf{v}_1, \mathbf{v}_2$  构成的平行四边形的底边长度为  $\|\mathbf{v}_1\| = |x_1|$ , 对应的高是  $\mathbf{v}_2$  的  $y$  分量的绝对值  $|y_2|$ 。面积 = 底  $\times$  高 =  $|x_1| |y_2| = |x_1 y_2| = |\det A|$ 。

**情况 2:** 一般情况。存在一个旋转矩阵  $Q$  (旋转角度  $-\theta$ ), 使得  $Q\mathbf{v}_1$  落在  $x$  轴上 (即  $Q\mathbf{v}_1 = (\tilde{x}_1, 0)^T$ )。旋转矩阵是正交矩阵, 且保持定向, 所以  $\det Q = 1$ 。旋转是一种刚体变换, 不改变几何形状的面积。因此, 由  $\mathbf{v}_1, \mathbf{v}_2$  生成的平行四边形面积等于由  $Q\mathbf{v}_1, Q\mathbf{v}_2$  生成的面积。令  $B = QA = [Q\mathbf{v}_1, Q\mathbf{v}_2]$ 。根据情况 1, 面积 =  $|\det B|$ 。又因为  $\det B = \det(QA) = \det(Q) \det(A) = 1 \cdot \det A = \det A$ 。所以, 面积 =  $|\det A|$ 。

7.5. 证明: 设  $A = [\mathbf{v}_1, \mathbf{v}_2]$ , 则  $D(\mathbf{v}_1, \mathbf{v}_2) = \det A$ 。选取旋转矩阵  $T_\alpha$  将  $\mathbf{v}_1$  旋转到正  $x$  轴方向。即  $T_\alpha \mathbf{v}_1 = (r, 0)^T$ , 其中  $r = \|\mathbf{v}_1\| > 0$  (假设  $\mathbf{v}_1 \neq \mathbf{0}$ )。设变换后的第二个向量为  $T_\alpha \mathbf{v}_2 = (x', y')^T$ 。此时变换后的矩阵为  $A' = [T_\alpha \mathbf{v}_1, T_\alpha \mathbf{v}_2] = \begin{pmatrix} r & x' \\ 0 & y' \end{pmatrix}$ 。我们知道  $\det T_\alpha = 1$ 。

$$\det A' = \det(T_\alpha A) = \det(T_\alpha) \det(A) = \det A$$

计算  $\det A'$ :

$$\det A' = ry' - 0 = ry'$$

因为  $r = \|\mathbf{v}_1\| > 0$ , 所以  $\det A > 0$  当且仅当  $y' > 0$ 。 $y' > 0$  意味着向量  $T_\alpha \mathbf{v}_2$  的第二个分量为正, 即该向量位于上半平面。得证。

## 第四章习题解答

1.1. 解: **a) 错误**。特征值可以重复 (例如单位矩阵  $I$ , 所有特征值都是 1)。

**b) 正确**。如果  $\mathbf{v}$  是特征向量, 则对于任何非零标量  $c$ ,  $c\mathbf{v}$  也是特征向量。

**c) 正确**。例如旋转  $90^\circ$  的矩阵  $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ , 其特征值为  $\pm i$ 。

**d) 错误**。根据代数基本定理, 复数域上的特征多项式总是有根, 因此总存在特征值和特征向量。

**e) 正确**。相似矩阵具有相同的特征多项式, 因此特征值相同。

**f) 错误**。如果  $A = SBS^{-1}$ , 且  $B\mathbf{v} = \lambda\mathbf{v}$ , 则  $A(S\mathbf{v}) = \lambda(S\mathbf{v})$ 。特征向量由  $S$  变换, 通常不相同。

**g) 错误**。只有当这两个特征向量对应于同一个特征值时, 它们的和才是特征向量。如果对应不同特征值, 和不是特征向量。

**h) 正确**。如果  $A\mathbf{v}_1 = \lambda\mathbf{v}_1$  且  $A\mathbf{v}_2 = \lambda\mathbf{v}_2$ , 则  $A(\mathbf{v}_1 + \mathbf{v}_2) = \lambda(\mathbf{v}_1 + \mathbf{v}_2)$ 。

1.2. 解: 1. 第一个矩阵  $A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$ : 特征多项式:

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{vmatrix} = (4 - \lambda)(-3 - \lambda) + 10 = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1)$$

特征值:  $\lambda_1 = 2, \lambda_2 = -1$ . 特征向量: 对于  $\lambda_1 = 2$ : 求解  $(A - 2I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} 2 & -5 \\ 2 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies 2x - 5y = 0$ . 取  $\mathbf{v}_1 = (5, 2)^T$ . 对于  $\lambda_2 = -1$ : 求解  $(A + I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies x - y = 0$ . 取  $\mathbf{v}_2 = (1, 1)^T$ .

2. 第二个矩阵  $B = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$ : 特征多项式:

$$\det(B - \lambda I) = \begin{vmatrix} 2 - \lambda & 1 \\ -1 & 4 - \lambda \end{vmatrix} = \lambda^2 - 6\lambda + 8 + 1 = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2$$

特征值:  $\lambda = 3$  (代数重数 2). 特征向量: 求解  $(B - 3I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies x = y$ . 取  $\mathbf{v} = (1, 1)^T$ . (几何重数为 1).

3. 第三个矩阵  $C = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}$ : 特征多项式:

$$\det(C - \lambda I) = \begin{vmatrix} 1 - \lambda & 3 & 3 \\ -3 & -5 - \lambda & -3 \\ 3 & 3 & 1 - \lambda \end{vmatrix}$$

利用  $R_3 + R_2$ :

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

提取  $(-2 - \lambda)$  从  $R_3$ , 然后展开, 最终得到  $\det = -(\lambda - 1)(\lambda + 2)^2$ . 特征值:  $\lambda_1 = 1, \lambda_2 = -2$  (代数重数 2). 特征向量:

对于  $\lambda_1 = 1$ :  $(C - I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0}$ . 解得  $\mathbf{v}_1 = (1, -1, 1)^T$ .

对于  $\lambda_2 = -2$ :  $(C + 2I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x + y + z = 0$ .

这是一个平面, 基向量可选为  $\mathbf{v}_2 = (1, -1, 0)^T$  和  $\mathbf{v}_3 = (1, 0, -1)^T$ .

1.3. 解: 特征多项式:

$$\begin{vmatrix} \cos \alpha - \lambda & -\sin \alpha \\ \sin \alpha & \cos \alpha - \lambda \end{vmatrix} = (\cos \alpha - \lambda)^2 + \sin^2 \alpha = \lambda^2 - 2\lambda \cos \alpha + (\cos^2 \alpha + \sin^2 \alpha) = \lambda^2 - 2\lambda \cos \alpha + 1 = 0$$

使用求根公式:

$$\lambda = \frac{2 \cos \alpha \pm \sqrt{4 \cos^2 \alpha - 4}}{2} = \cos \alpha \pm \sqrt{-\sin^2 \alpha} = \cos \alpha \pm i \sin \alpha = e^{\pm i \alpha}$$

特征向量: 对于  $\lambda_1 = \cos \alpha + i \sin \alpha$ :

$$\begin{pmatrix} -i \sin \alpha & -\sin \alpha \\ \sin \alpha & -i \sin \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies -ix - y = 0 \implies y = -ix$$

取  $\mathbf{v}_1 = (1, -i)^T$ . 对于  $\lambda_2 = \cos \alpha - i \sin \alpha$ : 同理可得  $\mathbf{v}_2 = (1, i)^T$ .

1.4. 解: 这些都是三角矩阵 (上三角或下三角). 特征值即为对角线元素.

1.  $p(\lambda) = (1 - \lambda)(2 - \lambda)(-2 - \lambda)(3 - \lambda)$ . 特征值:  $1, 2, -2, 3$ .

2.  $p(\lambda) = (2 - \lambda)(\pi - \lambda)(16 - \lambda)(54 - \lambda)$ . 特征值:  $2, \pi, 16, 54$ .  
 3. 下三角矩阵.  $p(\lambda) = (4 - \lambda)(3 - \lambda)(e - \lambda)(1 - \lambda)$ . 特征值:  $4, 3, e, 1$ .  
 4. 下三角矩阵.  $p(\lambda) = (4 - \lambda)(0 - \lambda)(0 - \lambda)(1 - \lambda) = \lambda^2(\lambda - 4)(\lambda - 1)$ . 特征值:  $4, 0, 0, 1$ .  
 1.5. 证明: 设  $A$  为三角矩阵 (不妨设为上三角). 矩阵  $A - \lambda I$  也是上三角矩阵, 其对角线元素为  $a_{kk} - \lambda$ . 三角矩阵的行列式等于其对角线元素的乘积. 因此, 特征多项式为:

$$\det(A - \lambda I) = \prod_{k=1}^n (a_{kk} - \lambda)$$

该多项式的根正是  $a_{11}, a_{22}, \dots, a_{nn}$ . 证毕。

- 1.6. 证明: 设  $\lambda$  是  $A$  的一个特征值,  $\mathbf{v}$  是对应的非零特征向量. 则  $A\mathbf{v} = \lambda\mathbf{v}$ . 反复应用  $A$ :

$$A^2\mathbf{v} = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda^2\mathbf{v}$$

归纳可得  $A^k\mathbf{v} = \lambda^k\mathbf{v}$ . 因为  $A$  是幂零的, 存在  $k$  使得  $A^k = \mathbf{0}$ . 所以:

$$\mathbf{0} = A^k\mathbf{v} = \lambda^k\mathbf{v}$$

由于  $\mathbf{v} \neq \mathbf{0}$ , 必须有  $\lambda^k = 0$ , 这蕴含  $\lambda = 0$ . 因此  $0$  是唯一的特征值。

- 1.7. 证明: 设  $M = \begin{pmatrix} A & C \\ \mathbf{0} & B \end{pmatrix}$ . 则  $M - \lambda I = \begin{pmatrix} A - \lambda I_A & C \\ \mathbf{0} & B - \lambda I_B \end{pmatrix}$ . 根据分块矩阵行列式的性质 (当左下角块为零时, 行列式等于对角块行列式的乘积):

$$\det(M - \lambda I) = \det(A - \lambda I_A) \cdot \det(B - \lambda I_B)$$

即  $M$  的特征多项式是  $A$  和  $B$  特征多项式的乘积。

- 1.8. 证明: 算子  $A$  在基  $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  下的矩阵表示的第  $j$  列是向量  $A\mathbf{v}_j$  在该基下的坐标. 对于  $j \leq k$ :

$$A\mathbf{v}_j = \lambda\mathbf{v}_j = 0\mathbf{v}_1 + \dots + \lambda\mathbf{v}_j + \dots + 0\mathbf{v}_n$$

因此, 矩阵的前  $k$  列中, 第  $j$  列只有第  $j$  个位置是  $\lambda$ , 其余为  $0$ . 这意味着矩阵的左上角  $k \times k$  块是对角矩阵  $\lambda I_k$ , 而左下角  $(n - k) \times k$  块全是  $0$ . 矩阵形式为  $\begin{pmatrix} \lambda I_k & C \\ \mathbf{0} & B \end{pmatrix}$ .

- 1.9. 证明: 设特征值  $\lambda_0$  的几何重数为  $k$ . 这意味着对应的特征空间  $\dim \text{Ker}(A - \lambda_0 I) = k$ . 选取该特征空间的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_k$ , 并将其扩充为全空间  $V$  的一组基. 根据练习 1.8, 在该基下矩阵  $A$  具有形式  $\begin{pmatrix} \lambda_0 I_k & * \\ \mathbf{0} & B \end{pmatrix}$ . 根据练习 1.7, 其特征多项式为:

$$p(z) = \det(\lambda_0 I_k - z I_k) \det(B - z I_{n-k}) = (\lambda_0 - z)^k \det(B - z I_{n-k})$$

这表明  $(\lambda_0 - z)^k$  是特征多项式的一个因式. 因此,  $\lambda_0$  作为特征多项式根的重数 (代数重数) 至少为  $k$ . 即: 几何重数  $\leq$  代数重数.

- 1.10. 证明: 设特征值为  $\lambda_1, \dots, \lambda_n$ . 特征多项式可以分解为:

$$p(\lambda) = \det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

在上式中令  $\lambda = 0$ :

$$p(0) = \det(A - 0I) = \det A$$

另一方面, 代入右边:

$$(\lambda_1 - 0)(\lambda_2 - 0) \dots (\lambda_n - 0) = \lambda_1 \lambda_2 \dots \lambda_n$$

因此  $\det A = \prod_{i=1}^n \lambda_i$ .

- 1.11. 证明: **步骤 1:** 考虑右侧多项式  $P(\lambda) = \prod_{i=1}^n (\lambda_i - \lambda)$ . 该多项式是  $n$  次的.  $\lambda^n$  的系数是  $(-1)^n$ .  $\lambda^{n-1}$  项是通过在  $n$  个因子中, 选择  $n - 1$  个  $(-\lambda)$  和 1 个常数项  $\lambda_k$  相乘得到的. 对所有可能的选择求和, 系数为:

$$\sum_{k=1}^n \lambda_k (-1)^{n-1} = (-1)^{n-1} \sum_{i=1}^n \lambda_i$$



**步骤 2:** 考虑行列式的定义

$$\det(A - \lambda I) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n (A_{i, \sigma(i)} - \delta_{i, \sigma(i)} \lambda)$$

当  $\sigma$  是恒等排列时, 该项为  $\prod_{i=1}^n (a_{ii} - \lambda)$ . 这是一个  $n$  次多项式。当  $\sigma$  不是恒等排列时, 它至少改变了两个元素的索引 (即至少有两个  $i$  使得  $\sigma(i) \neq i$ )。这意味着乘积中至少有两个因子取的是非对角元  $a_{i, \sigma(i)}$  (这些项不含  $\lambda$ )。因此, 剩下的含有  $\lambda$  的对角元因子最多只有  $n - 2$  个。所以, 对于  $\sigma \neq \text{id}$ , 对应的多项式次数最多为  $n - 2$ 。因此, 特征多项式可以写为:

$$\det(A - \lambda I) = \prod_{i=1}^n (a_{ii} - \lambda) + q(\lambda), \quad \deg(q) \leq n - 2$$

**步骤 3:** 比较  $\lambda^{n-1}$  的系数。在  $\prod_{i=1}^n (a_{ii} - \lambda)$  中,  $\lambda^{n-1}$  的系数同样由选择  $n - 1$  个  $(-\lambda)$  和 1 个  $a_{kk}$  得到:

$$\text{Coef} = (-1)^{n-1} \sum_{i=1}^n a_{ii} = (-1)^{n-1} \operatorname{trace}(A)$$

因为  $q(\lambda)$  不含  $\lambda^{n-1}$  项, 所以特征多项式中  $\lambda^{n-1}$  的系数就是  $(-1)^{n-1} \operatorname{trace}(A)$ 。结合步骤 1 的结果:

$$(-1)^{n-1} \sum_{i=1}^n \lambda_i = (-1)^{n-1} \operatorname{trace}(A)$$

消去  $(-1)^{n-1}$ , 得  $\operatorname{trace}(A) = \sum_{i=1}^n \lambda_i$ 。

2.1. 解: **a) 正确。** 特征多项式取决于行列式, 而转置不改变行列式的值:

$$\det(A^T - \lambda I) = \det((A - \lambda I)^T) = \det(A - \lambda I)$$

因为特征多项式相同, 所以特征值相同。

**b) 错误。** 例如, 设  $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ .  $A$  的特征向量满足  $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies y = 0$ , 即  $\mathbf{v} = (1, 0)^T$ .  $A^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$  的特征向量满足  $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies x = 0$ , 即  $\mathbf{w} = (0, 1)^T$ . 两者不同。

**c) 正确。** 如果  $A$  可对角化, 则存在可逆矩阵  $S$  和对角矩阵  $D$  使得  $A = SDS^{-1}$ . 取转置:

$$A^T = (SDS^{-1})^T = (S^{-1})^T D^T S^T = (S^T)^{-1} D S^T$$

因为  $D$  是对角矩阵, 所以  $D^T = D$ . 令  $P = (S^T)^{-1}$ , 则  $A^T = P^{-1} D P$  (或者更标准的  $Q D Q^{-1}$  形式, 取  $Q = (S^T)^{-1}$ )。这意味着  $A^T$  相似于同一个对角矩阵  $D$ , 因此是可对角化的。

2.2. 证明: 我们有方程  $A\mathbf{v} = \lambda\mathbf{v}$ . 对等式两边取复共轭:

$$\overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}}$$

由共轭的性质 (积的共轭等于共轭的积), 得:

$$\overline{A}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$$

因为  $A$  是实矩阵, 所以  $\overline{A} = A$ . 因此:

$$A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$$

这正是特征值和特征向量的定义。所以  $\overline{\lambda}$  是特征值,  $\overline{\mathbf{v}}$  是对应的特征向量。

## 2.3. 解: 步骤 1: 求特征值

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 3 \\ 1 & 2 - \lambda \end{vmatrix} = \lambda^2 - 6\lambda + 8 - 3 = \lambda^2 - 6\lambda + 5 = (\lambda - 5)(\lambda - 1)$$

特征值为  $\lambda_1 = 5, \lambda_2 = 1$ .

步骤 2: 求特征向量对于  $\lambda_1 = 5$ :

$$\begin{pmatrix} -1 & 3 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies x - 3y = 0$$

取  $\mathbf{v}_1 = (3, 1)^T$ . 对于  $\lambda_2 = 1$ :

$$\begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \implies x + y = 0$$

取  $\mathbf{v}_2 = (1, -1)^T$ .

步骤 3: 对角化令  $S = (\mathbf{v}_1, \mathbf{v}_2) = \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix}$ , 则  $D = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$ . 计算  $S^{-1}$ :

$$\det S = -3 - 1 = -4, \quad S^{-1} = \frac{1}{-4} \begin{pmatrix} -1 & -1 \\ -1 & 3 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & -3 \end{pmatrix}$$

步骤 4: 计算  $A^{2004}$

$$\begin{aligned} A^{2004} &= S D^{2004} S^{-1} = \frac{1}{4} \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5^{2004} & 0 \\ 0 & 1^{2004} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -3 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 3 \cdot 5^{2004} & 1 \\ 5^{2004} & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -3 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 3 \cdot 5^{2004} + 1 & 3 \cdot 5^{2004} - 3 \\ 5^{2004} - 1 & 5^{2004} + 3 \end{pmatrix} \end{aligned}$$

2.4. 解: 设  $\lambda_1 = 1, \mathbf{v}_1 = (1, 2)^T$ ;  $\lambda_2 = 3, \mathbf{v}_2 = (1, 1)^T$ . 令  $S = [\mathbf{v}_1, \mathbf{v}_2] = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$ . 令

$D = \text{diag}(1, 3) = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ . 我们要求  $A = S D S^{-1}$ .

$$\det S = 1 - 2 = -1, \quad S^{-1} = \frac{1}{-1} \begin{pmatrix} 1 & -1 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}$$

计算  $A$ :

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

这样的矩阵是唯一的, 因为线性变换由其基上的作用唯一确定 (这里  $\mathbf{v}_1, \mathbf{v}_2$  构成了  $\mathbb{R}^2$  的一组基).

## 2.5. 解: a)

$$\det(A - \lambda I) = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3)$$

$$\lambda_1 = 2 \implies \mathbf{v}_1 = (1, 1)^T. \quad \lambda_2 = 3 \implies \mathbf{v}_2 = (2, 1)^T.$$

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}^{-1}$$

b)

$$\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2)$$

$$\lambda_1 = 1 \implies \mathbf{v}_1 = (1, -2)^T. \quad \lambda_2 = 2 \implies \mathbf{v}_2 = (1, -3)^T.$$

$$A = \begin{pmatrix} 1 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & -3 \end{pmatrix}^{-1}$$

c) 已知  $\lambda_1 = 2$ . 计算特征多项式 (可利用迹  $\text{tr}(A) = 8$  和行列式或直接行变换):

$$\det(A - \lambda I) = -(\lambda - 2)(\lambda - 3)^2$$

特征值为 2, 3, 3. 检查  $\lambda = 3$  的几何重数:

$$A - 3I = \begin{pmatrix} -5 & 2 & 6 \\ 5 & -2 & -6 \\ -5 & 2 & 6 \end{pmatrix}$$

秩为 1 (所有行成比例), 所以零空间维数为  $3 - 1 = 2$ . 几何重数等于代数重数, 可对角化。

$\lambda = 2$ :  $(A - 2I)\mathbf{v} = \mathbf{0} \implies \mathbf{v}_1 = (1, -1, 1)^T$  (验证:  $-2 - 2 + 6 = 2, 5 - 1 - 6 = -2, -5 - 2 + 9 = 2$ , 符合  $2\mathbf{v}$ ).

$\lambda = 3$ :  $5x - 2y - 6z = 0$ . 基向量可取  $\mathbf{v}_2 = (2, 5, 0)^T, \mathbf{v}_3 = (6, 0, 5)^T$  (或  $(0, 3, -1)^T$  等).

$$A = S \text{diag}(2, 3, 3) S^{-1}$$

2.6. 解: a) 特征值是 **2, 5, 4**. 因为这是一个上三角矩阵, 特征值即对角线元素。

b) 可以. 因为  $A$  有 3 个互不相同的特征值, 所以它必定有 3 个线性无关的特征向量, 构成一组基。

$$\text{c) } \lambda_1 = 2: (A - 2I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} 0 & 6 & -6 \\ 0 & 3 & -2 \\ 0 & 0 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies y = z = 0. \mathbf{v}_1 = (1, 0, 0)^T.$$

$$\lambda_2 = 5: (A - 5I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} -3 & 6 & -6 \\ 0 & 0 & -2 \\ 0 & 0 & -1 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies z = 0, -3x + 6y = 0.$$

$$\mathbf{v}_2 = (2, 1, 0)^T. \lambda_3 = 4: (A - 4I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} -2 & 6 & -6 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0}. \text{ 设 } z = 1 \implies y = 2 \implies -2x + 12 - 6 = 0 \implies x = 3. \mathbf{v}_3 = (3, 2, 1)^T.$$

$$S = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

2.7. 解: 特征值 (对角元):  $\lambda = 2$  (代数重数 2),  $\lambda = 4$  (代数重数 1). 检查  $\lambda = 2$  的几何重数:

$$A - 2I = \begin{pmatrix} 0 & 0 & 6 \\ 0 & 0 & 4 \\ 0 & 0 & 2 \end{pmatrix}$$

该矩阵显然秩为 1 (只有第 3 列非零). 因此  $\dim \text{Ker}(A - 2I) = 3 - 1 = 2$ . 几何重数 = 代数重数, 矩阵可对角化.  $\lambda = 2$  的特征向量基:  $\mathbf{e}_1 = (1, 0, 0)^T, \mathbf{e}_2 = (0, 1, 0)^T$ .  $\lambda = 4$  的特征

$$\text{向量: } (A - 4I)\mathbf{x} = \mathbf{0} \implies \begin{pmatrix} -2 & 0 & 6 \\ 0 & -2 & 4 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x = 3z, y = 2z. \text{ 取 } \mathbf{v}_3 = (3, 2, 1)^T.$$

$$S = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

2.8. 解: 首先对角化  $A$ .  $\det(A - \lambda I) = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3)$ .  $\lambda_1 = 2, \mathbf{v}_1 = (2, -3)^T$  (由  $3x + 2y = 0$ ).  $\lambda_2 = 3, \mathbf{v}_2 = (1, -1)^T$  (由  $2x + 2y = 0$ ).  $S = \begin{pmatrix} 2 & 1 \\ -3 & -1 \end{pmatrix}, D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ .

$A = SDS^{-1}$ . 若  $B^2 = A$ , 则  $B = S\sqrt{D}S^{-1}$ , 其中  $\sqrt{D}$  是任意平方等于  $D$  的矩阵. 对于对角矩阵, 平方根可以是  $\text{diag}(\pm\sqrt{2}, \pm\sqrt{3})$ . 共有 4 个解:

$$B = \begin{pmatrix} 2 & 1 \\ -3 & -1 \end{pmatrix} \begin{pmatrix} \pm\sqrt{2} & 0 \\ 0 & \pm\sqrt{3} \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 3 & 2 \end{pmatrix}$$

(注:  $S^{-1} = \frac{1}{1} \begin{pmatrix} -1 & -1 \\ 3 & 2 \end{pmatrix}$ ).

2.9. 解: **a)** 由  $\phi_{n+2} = 1 \cdot \phi_{n+1} + 1 \cdot \phi_n$  和  $\phi_{n+1} = 1 \cdot \phi_{n+1} + 0 \cdot \phi_n$ , 得到  $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ .

**b)** 特征方程  $\lambda^2 - \lambda - 1 = 0$ . 特征值  $\lambda_1 = \frac{1+\sqrt{5}}{2} = \varphi$  (黄金分割比),  $\lambda_2 = \frac{1-\sqrt{5}}{2} = \psi$ . 特征向量满足  $\lambda x - y = 0 \implies y = \lambda x$ . 取  $\mathbf{v}_1 = (\lambda_1, 1)^T$ ,  $\mathbf{v}_2 = (\lambda_2, 1)^T$ .  $S = \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix}$ .  $\det S = \lambda_1 - \lambda_2 = \sqrt{5}$ .  $S^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{pmatrix}$ .  $A^n = S \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} S^{-1}$ .

**c)**

$$\begin{pmatrix} \phi_{n+1} \\ \phi_n \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \begin{pmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

只关注右边向量的乘积部分:

$$S^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$D^n \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^n \\ -\lambda_2^n \end{pmatrix}$$

最后左乘  $S$ :

$$\begin{pmatrix} \phi_{n+1} \\ \phi_n \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^n \\ -\lambda_2^n \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \dots \\ \lambda_1^n - \lambda_2^n \end{pmatrix}$$

提取第二行:

$$\phi_n = \frac{1}{\sqrt{5}}(\lambda_1^n - \lambda_2^n) = \frac{1}{\sqrt{5}} \left[ \left( \frac{1+\sqrt{5}}{2} \right)^n - \left( \frac{1-\sqrt{5}}{2} \right)^n \right]$$

**d)**

$$\frac{\phi_{n+1}}{\phi_n} \approx \frac{c_1 \lambda_1^{n+1}}{c_1 \lambda_1^n} = \lambda_1$$

因为  $|\lambda_1| > 1 > |\lambda_2|$ , 随着  $n \rightarrow \infty$ ,  $\lambda_1^n$  项占主导地位. 向量  $(\frac{\phi_{n+1}}{\phi_n}, 1)^T \rightarrow (\lambda_1, 1)^T$ , 这正是对应于主特征值  $\lambda_1$  的特征向量. 这不是巧合, 这是幂法 (Power Iteration) 的一个实例.

2.10. 解: **是的,  $A$  必定可对角化.** 理由如下: 设这三个互不相同的特征值为  $\lambda_1, \lambda_2, \lambda_3$ . 设它们的代数重数为  $m_1, m_2, m_3$ , 几何重数为  $d_1, d_2, d_3$ . 已知  $A$  是  $5 \times 5$ , 所以  $\sum m_i = 5$ . 已知其中一个特征子空间维数为 3, 不妨设  $d_1 = 3$ . 因为代数重数总是大于等于几何重数, 所以  $m_1 \geq 3$ . 剩下两个特征值  $\lambda_2, \lambda_3$  各自至少有 1 的代数重数 ( $m_2 \geq 1, m_3 \geq 1$ ). 因为  $m_1 + m_2 + m_3 = 5$  且  $m_1 \geq 3, m_2 \geq 1, m_3 \geq 1$ , 唯一可能的整数解是  $m_1 = 3, m_2 = 1, m_3 = 1$ . 对于几何重数, 我们有  $1 \leq d_i \leq m_i$ . 因此  $d_1 = 3$  (已知),  $d_2 = 1$  (因为  $m_2 = 1$ ),  $d_3 = 1$  (因为  $m_3 = 1$ ). 总的几何重数之和  $\sum d_i = 3 + 1 + 1 = 5$ . 因为特征向量的总维数等于空间维数  $n = 5$ , 所以  $A$  可对角化.

2.11. 解: 最简单的例子是若尔当块 (Jordan block):

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

这里  $\lambda = 1$  的代数重数为 2, 但几何重数为 1 (秩为 1 的矩阵  $A - I$  的零空间维数), 所以不可对角化. 为了使其“通用”, 取任意可逆矩阵  $S$ , 计算  $A = SJS^{-1}$ . 例如取

$S = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ , 算出  $A$  后, 其不可对角化的性质保持不变, 但矩阵元素看起来不再那么稀疏.

2.12. 证明: 假设  $A$  可对角化, 即  $A = SDS^{-1}$ . 因为  $A^N = 0$ , 所以:

$$(SDS^{-1})^N = SD^N S^{-1} = 0$$

两边左乘  $S^{-1}$  右乘  $S$ , 得到  $D^N = 0$ .  $D$  是对角矩阵, 其  $N$  次方也是对角矩阵, 对角线上是特征值的  $N$  次方  $\lambda_i^N$ .  $\lambda_i^N = 0 \implies \lambda_i = 0$ . 这意味着  $D$  必须是零矩阵. 如果  $D = 0$ , 则  $A = S \cdot 0 \cdot S^{-1} = 0$ . 这与题目假设  $A$  是非零矩阵矛盾. 因此  $A$  不能被对角化.

2.13. 解: **a)** 我们要解  $T(A) = \lambda A$ , 即  $A^T = \lambda A$ . 对两边再取转置:  $(A^T)^T = \lambda A^T \implies A = \lambda(\lambda A) = \lambda^2 A$ . 因为  $A$  是非零向量 (矩阵空间中的非零向量), 所以  $\lambda^2 = 1$ , 即  $\lambda = 1$  或  $\lambda = -1$ .

对于  $\lambda = 1$ :  $A^T = A$  (对称矩阵). 基:  $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . 维数为 3.

对于  $\lambda = -1$ :  $A^T = -A$  (反对称矩阵). 基:  $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ . 维数为 1. 总维数  $3 + 1 = 4$ , 等于空间  $M_{2 \times 2}$  的维数. 所以该变换可对角化.

**b)** 对于  $M_{n \times n}$ , 同样的逻辑适用. 特征值只能是  $\pm 1$ . 特征空间为对称矩阵空间 ( $\lambda = 1$ , 维数  $\frac{n(n+1)}{2}$ ) 和反对称矩阵空间 ( $\lambda = -1$ , 维数  $\frac{n(n-1)}{2}$ ). 维数之和:

$$\frac{n^2 + n}{2} + \frac{n^2 - n}{2} = n^2 = \dim M_{n \times n}$$

因此总是可对角化的.

2.14. 证明: ( $\implies$ ) 假设  $V_1, V_2$  线性无关. 这意味着方程  $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$  ( $\mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2$ ) 只有平凡解  $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$ . 设  $\mathbf{v} \in V_1 \cap V_2$ . 因为  $\mathbf{v} \in V_1$ , 我们可写  $\mathbf{v}_1 = \mathbf{v}$ . 因为  $\mathbf{v} \in V_2$ , 我们可写  $\mathbf{v}_2 = -\mathbf{v}$ . 考虑和:  $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ . 由线性无关性, 必须有  $\mathbf{v}_1 = \mathbf{0}$  和  $\mathbf{v}_2 = \mathbf{0}$ . 所以  $\mathbf{v} = \mathbf{0}$ . 即交集只有零向量.

( $\impliedby$ ) 假设  $V_1 \cap V_2 = \{\mathbf{0}\}$ . 考虑方程  $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$ , 其中  $\mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2$ . 这暗示  $\mathbf{v}_1 = -\mathbf{v}_2$ . 左边属于  $V_1$ , 右边属于  $V_2$  (因为  $V_2$  是子空间, 对数乘封闭). 所以  $\mathbf{v}_1$  既在  $V_1$  也在  $V_2$  中, 即  $\mathbf{v}_1 \in V_1 \cap V_2$ . 由假设,  $\mathbf{v}_1 = \mathbf{0}$ . 进而  $\mathbf{v}_2 = -\mathbf{v}_1 = \mathbf{0}$ . 解是唯一的 (平凡的), 所以子空间线性无关.

## 第五章习题解答

1.1. 解: 1)

$$(3 + 2i)(5 - 3i) = 15 - 9i + 10i - 6i^2 = 15 + i - 6(-1) = 21 + i$$

2)

$$\frac{2 - 3i}{1 - 2i} = \frac{(2 - 3i)(1 + 2i)}{(1 - 2i)(1 + 2i)} = \frac{2 + 4i - 3i - 6i^2}{1^2 + 2^2} = \frac{2 + i + 6}{5} = \frac{8 + i}{5} = 1.6 + 0.2i$$

3)

$$\operatorname{Re}\left(\frac{2 - 3i}{1 - 2i}\right) = \operatorname{Re}(1.6 + 0.2i) = 1.6$$

4)

$$(1 + 2i)^3 = 1^3 + 3(1)^2(2i) + 3(1)(2i)^2 + (2i)^3 = 1 + 6i + 3(-4) + (-8i) = 1 + 6i - 12 - 8i = -11 - 2i$$

5)

$$\operatorname{Im}((1 + 2i)^3) = \operatorname{Im}(-11 - 2i) = -2$$

1.2. 解: 注意: 复数空间  $\mathbb{C}^n$  上的标准内积定义为  $(\mathbf{x}, \mathbf{y}) = \sum x_k \overline{y_k}$  (关于第一个变量线性, 关于第二个变量共轭线性).

a)

$$\begin{aligned}
 (\mathbf{x}, \mathbf{y}) &= 1 \cdot \bar{i} + 2i \cdot \overline{(2-i)} + (1+i) \cdot \bar{3} \\
 &= 1(-i) + 2i(2+i) + (1+i)(3) \\
 &= -i + 4i + 2i^2 + 3 + 3i \\
 &= -i + 4i - 2 + 3 + 3i \\
 &= 1 + 6i
 \end{aligned}$$

$$\begin{aligned}
 \|\mathbf{x}\|^2 &= |1|^2 + |2i|^2 + |1+i|^2 = 1 + 4 + (1^2 + 1^2) = 1 + 4 + 2 = 7 \\
 \|\mathbf{y}\|^2 &= |i|^2 + |2-i|^2 + |3|^2 = 1 + (2^2 + (-1)^2) + 9 = 1 + 5 + 9 = 15 \\
 \|\mathbf{y}\| &= \sqrt{15}
 \end{aligned}$$

b) 利用内积的性质:  $(\alpha\mathbf{x}, \beta\mathbf{y}) = \alpha\bar{\beta}(\mathbf{x}, \mathbf{y})$ .

$$\begin{aligned}
 (3\mathbf{x}, 2i\mathbf{y}) &= 3 \cdot \bar{2i}(\mathbf{x}, \mathbf{y}) = 3(-2i)(1+6i) = -6i(1+6i) = -6i - 36i^2 = 36 - 6i \\
 (2\mathbf{x}, i\mathbf{x} + 2\mathbf{y}) &= (2\mathbf{x}, i\mathbf{x}) + (2\mathbf{x}, 2\mathbf{y}) \\
 &= 2 \cdot \bar{i}(\mathbf{x}, \mathbf{x}) + 2 \cdot \bar{2}(\mathbf{x}, \mathbf{y}) \\
 &= -2i\|\mathbf{x}\|^2 + 4(\mathbf{x}, \mathbf{y}) \\
 &= -2i(7) + 4(1+6i) \\
 &= -14i + 4 + 24i \\
 &= 4 + 10i
 \end{aligned}$$

c)

$$\begin{aligned}
 \|\mathbf{x} + 2\mathbf{y}\|^2 &= (\mathbf{x} + 2\mathbf{y}, \mathbf{x} + 2\mathbf{y}) \\
 &= \|\mathbf{x}\|^2 + (\mathbf{x}, 2\mathbf{y}) + (2\mathbf{y}, \mathbf{x}) + \|2\mathbf{y}\|^2 \\
 &= \|\mathbf{x}\|^2 + \bar{2}(\mathbf{x}, \mathbf{y}) + 2\overline{(\mathbf{x}, \mathbf{y})} + 4\|\mathbf{y}\|^2 \\
 &= 7 + 2(1+6i) + 2(1-6i) + 4(15) \\
 &= 7 + 2 + 12i + 2 - 12i + 60 \\
 &= 71
 \end{aligned}$$

所以  $\|\mathbf{x} + 2\mathbf{y}\| = \sqrt{71}$ .

1.3. 解: 1)

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\operatorname{Re}(\mathbf{u}, \mathbf{v}) = 4 + 9 + 2(2) = 17$$

2)

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\operatorname{Re}(\mathbf{u}, \mathbf{v}) = 4 + 9 - 4 = 9$$

3)

$$\begin{aligned}
 (\mathbf{u} + \mathbf{v}, \mathbf{u} - i\mathbf{v}) &= (\mathbf{u}, \mathbf{u}) + (\mathbf{u}, -i\mathbf{v}) + (\mathbf{v}, \mathbf{u}) + (\mathbf{v}, -i\mathbf{v}) \\
 &= \|\mathbf{u}\|^2 + \overline{(-i)}(\mathbf{u}, \mathbf{v}) + \overline{(\mathbf{u}, \mathbf{v})} + \overline{(-i)}\|\mathbf{v}\|^2 \\
 &= 4 + i(2+i) + (2-i) + i(9) \\
 &= 4 + 2i - 1 + 2 - i + 9i \\
 &= 5 + 10i
 \end{aligned}$$

4)

$$\begin{aligned}
 (\mathbf{u} + 3i\mathbf{v}, 4i\mathbf{u}) &= \bar{4i}(\mathbf{u} + 3i\mathbf{v}, \mathbf{u}) \\
 &= -4i[(\mathbf{u}, \mathbf{u}) + 3i(\mathbf{v}, \mathbf{u})] \\
 &= -4i[4 + 3i(2-i)] \\
 &= -4i[4 + 6i + 3] \\
 &= -4i[7 + 6i] \\
 &= -28i - 24i^2 \\
 &= 24 - 28i
 \end{aligned}$$

1.4. 证明:

$$\begin{aligned}\|\mathbf{x} \pm \mathbf{y}\|^2 &= (\mathbf{x} \pm \mathbf{y}, \mathbf{x} \pm \mathbf{y}) \\ &= (\mathbf{x}, \mathbf{x}) \pm (\mathbf{x}, \mathbf{y}) \pm (\mathbf{y}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) \\ &= \|\mathbf{x}\|^2 \pm (\mathbf{x}, \mathbf{y}) \pm \overline{(\mathbf{x}, \mathbf{y})} + \|\mathbf{y}\|^2 \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \pm \left( (\mathbf{x}, \mathbf{y}) + \overline{(\mathbf{x}, \mathbf{y})} \right)\end{aligned}$$

因为  $z + \bar{z} = 2\operatorname{Re}(z)$ , 所以

$$\|\mathbf{x} \pm \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \pm 2\operatorname{Re}(\mathbf{x}, \mathbf{y})$$

1.5. 解: **a)** 违反非负性 (positivity). 取  $\mathbf{x} = (0, 1)^T$ , 则  $(\mathbf{x}, \mathbf{x}) = 0 \cdot 0 - 1 \cdot 1 = -1 < 0$ . 内积要求对所有  $\mathbf{x} \neq \mathbf{0}$ ,  $(\mathbf{x}, \mathbf{x}) > 0$ .

**b)** 违反非负性. 取  $A = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ .  $(A, A) = \operatorname{trace}(A + A) = \operatorname{trace}(-2I) = -4 < 0$ . (注: 此外, 内积通常要求关于第一个变量线性,  $(A, B) = \operatorname{trace}(A + B)$  甚至不是双线性的. 例如  $(2A, B) = \operatorname{trace}(2A + B) \neq 2\operatorname{trace}(A + B)$ .)

**c)** 违反非退化性 (definiteness) 和共轭对称性. **非退化性**: 取  $f(t) = 1$  (常数函数),  $f$  不是零向量.  $f'(t) = 0$ , 因此  $(f, f) = \int_0^1 0 \cdot \bar{1} dt = 0$ . 但  $f \neq \mathbf{0}$ . **共轭对称性**: 一般情况下,  $\int f' \bar{g} \neq \overline{\int g' \bar{f}} = \int \bar{g}' f$ .

1.6. 证明: 如果  $\mathbf{y} = \mathbf{0}$ , 则两边均为 0, 等式成立, 且  $\mathbf{y} = 0\mathbf{x}$ , 结论成立. 假设  $\mathbf{y} \neq \mathbf{0}$ . 回顾柯西-施瓦茨不等式的证明, 我们考虑函数  $f(t) = \|\mathbf{x} - t\mathbf{y}\|^2$  (对于实数情况) 或令  $t = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2}$ . 关键步骤在于正交投影的误差项:

$$0 \leq \left\| \mathbf{x} - \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2} \mathbf{y} \right\|^2 = \|\mathbf{x}\|^2 - \frac{|(\mathbf{x}, \mathbf{y})|^2}{\|\mathbf{y}\|^2}$$

等号  $|(\mathbf{x}, \mathbf{y})| = \|\mathbf{x}\|\|\mathbf{y}\|$  成立当且仅当

$$\left\| \mathbf{x} - \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2} \mathbf{y} \right\|^2 = 0$$

根据范数的非退化性, 这等价于

$$\mathbf{x} - \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2} \mathbf{y} = \mathbf{0}$$

即  $\mathbf{x} = c\mathbf{y}$ , 其中  $c = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{y}\|^2}$ . 这表明  $\mathbf{x}$  是  $\mathbf{y}$  的倍数.

1.7. 证明: 根据习题 1.4 的结论:

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\operatorname{Re}(\mathbf{x}, \mathbf{y})$$

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\operatorname{Re}(\mathbf{x}, \mathbf{y})$$

将上述两式相加:

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 &= 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 + (2\operatorname{Re}(\mathbf{x}, \mathbf{y}) - 2\operatorname{Re}(\mathbf{x}, \mathbf{y})) \\ &= 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)\end{aligned}$$

1.8. 证明: **a)** 因为  $(\mathbf{x}, \mathbf{v}) = 0$  对所有  $\mathbf{v}$  成立, 我们可以取  $\mathbf{v} = \mathbf{x}$ . 则  $(\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|^2 = 0$ . 由内积的非退化性可知,  $\mathbf{x} = \mathbf{0}$ .

**b)** 设  $\mathbf{v}$  是  $V$  中的任意向量. 因为  $\mathbf{v}_k$  生成  $V$ , 所以存在标量  $c_1, \dots, c_n$  使得  $\mathbf{v} = \sum_{k=1}^n c_k \mathbf{v}_k$ . 计算内积:

$$(\mathbf{x}, \mathbf{v}) = \left( \mathbf{x}, \sum_{k=1}^n c_k \mathbf{v}_k \right) = \sum_{k=1}^n \bar{c}_k (\mathbf{x}, \mathbf{v}_k)$$

已知  $(\mathbf{x}, \mathbf{v}_k) = 0$ , 所以对于任意  $\mathbf{v} \in V$  都有  $(\mathbf{x}, \mathbf{v}) = 0$ . 由 **a)** 部分结论可知,  $\mathbf{x} = \mathbf{0}$ .

**c)** 已知  $(\mathbf{x}, \mathbf{v}_k) = (\mathbf{y}, \mathbf{v}_k)$ , 即  $(\mathbf{x}, \mathbf{v}_k) - (\mathbf{y}, \mathbf{v}_k) = 0$ . 利用内积的线性性质:

$$(\mathbf{x} - \mathbf{y}, \mathbf{v}_k) = 0 \quad \forall k$$

应用 b) 部分结论 (令新向量  $\mathbf{z} = \mathbf{x} - \mathbf{y}$ ), 可得  $\mathbf{z} = \mathbf{0}$ , 即  $\mathbf{x} = \mathbf{y}$ .

### 1.9. 解: 1) 绘制形状

$p = 1$ : 定义为  $|x_1| + |x_2| \leq 1$ . 这是一个以原点为中心, 顶点在  $(\pm 1, 0)$  和  $(0, \pm 1)$  的正方形 (菱形), 其边相对于坐标轴旋转了  $45^\circ$ .

$p = 2$ : 定义为  $\sqrt{x_1^2 + x_2^2} \leq 1 \implies x_1^2 + x_2^2 \leq 1$ . 这是一个以原点为中心的单位圆盘.

$p = \infty$ : 定义为  $\max(|x_1|, |x_2|) \leq 1$ , 即  $|x_1| \leq 1$  且  $|x_2| \leq 1$ . 这是一个顶点在  $(\pm 1, \pm 1)$  的正方形, 其边平行于坐标轴.

2) 猜测其他  $p$  对于  $1 < p < \infty$ , 单位球  $B_p$  是介于  $p = 1$  的菱形和  $p = \infty$  的正方形之间的凸集. 随着  $p$  增大, 形状从菱形逐渐向外膨胀, 经过圆形 ( $p = 2$ ), 并随着  $p \rightarrow \infty$  逐渐充满整个  $p = \infty$  的正方形 (变得越来越像圆角正方形, 即超椭圆).

### 2.1. 解: 设所求向量为 $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ . 根据正交性的定义, 我们需要满足:

$$\mathbf{x} \cdot \mathbf{v}_1 = 0 \quad \text{且} \quad \mathbf{x} \cdot \mathbf{v}_2 = 0$$

这等价于求解齐次线性方程组  $A\mathbf{x} = \mathbf{0}$ , 其中  $A$  的行由这两个向量组成:

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

对矩阵进行行约简:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} \xrightarrow{R_2 - R_1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \xrightarrow{R_1 - R_2} \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$

这对应于方程组:

$$\begin{cases} x_1 - x_3 - 2x_4 = 0 \\ x_2 + 2x_3 + 3x_4 = 0 \end{cases} \implies \begin{cases} x_1 = x_3 + 2x_4 \\ x_2 = -2x_3 - 3x_4 \end{cases}$$

设  $x_3 = s, x_4 = t$  为自由变量, 则通解为:

$$\mathbf{x} = \begin{pmatrix} s + 2t \\ -2s - 3t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix}$$

即所有满足条件的向量构成的集合是由  $(1, -2, 1, 0)^T$  和  $(2, -3, 0, 1)^T$  生成的子空间.

### 2.2. 解: 根据基本子空间定理 (或正交补的性质): 1) $(\text{Ran } A^T)^\perp = \text{Ker } A$ . 证明思路: $\mathbf{x} \in (\text{Ran } A^T)^\perp$ 当且仅当 $\mathbf{x}$ 正交于 $A^T$ 的列空间, 即 $A\mathbf{x} = \mathbf{0}$ .

2)  $(\text{Ran } A)^\perp = \text{Ker } A^T$ . 证明思路:  $\mathbf{y} \in (\text{Ran } A)^\perp$  当且仅当  $\mathbf{y}$  正交于  $A$  的每一列, 即  $A^T\mathbf{y} = \mathbf{0}$ .

### 2.3. 证明: a) 利用内积的双线性 (对第一变量线性, 对第二变量共轭线性):

$$\begin{aligned} (\mathbf{x}, \mathbf{y}) &= \left( \sum_{j=1}^n \alpha_j \mathbf{v}_j, \sum_{k=1}^n \beta_k \mathbf{v}_k \right) \\ &= \sum_{j=1}^n \alpha_j \left( \mathbf{v}_j, \sum_{k=1}^n \beta_k \mathbf{v}_k \right) \\ &= \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\beta_k} (\mathbf{v}_j, \mathbf{v}_k) \end{aligned}$$

因为  $\mathbf{v}_k$  是标准正交基, 所以  $(\mathbf{v}_j, \mathbf{v}_k) = \delta_{jk}$  (当  $j = k$  时为 1, 否则为 0). 双重求和中只有  $j = k$  的项非零, 因此:

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n \alpha_k \overline{\beta_k}$$



b) 对于标准正交基, 向量的坐标由傅里叶系数给出:  $\alpha_k = (\mathbf{x}, \mathbf{v}_k)$  且  $\beta_k = (\mathbf{y}, \mathbf{v}_k)$ . 直接代入 a) 中的公式即得:

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n (\mathbf{x}, \mathbf{v}_k) \overline{(\mathbf{y}, \mathbf{v}_k)}$$

c) 如果基只是正交的 ( $\|\mathbf{v}_k\| \neq 1$ ), 则  $(\mathbf{v}_j, \mathbf{v}_k) = \|\mathbf{v}_k\|^2 \delta_{jk}$ . 此时, 坐标系数为  $\alpha_k = \frac{(\mathbf{x}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}$ ,  $\beta_k = \frac{(\mathbf{y}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}$ . 代入求和公式:

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n \alpha_k \overline{\beta_k} \|\mathbf{v}_k\|^2 = \sum_{k=1}^n \frac{(\mathbf{x}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \overline{\frac{(\mathbf{y}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2}} \|\mathbf{v}_k\|^2$$

化简得正交基下的帕塞瓦尔恒等式:

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n \frac{(\mathbf{x}, \mathbf{v}_k) \overline{(\mathbf{y}, \mathbf{v}_k)}}{\|\mathbf{v}_k\|^2}$$

2.4. 证明: 我们需要验证内积的三个性质: 1. 共轭对称性:

$$\langle \mathbf{y}, \mathbf{x} \rangle = \sum_{k=1}^n \beta_k \overline{\alpha_k} = \sum_{k=1}^n \overline{\alpha_k \beta_k} = \overline{\sum_{k=1}^n \alpha_k \beta_k} = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$$

2. 第一变量的线性性: 设  $\mathbf{z} = \sum \gamma_k \mathbf{v}_k$ . 则  $\mathbf{x} + c\mathbf{z}$  的第  $k$  个系数为  $\alpha_k + c\gamma_k$ .

$$\langle \mathbf{x} + c\mathbf{z}, \mathbf{y} \rangle = \sum_{k=1}^n (\alpha_k + c\gamma_k) \overline{\beta_k} = \sum_{k=1}^n \alpha_k \overline{\beta_k} + c \sum_{k=1}^n \gamma_k \overline{\beta_k} = \langle \mathbf{x}, \mathbf{y} \rangle + c \langle \mathbf{z}, \mathbf{y} \rangle$$

3. 非退化性 (正定性):

$$\langle \mathbf{x}, \mathbf{x} \rangle = \sum_{k=1}^n \alpha_k \overline{\alpha_k} = \sum_{k=1}^n |\alpha_k|^2 \geq 0$$

若  $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ , 则所有  $|\alpha_k|^2 = 0$ , 即  $\alpha_k = 0$  对所有  $k$  成立. 因为  $\mathbf{v}_k$  是一组基, 零系数意味着  $\mathbf{x} = \sum 0\mathbf{v}_k = \mathbf{0}$ .

综上,  $\langle \mathbf{x}, \mathbf{y} \rangle$  定义了一个内积. (注: 这实际上是使得  $\mathbf{v}_k$  成为标准正交基的那个内积)。

2.5. 解: 这等同于求  $(\text{Ran } A)^\perp$ . 根据四个基本子空间的性质 (或习题 2.2 的结论): 该集合是  $A^T$  的核 (零空间), 即  $\text{Ker } A^T$ .

$$\{\mathbf{y} \in \mathbb{F}^m : \mathbf{y} \perp \text{Ran } A\} = \text{Ker } A^T$$

(注: 如果是在  $\mathbb{C}^m$  中且考虑标准复内积, 通常定义为  $\text{Ker } A^*$ , 但因为  $A$  是实矩阵, 所以  $A^* = A^T$ .)

3.1. 解: 设原向量为  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ . 1) 令  $\mathbf{u}_1 = \mathbf{v}_1 = (1, 2, -2)^T$ . 计算范数平方:  $\|\mathbf{u}_1\|^2 = 1^2 + 2^2 + (-2)^2 = 1 + 4 + 4 = 9$ .

2) 计算  $\mathbf{u}_2$ :

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{(\mathbf{v}_2, \mathbf{u}_1)}{\|\mathbf{u}_1\|^2} \mathbf{u}_1$$

内积  $(\mathbf{v}_2, \mathbf{u}_1) = 1(1) + (-1)(2) + 4(-2) = 1 - 2 - 8 = -9$ .

$$\mathbf{u}_2 = (1, -1, 4)^T - \frac{-9}{9} (1, 2, -2)^T = (1, -1, 4)^T + (1, 2, -2)^T = (2, 1, 2)^T$$

检查正交性:  $(\mathbf{u}_2, \mathbf{u}_1) = 2 + 2 - 4 = 0$ . 计算范数平方:  $\|\mathbf{u}_2\|^2 = 2^2 + 1^2 + 2^2 = 9$ .

3) 计算  $\mathbf{u}_3$ :

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{(\mathbf{v}_3, \mathbf{u}_1)}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 - \frac{(\mathbf{v}_3, \mathbf{u}_2)}{\|\mathbf{u}_2\|^2} \mathbf{u}_2$$

内积  $(\mathbf{v}_3, \mathbf{u}_1) = 2(1) + 1(2) + 1(-2) = 2$ . 内积  $(\mathbf{v}_3, \mathbf{u}_2) = 2(2) + 1(1) + 1(2) = 7$ .

$$\mathbf{u}_3 = (2, 1, 1)^T - \frac{2}{9}(1, 2, -2)^T - \frac{7}{9}(2, 1, 2)^T$$

为了消除分母, 我们可以先计算  $9\mathbf{u}_3$ :

$$9\mathbf{u}_3 = (18, 9, 9)^T - (2, 4, -4)^T - (14, 7, 14)^T = (18-2-14, 9-4-7, 9+4-14)^T = (2, -2, -1)^T$$

取  $\mathbf{u}_3 = (2, -2, -1)^T$  (或者  $\frac{1}{9}(2, -2, -1)^T$ ).

正交系统为:  $\{(1, 2, -2)^T, (2, 1, 2)^T, (2, -2, -1)^T\}$ .

3.2. 解: 1) 格拉姆-施密特正交化 设  $\mathbf{v}_1 = (1, 2, 3)^T, \mathbf{v}_2 = (1, 3, 1)^T$ .  $\mathbf{u}_1 = \mathbf{v}_1 = (1, 2, 3)^T$ .  $\|\mathbf{u}_1\|^2 = 1 + 4 + 9 = 14$ .

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{(\mathbf{v}_2, \mathbf{u}_1)}{\|\mathbf{u}_1\|^2} \mathbf{u}_1$$

$$(\mathbf{v}_2, \mathbf{u}_1) = 1 + 6 + 3 = 10.$$

$$\mathbf{u}_2 = (1, 3, 1)^T - \frac{10}{14}(1, 2, 3)^T = (1, 3, 1)^T - \frac{5}{7}(1, 2, 3)^T = \frac{1}{7}((7, 21, 7)^T - (5, 10, 15)^T) = \frac{1}{7}(2, 11, -8)^T$$

我们可以丢弃系数  $1/7$ , 取  $\mathbf{u}_2 = (2, 11, -8)^T$ . 验证:  $(\mathbf{u}_1, \mathbf{u}_2) = 2 + 22 - 24 = 0$ .

2) 正交投影矩阵方法一:  $P = \frac{\mathbf{u}_1 \mathbf{u}_1^T}{\|\mathbf{u}_1\|^2} + \frac{\mathbf{u}_2 \mathbf{u}_2^T}{\|\mathbf{u}_2\|^2}$ . 方法二:  $P = I - P_{E^\perp}$ . 子空间  $E$  在  $\mathbb{R}^3$

中是二维的, 其正交补  $E^\perp$  是一维的, 由法向量  $\mathbf{n}$  生成.  $\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 3 \\ 1 & 3 & 1 \end{vmatrix} =$

$$(2-9, 3-1, 3-2)^T = (-7, 2, 1)^T.$$

$$P_{E^\perp} = \frac{\mathbf{n} \mathbf{n}^T}{\|\mathbf{n}\|^2}$$

$$\|\mathbf{n}\|^2 = 49 + 4 + 1 = 54.$$

$$\mathbf{n} \mathbf{n}^T = \begin{pmatrix} -7 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} -7 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 49 & -14 & -7 \\ -14 & 4 & 2 \\ -7 & 2 & 1 \end{pmatrix}$$

$$P_E = I - \frac{1}{54} \begin{pmatrix} 49 & -14 & -7 \\ -14 & 4 & 2 \\ -7 & 2 & 1 \end{pmatrix} = \frac{1}{54} \begin{pmatrix} 54-49 & 14 & 7 \\ 14 & 54-4 & -2 \\ 7 & -2 & 54-1 \end{pmatrix} = \frac{1}{54} \begin{pmatrix} 5 & 14 & 7 \\ 14 & 50 & -2 \\ 7 & -2 & 53 \end{pmatrix}$$

3.3. 解: 1) 补全  $\mathbb{R}^3$  我们在 3.2 中通过叉积找到了正交于前两个向量的向量  $\mathbf{u}_3 = (-7, 2, 1)^T$ . 所以补全后的基为:  $\{(1, 2, 3)^T, (2, 11, -8)^T, (-7, 2, 1)^T\}$ .

2) 一般情况描述 设当前正交系为  $\mathbf{u}_1, \dots, \mathbf{u}_k$ . 步骤: 1. 在空间中选取一个不属于  $\text{span}(\mathbf{u}_1, \dots, \mathbf{u}_k)$  的向量  $\mathbf{v}$  (例如从标准基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  中尝试).

2. 应用格拉姆-施密特过程, 将  $\mathbf{v}$  对  $\mathbf{u}_1, \dots, \mathbf{u}_k$  进行正交化, 得到  $\mathbf{u}_{k+1}$ .

3. 重复此过程直到得到  $n$  个向量。(或者, 可以构建矩阵  $A$  以  $\mathbf{u}_i$  为行, 求其零空间  $\text{Null}(A)$  的基, 并对该基进行正交化。)

3.4. 解: 设子空间为  $E$ . 向量  $\mathbf{x} = (2, 3, 1)^T$  到  $E$  的距离等于  $\mathbf{x}$  在  $E^\perp$  上投影的长度, 即  $\|P_{E^\perp} \mathbf{x}\|$ . 由 3.2 可知,  $E^\perp$  由法向量  $\mathbf{n} = (-7, 2, 1)^T$  生成。

$$P_{E^\perp} \mathbf{x} = \frac{(\mathbf{x}, \mathbf{n})}{\|\mathbf{n}\|^2} \mathbf{n}$$

$$(\mathbf{x}, \mathbf{n}) = 2(-7) + 3(2) + 1(1) = -14 + 6 + 1 = -7. \quad \|\mathbf{n}\|^2 = 54.$$

$$\|P_{E^\perp} \mathbf{x}\| = \frac{|(\mathbf{x}, \mathbf{n})|}{\|\mathbf{n}\|^2} \|\mathbf{n}\| = \frac{7}{54} \sqrt{54} = \frac{7}{\sqrt{54}} = \frac{7}{3\sqrt{6}}$$

3.5. 解: 因为  $\mathbf{v}_1$  和  $\mathbf{v}_2$  已经正交 (验证:  $2-3+1+0=0$ ), 我们可以直接使用正交投影公式:

$$P\mathbf{x} = \frac{(\mathbf{x}, \mathbf{v}_1)}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \frac{(\mathbf{x}, \mathbf{v}_2)}{\|\mathbf{v}_2\|^2} \mathbf{v}_2$$

设  $\mathbf{x} = (1, 1, 1, 1)^T$ .  $(\mathbf{x}, \mathbf{v}_1) = 1 + 3 + 1 + 1 = 6$ .  $\|\mathbf{v}_1\|^2 = 1 + 9 + 1 + 1 = 12$ . 第一项系数:  $6/12 = 1/2$ .

$(\mathbf{x}, \mathbf{v}_2) = 2 - 1 + 1 + 0 = 2$ .  $\|\mathbf{v}_2\|^2 = 4 + 1 + 1 + 0 = 6$ . 第二项系数:  $2/6 = 1/3$ .

$$P\mathbf{x} = \frac{1}{2} \begin{pmatrix} 1 \\ 3 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 + 2/3 \\ 3/2 - 1/3 \\ 1/2 + 1/3 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 7/6 \\ 7/6 \\ 5/6 \\ 1/2 \end{pmatrix}$$

3.6. 解: 验证正交性:  $1 - 2 + 1 + 0 = 0$ .  $\mathbf{v}_1 \perp \mathbf{v}_2$ . 使用勾股定理:  $\text{distance}(\mathbf{x}, E)^2 = \|\mathbf{x}\|^2 - \|P_E \mathbf{x}\|^2$ . 设  $\mathbf{x} = (1, 2, 3, 4)^T$ .  $\|\mathbf{x}\|^2 = 1 + 4 + 9 + 16 = 30$ .

由于  $\mathbf{v}_1, \mathbf{v}_2$  正交,  $\|P_E \mathbf{x}\|^2 = \frac{|\langle \mathbf{x}, \mathbf{v}_1 \rangle|^2}{\|\mathbf{v}_1\|^2} + \frac{|\langle \mathbf{x}, \mathbf{v}_2 \rangle|^2}{\|\mathbf{v}_2\|^2} = \frac{|\langle \mathbf{x}, \mathbf{v}_1 \rangle|^2}{\|\mathbf{v}_1\|^2} + \frac{|\langle \mathbf{x}, \mathbf{v}_2 \rangle|^2}{\|\mathbf{v}_2\|^2}$ .

$(\mathbf{x}, \mathbf{v}_1) = 1 - 2 + 3 + 0 = 2$ .  $\|\mathbf{v}_1\|^2 = 1 + 1 + 1 = 3$ . 第一项贡献:  $2^2/3 = 4/3$ .

$(\mathbf{x}, \mathbf{v}_2) = 1 + 4 + 3 + 4 = 12$ .  $\|\mathbf{v}_2\|^2 = 1 + 4 + 1 + 1 = 7$ . 第二项贡献:  $12^2/7 = 144/7$ .

$$\|P_E \mathbf{x}\|^2 = \frac{4}{3} + \frac{144}{7} = \frac{28 + 432}{21} = \frac{460}{21}$$

$$\text{dist}^2 = 30 - \frac{460}{21} = \frac{630 - 460}{21} = \frac{170}{21}$$

距离为  $\sqrt{\frac{170}{21}}$ .

3.7. 解: **正确** (假设  $V$  是有限维内积空间). 证明: 根据正交分解定理, 任何向量  $\mathbf{v} \in V$  可以唯一表示为  $\mathbf{v} = \mathbf{x} + \mathbf{y}$ , 其中  $\mathbf{x} \in E, \mathbf{y} \in E^\perp$ . 这意味着  $V = E \oplus E^\perp$  (直和). 对于直和, 维数相加:  $\dim V = \dim E + \dim E^\perp$ .

3.8. 解: 正交投影满足  $P^2 = P$ . **特征值**:  $\lambda^2 = \lambda \implies \lambda = 1$  或  $\lambda = 0$ .

**特征向量**: 1. 对于任何  $\mathbf{x} \in E$ , 有  $P\mathbf{x} = \mathbf{x} = 1 \cdot \mathbf{x}$ . 所以  $E$  是对应于  $\lambda = 1$  的特征子空间.

2. 对于任何  $\mathbf{y} \in E^\perp$ , 有  $P\mathbf{y} = \mathbf{0} = 0 \cdot \mathbf{y}$ . 所以  $E^\perp$  是对应于  $\lambda = 0$  的特征子空间.

**重数**:  $\lambda = 1$ : 几何重数  $= \dim E = r$ .  $\lambda = 0$ : 几何重数  $= \dim E^\perp = n - r$ . 由于几何重数之和  $r + (n - r) = n = \dim V$ , 存在特征向量基, 因此矩阵可对角化, 代数重数等于几何重数.

3.9. 解: **a)** 设  $\mathbf{u} = (1, 1, \dots, 1)^T$ .  $P = \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2}$ .  $\|\mathbf{u}\|^2 = n$ .  $\mathbf{u}\mathbf{u}^T$  是全 1 矩阵 (记为  $J$ ). 所以  $P = \frac{1}{n}J$ .

**b)** 题目描述的矩阵  $A$  实际上就是全 1 矩阵  $J$ . 由 **a)** 知  $J = nP$ .  $P$  的特征值是 1 (重数 1) 和 0 (重数  $n - 1$ ). 所以  $A = nP$  的特征值是  $n \times 1 = n$  (重数 1) 和  $n \times 0 = 0$  (重数  $n - 1$ ).

**c)**  $A - I = J - I$  (这是全 1 矩阵减去单位矩阵, 即对角线为 0, 其余为 1). 其特征值为  $J$  的特征值减去 1. 特征值:  $n - 1$  (重数 1) 和  $0 - 1 = -1$  (重数  $n - 1$ ).

**d)** 行列式等于特征值之积.

$$\det(A - I) = (n - 1)^1 \cdot (-1)^{n-1} = (-1)^{n-1}(n - 1)$$

3.10. 解: 令  $v_0 = 1, v_1 = t, v_2 = t^2, v_3 = t^3$ . **1)**  $u_0 = 1$ .  $\|u_0\|^2 = \int_{-1}^1 1 dt = 2$ .

**2)**  $u_1 = t - \frac{(t, 1)}{\|u_0\|^2} 1$ .  $(t, 1) = \int_{-1}^1 t dt = 0$  (奇函数). 所以  $u_1 = t$ .  $\|u_1\|^2 = \int_{-1}^1 t^2 dt = [\frac{t^3}{3}]_{-1}^1 = \frac{2}{3}$ .

**3)**  $u_2 = t^2 - \frac{(t^2, 1)}{\|u_0\|^2} 1 - \frac{(t^2, t)}{\|u_1\|^2} t$ .  $(t^2, 1) = \int_{-1}^1 t^2 dt = \frac{2}{3}$ .  $(t^2, t) = \int_{-1}^1 t^3 dt = 0$ . 所以  $u_2 = t^2 - \frac{2/3}{2}(1) - 0 = t^2 - \frac{1}{3}$ .  $\|u_2\|^2 = \int_{-1}^1 (t^2 - 1/3)^2 dt = \int_{-1}^1 (t^4 - \frac{2}{3}t^2 + \frac{1}{9}) dt = 2(\frac{1}{5} - \frac{2}{9} + \frac{1}{9}) = 2(\frac{9-10+5}{45}) = \frac{8}{45}$ .

**4)**  $u_3 = t^3 - \text{proj}_{u_0} - \text{proj}_{u_1} - \text{proj}_{u_2}$ . 由于区间对称, 奇偶性不同正交.  $t^3$  与  $u_0(1)$  和  $u_2(t^2 - 1/3)$  正交. 只剩下在  $u_1(t)$  上的投影.  $(t^3, t) = \int_{-1}^1 t^4 dt = \frac{2}{5}$ . 系数  $= \frac{2/5}{2/3} = \frac{3}{5}$ . 所以  $u_3 = t^3 - \frac{3}{5}t$ .

正交基为  $\{1, t, t^2 - \frac{1}{3}, t^3 - \frac{3}{5}t\}$ .

3.11. 证明: **a)** 对于任意  $\mathbf{v}, \mathbf{w} \in V$ , 我们可以分解  $\mathbf{v} = \mathbf{v}_E + \mathbf{v}_{E^\perp}$ ,  $\mathbf{w} = \mathbf{w}_E + \mathbf{w}_{E^\perp}$ .  $(P\mathbf{v}, \mathbf{w}) = (\mathbf{v}_E, \mathbf{w}_E + \mathbf{w}_{E^\perp}) = (\mathbf{v}_E, \mathbf{w}_E)$  (因为  $\mathbf{v}_E \perp \mathbf{w}_{E^\perp}$ ).  $(\mathbf{v}, P\mathbf{w}) = (\mathbf{v}_E + \mathbf{v}_{E^\perp}, \mathbf{w}_E) = (\mathbf{v}_E, \mathbf{w}_E)$

(因为  $\mathbf{v}_{E^\perp} \perp \mathbf{w}_E$ )。所以  $(P\mathbf{v}, \mathbf{w}) = (\mathbf{v}, P\mathbf{w})$ , 即  $P^* = P$ 。

**b)** 对于任意  $\mathbf{v}$ , 令  $\mathbf{y} = P\mathbf{v}$ . 根据定义  $\mathbf{y} \in E$ . 再次投影:  $P\mathbf{y} = \mathbf{y}$  (因为  $\mathbf{y}$  已经在  $E$  中了)。所以  $P(P\mathbf{v}) = P\mathbf{v}$ , 即  $P^2 = P$ 。

3.12. 证明: **1)**  $E \subseteq (E^\perp)^\perp$ : 对于任意  $\mathbf{x} \in E$ ,  $\mathbf{x}$  正交于  $E^\perp$  中的任何向量 (根据  $E^\perp$  的定义)。这意味着  $\mathbf{x} \in (E^\perp)^\perp$ 。

**2)**  $(E^\perp)^\perp \subseteq E$ : 对于任意  $\mathbf{y} \in (E^\perp)^\perp$ , 利用正交分解  $V = E \oplus E^\perp$ , 我们可以写  $\mathbf{y} = \mathbf{y}_E + \mathbf{y}_{E^\perp}$ . 因为  $\mathbf{y} \in (E^\perp)^\perp$ , 所以  $\mathbf{y}$  垂直于任何在  $E^\perp$  中的向量。特别是, 它垂直于  $\mathbf{y}_{E^\perp}$ . 所以  $(\mathbf{y}, \mathbf{y}_{E^\perp}) = 0$ . 代入分解式:  $(\mathbf{y}_E + \mathbf{y}_{E^\perp}, \mathbf{y}_{E^\perp}) = 0$ .  $(\mathbf{y}_E, \mathbf{y}_{E^\perp}) + (\mathbf{y}_{E^\perp}, \mathbf{y}_{E^\perp}) = 0$ . 第一项为 0, 所以  $\|\mathbf{y}_{E^\perp}\|^2 = 0 \implies \mathbf{y}_{E^\perp} = \mathbf{0}$ . 因此  $\mathbf{y} = \mathbf{y}_E \in E$ 。

3.13. 解: **a)** 对于任意  $\mathbf{v}$ ,  $\mathbf{v} = \mathbf{v}_E + \mathbf{v}_{E^\perp} = P\mathbf{v} + Q\mathbf{v} = (P + Q)\mathbf{v}$ . 所以  $P + Q = I$  (单位算子)。

$PQ$  是先投影到  $E^\perp$  再投影到  $E$ .  $Q\mathbf{v} \in E^\perp$ , 而  $E \perp E^\perp$ , 所以  $Q\mathbf{v}$  在  $E$  上的投影为  $\mathbf{0}$ . 所以  $PQ = 0$  (零算子)。

**b)** 我们要证明  $(P - Q)(P - Q) = I$ .

$$(P - Q)^2 = P^2 - PQ - QP + Q^2$$

已知  $P^2 = P, Q^2 = Q$ , 且  $PQ = 0$ . 由于  $P$  和  $Q$  自伴随,  $(PQ)^* = Q^*P^* = QP = 0^* = 0$ . 所以  $(P - Q)^2 = P - 0 - 0 + Q = P + Q = I$ . 得证。

4.1. 解: 这是一个超定方程组  $A\mathbf{x} = \mathbf{b}$ , 其中

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

我们需要求解正规方程  $A^T A\mathbf{x} = A^T \mathbf{b}$ . 计算  $A^T A$ :

$$A^T A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1+1 & 1 \\ 1 & 1+1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

计算  $A^T \mathbf{b}$ :

$$A^T \mathbf{b} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

现在求解线性方程组:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

由对称性易得  $x_1 = x_2$ .  $2x_1 + x_1 = 1 \implies 3x_1 = 1 \implies x_1 = 1/3$ . 所以最小二乘解为  $\mathbf{x} = (1/3, 1/3)^T$ .

4.2. 解: 设矩阵为  $A$ , 列向量为  $\mathbf{v}_1 = (1, 2, -2)^T, \mathbf{v}_2 = (1, -1, 4)^T$ .

**方法一: 格拉姆-施密特正交化** 我们在习题 3.1 中已经对这两个向量进行了正交化 (注意  $\mathbf{v}_2$  与 3.1 中的第二个向量略有不同, 需重新计算, 或者观察到它们其实是一样的向量  $(1, -1, 4)^T$  与  $(1, 2, -2)^T$  交换了位置? 不, 完全一致, 只是 3.1 是三个向量)。回顾 3.1 的计算结果或重新计算:  $\mathbf{u}_1 = \mathbf{v}_1 = (1, 2, -2)^T$ .  $\|\mathbf{u}_1\|^2 = 9$ .  $\mathbf{u}_2 = \mathbf{v}_2 - \frac{(\mathbf{v}_2, \mathbf{u}_1)}{\|\mathbf{u}_1\|^2} \mathbf{u}_1$ .  $(\mathbf{v}_2, \mathbf{u}_1) = 1 - 2 - 8 = -9$ .  $\mathbf{u}_2 = (1, -1, 4)^T - \frac{-9}{9}(1, 2, -2)^T = (1, -1, 4)^T + (1, 2, -2)^T = (2, 1, 2)^T$ .  $\|\mathbf{u}_2\|^2 = 4 + 1 + 4 = 9$ . 投影矩阵公式为  $P = \frac{\mathbf{u}_1 \mathbf{u}_1^T}{\|\mathbf{u}_1\|^2} + \frac{\mathbf{u}_2 \mathbf{u}_2^T}{\|\mathbf{u}_2\|^2}$ .

$$\mathbf{u}_1 \mathbf{u}_1^T = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} (1 \ 2 \ -2) = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 4 & -4 \\ -2 & -4 & 4 \end{pmatrix}$$

$$\mathbf{u}_2 \mathbf{u}_2^T = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} (2 \ 1 \ 2) = \begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix}$$

$$P = \frac{1}{9} \left[ \begin{pmatrix} 1 & 2 & -2 \\ 2 & 4 & -4 \\ -2 & -4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix} \right] = \frac{1}{9} \begin{pmatrix} 5 & 4 & 2 \\ 4 & 5 & -2 \\ 2 & -2 & 8 \end{pmatrix}.$$

方法二：投影公式  $P = A(A^T A)^{-1} A^T$

$$A^T A = \begin{pmatrix} 1 & 2 & -2 \\ 1 & -1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 9 & -9 \\ -9 & 18 \end{pmatrix} = 9 \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

求逆：

$$(A^T A)^{-1} = \frac{1}{9(2-1)} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

计算  $A(A^T A)^{-1}$ ：

$$A(A^T A)^{-1} = \frac{1}{9} \begin{pmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 3 & 2 \\ 3 & 1 \\ 0 & 2 \end{pmatrix}.$$

最后乘以  $A^T$ ：

$$P = \frac{1}{9} \begin{pmatrix} 3 & 2 \\ 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & -2 \\ 1 & -1 & 4 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 5 & 4 & 2 \\ 4 & 5 & -2 \\ 2 & -2 & 8 \end{pmatrix}.$$

结论：两种方法得到的结果完全一致。

4.3. 解：我们寻找直线  $y = a + bx$ . 对于给定的数据点  $(x_k, y_k)$ ，我们希望求解方程组：

$$\begin{cases} a - 2b = 4 \\ a - b = 3 \\ a + 0b = 1 \\ a + 2b = 0 \end{cases} \Rightarrow \underbrace{\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \end{pmatrix}}_A \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 4 \\ 3 \\ 1 \\ 0 \end{pmatrix}}_y.$$

建立正规方程  $A^T A \mathbf{x} = A^T \mathbf{y}$ ：

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix}.$$

$$A^T \mathbf{y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ -11 \end{pmatrix}.$$

求解  $\begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8 \\ -11 \end{pmatrix}$ . 由第一个方程： $b = 4a - 8$ . 代入第二个方程： $-a + 9(4a - 8) = -11 \Rightarrow 35a - 72 = -11 \Rightarrow 35a = 61 \Rightarrow a = 61/35$ .  $b = 4(61/35) - 8 = (244 - 280)/35 = -36/35$ . 所以最佳拟合直线为：

$$y = \frac{61}{35} - \frac{36}{35}x \quad (\approx 1.74 - 1.03x).$$

4.4. 解: **a)** 将点  $(x_k, y_k, z_k)$  代入方程  $a + bx_k + cy_k = z_k$ :

$$\begin{cases} a + 1b + 1c = 3 \\ a + 0b + 3c = 6 \\ a + 2b + 1c = 5 \\ a + 0b + 0c = 0 \end{cases}$$

写成矩阵形式  $A\mathbf{x} = \mathbf{z}$ :

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 3 \\ 6 \\ 5 \\ 0 \end{pmatrix}.$$

**b)** 求解正规方程  $A^T A \mathbf{x} = A^T \mathbf{z}$ .

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 0 \\ 1 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 5 \\ 3 & 5 & 3 \\ 5 & 3 & 11 \end{pmatrix}.$$

$$A^T \mathbf{z} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 0 \\ 1 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 26 \end{pmatrix}.$$

我们需要求解:

$$\begin{pmatrix} 4 & 3 & 5 \\ 3 & 5 & 3 \\ 5 & 3 & 11 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 26 \end{pmatrix}.$$

进行行约简或高斯消元: 1.  $R_2 \rightarrow 4R_2 - 3R_1: 0, 11, -3 | 10$ . 2.  $R_3 \rightarrow 4R_3 - 5R_1: 0, -3, 19 | 34$ .

解子系统  $\begin{cases} 11b - 3c = 10 \\ -3b + 19c = 34 \end{cases}$ .  $3 \times (Eq1) + 11 \times (Eq2): (33b - 9c) + (-33b + 209c) = 30 + 374$ .

$200c = 404 \implies c = 2.02$ .  $11b = 10 + 3(2.02) = 16.06 \implies b = 1.46$ . 代入  $R_1$ :  $4a + 3(1.46) + 5(2.02) = 14 \implies 4a + 14.48 = 14 \implies 4a = -0.48 \implies a = -0.12$ . 所以最佳拟合平面为:

$$z = -0.12 + 1.46x + 2.02y \quad \text{或者} \quad z = -\frac{6}{50} + \frac{73}{50}x + \frac{101}{50}y.$$

4.5. 证明: 我们知道  $\mathbb{R}^n$  可以分解为正交直和  $V = (\text{Ker } A)^\perp \oplus \text{Ker } A$ . 这意味着任何向量  $\mathbf{x}$  都可以唯一地写成  $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$ , 其中  $\mathbf{x}_r \in (\text{Ker } A)^\perp$ ,  $\mathbf{x}_n \in \text{Ker } A$ . 由勾股定理,  $\|\mathbf{x}\|^2 = \|\mathbf{x}_r\|^2 + \|\mathbf{x}_n\|^2$ .

假设  $\mathbf{x}$  是方程  $A\mathbf{x} = \mathbf{b}$  的一个解. 因为  $\mathbf{x}_n \in \text{Ker } A$ , 我们有  $A\mathbf{x} = A(\mathbf{x}_r + \mathbf{x}_n) = A\mathbf{x}_r + \mathbf{0} = A\mathbf{x}_r$ . 这意味着  $\mathbf{x}_r$  也是方程的一个解. 而且, 对于任何解  $\mathbf{x}$ , 其在  $(\text{Ker } A)^\perp$  上的投影  $\mathbf{x}_r$  是完全相同的.

**理由:** 如果  $\mathbf{x}$  和  $\mathbf{y}$  都是解, 则  $\mathbf{x} - \mathbf{y} \in \text{Ker } A$ . 这意味着它们的差在  $(\text{Ker } A)^\perp$  上的投影为  $\mathbf{0}$ , 即  $P_{(\text{Ker } A)^\perp}(\mathbf{x} - \mathbf{y}) = \mathbf{0} \implies P_{(\text{Ker } A)^\perp} \mathbf{x} = P_{(\text{Ker } A)^\perp} \mathbf{y}$ .

令  $\mathbf{x}_0 = \mathbf{x}_r = P_{(\text{Ker } A)^\perp} \mathbf{x}$ . 由上可知  $A\mathbf{x}_0 = \mathbf{b}$ . 对于任何解  $\mathbf{x}$ , 我们有  $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_n$ , 其中  $\mathbf{x}_n \in \text{Ker } A$ .

$$\|\mathbf{x}\|^2 = \|\mathbf{x}_0\|^2 + \|\mathbf{x}_n\|^2 \geq \|\mathbf{x}_0\|^2.$$

等号成立当且仅当  $\|\mathbf{x}_n\| = 0$ , 即  $\mathbf{x} = \mathbf{x}_0$ . 因此,  $\mathbf{x}_0$  是唯一的最小范数解, 且由投影  $\mathbf{x}_0 = P_{(\text{Ker } A)^\perp} \mathbf{x}$  给出.

4.6. 证明: 最小二乘解的定义是方程  $A\mathbf{x} = P_{\text{Ran } A} \mathbf{b}$  的解 (或者等价地, 正规方程  $A^T A \mathbf{x} = A^T \mathbf{b}$  的解). 令  $\hat{\mathbf{b}} = P_{\text{Ran } A} \mathbf{b}$ . 方程变为  $A\mathbf{x} = \hat{\mathbf{b}}$ . 由于投影总是位于列空间  $\text{Ran } A$  中, 该方程

**总是相容的** (即有解)。设  $S$  为所有最小二乘解的集合。由于方程有解, 我们可以直接应用习题 4.5 的结论:

1. 存在解  $\mathbf{x}_0 \in S$  使得  $\|\mathbf{x}_0\| \leq \|\mathbf{x}\|$  对所有  $\mathbf{x} \in S$  成立。
2. 这个解是唯一的, 并且由  $\mathbf{x}_0 = P_{(\text{Ker } A)^\perp} \mathbf{x}$  给出, 其中  $\mathbf{x}$  是任意一个最小二乘解。

**注:** 这就是著名的摩尔-彭罗斯伪逆的定义基础, 记为  $\mathbf{x}_0 = A^+ \mathbf{b}$ .

- 5.1. 证明: 回顾伴随矩阵的定义  $A^* = (\overline{A})^T$  (即先取共轭再转置, 或者先转置再取共轭)。利用行列式的性质:

1. 转置不改变行列式:  $\det(M^T) = \det(M)$ .
2. 共轭矩阵的行列式等于行列式的共轭:  $\det(\overline{M}) = \overline{\det(M)}$ . 因此,

$$\det(A^*) = \det((\overline{A})^T) = \det(\overline{A}) = \overline{\det(A)}.$$

得证。

- 5.2. 解: 首先分析矩阵  $A$  的秩和基本子空间。观察行向量:  $R_3 = R_1 + R_2$  ( $1+1=2, 1+3=4, 1+2=3$ )。观察列向量:  $C_3$  不是  $C_1, C_2$  的简单倍数, 且前两列线性无关。因此,  $\text{rank } A = 2$ 。

我们需要计算四个投影矩阵:  $P_{\text{Ran } A}, P_{\text{Ker } A^*}, P_{\text{Ran } A^*}, P_{\text{Ker } A}$ . 利用互补关系:  $P_{\text{Ran } A} + P_{\text{Ker } A^*} = I$  且  $P_{\text{Ran } A^*} + P_{\text{Ker } A} = I$ . 策略是先计算维数为 1 的子空间的投影 (计算量小), 然后利用互补关系求另一个。由于  $\text{rank } A = 2$  且  $A$  是  $3 \times 3$  矩阵, 零空间  $\text{Ker } A$  和左零空间  $\text{Ker } A^*$  的维数均为  $3 - 2 = 1$ 。

1. 计算  $P_{\text{Ker } A}$  求解  $A\mathbf{x} = \mathbf{0}$ :

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

$R_2 - R_1 \implies 2x_2 + x_3 = 0 \implies x_3 = -2x_2$ . 代入  $R_1$ :  $x_1 + x_2 - 2x_2 = 0 \implies x_1 = x_2$ . 取  $x_2 = 1$ , 得基向量  $\mathbf{v} = (1, 1, -2)^T$ .  $\|\mathbf{v}\|^2 = 1 + 1 + 4 = 6$ .

$$P_{\text{Ker } A} = \frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2} = \frac{1}{6} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} (1 \ 1 \ -2) = \frac{1}{6} \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{pmatrix}.$$

2. 计算  $P_{\text{Ran } A^*}$  利用  $P_{\text{Ran } A^*} = I - P_{\text{Ker } A}$ :

$$P_{\text{Ran } A^*} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 5 & -1 & 2 \\ -1 & 5 & 2 \\ 2 & 2 & 2 \end{pmatrix}.$$

3. 计算  $P_{\text{Ker } A^*}$   $\text{Ker } A^* = (\text{Ran } A)^\perp$ . 我们需要找到垂直于  $A$  的列空间的向量。 $A$  的列空间由  $\mathbf{c}_1 = (1, 1, 2)^T$  和  $\mathbf{c}_2 = (1, 3, 4)^T$  生成。求叉积  $\mathbf{n} = \mathbf{c}_1 \times \mathbf{c}_2$ :

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 2 \\ 1 & 3 & 4 \end{vmatrix} = (4 - 6, 2 - 4, 3 - 1)^T = (-2, -2, 2)^T.$$

取更简单的基向量  $\mathbf{u} = (1, 1, -1)^T$ . 验证:  $1(1) + 1(1) + 2(-1) = 0, 1(1) + 3(1) + 4(-1) = 0$ . 正确。  $\|\mathbf{u}\|^2 = 1 + 1 + 1 = 3$ .

$$P_{\text{Ker } A^*} = \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} (1 \ 1 \ -1) = \frac{1}{3} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

4. 计算  $P_{\text{Ran } A}$  利用  $P_{\text{Ran } A} = I - P_{\text{Ker } A^*}$ :

$$P_{\text{Ran } A} = I - \frac{1}{3} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 3-1 & -1 & 1 \\ -1 & 3-1 & 1 \\ 1 & 1 & 3-1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

5.3. 证明: 1.  $\text{Ker } A \subseteq \text{Ker}(A^*A)$  假设  $\mathbf{x} \in \text{Ker } A$ , 即  $A\mathbf{x} = \mathbf{0}$ . 那么  $A^*A\mathbf{x} = A^*(A\mathbf{x}) = A^*\mathbf{0} = \mathbf{0}$ . 所以  $\mathbf{x} \in \text{Ker}(A^*A)$ . 这部分是平凡的。

2.  $\text{Ker}(A^*A) \subseteq \text{Ker } A$  假设  $\mathbf{x} \in \text{Ker}(A^*A)$ , 即  $A^*A\mathbf{x} = \mathbf{0}$ . 考虑  $A\mathbf{x}$  的范数平方:

$$\|A\mathbf{x}\|^2 = (A\mathbf{x}, A\mathbf{x}) = (A^*(A\mathbf{x}), \mathbf{x}) = (A^*A\mathbf{x}, \mathbf{x}).$$

将假设  $A^*A\mathbf{x} = \mathbf{0}$  代入:

$$\|A\mathbf{x}\|^2 = (\mathbf{0}, \mathbf{x}) = 0.$$

由于只有零向量的范数为 0, 这蕴含  $A\mathbf{x} = \mathbf{0}$ , 即  $\mathbf{x} \in \text{Ker } A$ . 得证。

5.4. 证明: a) 对于任何矩阵  $M$  (这里设定定义域维数为  $n$ ), 根据秩-零化度定理 (Rank-Nullity Theorem), 有  $\text{rank } M + \dim(\text{Ker } M) = n$ . 因为  $A$  是  $m \times n$  矩阵,  $A^*A$  是  $n \times n$  矩阵, 它们的定义域都是  $\mathbb{F}^n$ . 由 5.3 可知  $\text{Ker } A = \text{Ker}(A^*A)$ , 所以  $\dim(\text{Ker } A) = \dim(\text{Ker}(A^*A))$ . 因此:

$$\text{rank } A = n - \dim(\text{Ker } A) = n - \dim(\text{Ker}(A^*A)) = \text{rank}(A^*A).$$

b) 如果  $A\mathbf{x} = \mathbf{0}$  只有平凡解, 意味着  $\text{Ker } A = \{\mathbf{0}\}$ . 由 5.3,  $\text{Ker}(A^*A) = \{\mathbf{0}\}$ . 由于  $A^*A$  是  $n \times n$  方阵, 且只有平凡的核, 这意味着  $A^*A$  是可逆的。在方程  $A\mathbf{x} = \mathbf{b}$  两边左乘  $A^*$  得  $A^*A\mathbf{x} = A^*\mathbf{b}$ . 两边左乘  $(A^*A)^{-1}$  得:

$$(A^*A)^{-1}A^*A\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b} \implies I\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}.$$

因此, 矩阵  $L = (A^*A)^{-1}A^*$  满足  $LA = I$ , 即  $L$  是  $A$  的左逆。

5.5. 解: 题目假设  $A^*A$  可逆, 这意味着  $A$  具有满列秩 ( $\text{rank } A = n$ ), 即  $\text{Ker } A = \{\mathbf{0}\}$ .

1. 到  $\text{Ker } A$  的投影  $P_{\text{Ker } A}$ : 由于  $\text{Ker } A = \{\mathbf{0}\}$ , 这只是零空间。

$$P_{\text{Ker } A} = \mathbf{0} \quad (\text{零矩阵})$$

2. 到  $\text{Ran } A^*$  的投影  $P_{\text{Ran } A^*}$ : 回想  $P_{\text{Ran } A^*} = I - P_{\text{Ker } A}$  (因为  $\text{Ran } A^* = (\text{Ker } A)^\perp$ ).

$$P_{\text{Ran } A^*} = I - \mathbf{0} = I \quad (\text{单位矩阵})$$

(注: 这是有道理的, 因为如果  $A$  列满秩, 其行空间  $\text{Ran } A^*$  就是整个  $\mathbb{F}^n$ ).

3. 到  $\text{Ker } A^*$  的投影  $P_{\text{Ker } A^*}$ : 回想  $P_{\text{Ker } A^*} = I - P_{\text{Ran } A}$  (因为  $\text{Ker } A^* = (\text{Ran } A)^\perp$ ). 利用题目给出的  $P_{\text{Ran } A}$  公式:

$$P_{\text{Ker } A^*} = I - A(A^*A)^{-1}A^*$$

5.6. 证明: 我们需要证明  $P$  是到  $\text{Ran } P$  沿着  $(\text{Ran } P)^\perp$  的投影。对于任意  $\mathbf{x}$ , 我们可以正交分解为  $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ , 其中  $\mathbf{x}_1 \in \text{Ran } P$ ,  $\mathbf{x}_2 \in (\text{Ran } P)^\perp$ . 我们要证明  $P\mathbf{x} = \mathbf{x}_1$  (这等价于  $P\mathbf{x}_1 = \mathbf{x}_1$  且  $P\mathbf{x}_2 = \mathbf{0}$ ).

1. 证明  $P\mathbf{x}_1 = \mathbf{x}_1$  由于  $\mathbf{x}_1 \in \text{Ran } P$ , 存在  $\mathbf{y}$  使得  $\mathbf{x}_1 = P\mathbf{y}$ . 应用  $P$ :

$$P\mathbf{x}_1 = P(P\mathbf{y}) = P^2\mathbf{y}.$$

利用性质  $P^2 = P$ , 我们有  $P^2\mathbf{y} = P\mathbf{y} = \mathbf{x}_1$ . 所以  $P\mathbf{x}_1 = \mathbf{x}_1$ .

2. 证明  $P\mathbf{x}_2 = \mathbf{0}$  已知  $\mathbf{x}_2 \perp \text{Ran } P$ , 即对于任何  $\mathbf{z} \in \text{Ran } P$ , 有  $(\mathbf{z}, \mathbf{x}_2) = 0$ . 我们想计算  $P\mathbf{x}_2$ . 考虑它与任意向量  $\mathbf{y}$  的内积  $(\mathbf{y}, P\mathbf{x}_2)$ . 利用伴随性质:

$$(\mathbf{y}, P\mathbf{x}_2) = (P^*\mathbf{y}, \mathbf{x}_2).$$

利用自伴随性  $P^* = P$ :

$$(P^*\mathbf{y}, \mathbf{x}_2) = (P\mathbf{y}, \mathbf{x}_2).$$

观察  $P\mathbf{y}$ , 根据定义它属于  $\text{Ran } P$ . 因为  $\mathbf{x}_2 \perp \text{Ran } P$ , 所以  $(P\mathbf{y}, \mathbf{x}_2) = 0$ . 既然对于所有  $\mathbf{y}$ ,  $(\mathbf{y}, P\mathbf{x}_2) = 0$ , 则必须有  $P\mathbf{x}_2 = \mathbf{0}$ .

综上所述,  $P(\mathbf{x}_1 + \mathbf{x}_2) = P\mathbf{x}_1 + P\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{0} = \mathbf{x}_1$ . 这正是正交投影的定义。

6.1. 解: 1.  $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$  这是一个实对称矩阵, 特征值必为实数。计算特征多项式:

$$\det(A - \lambda I) = (1 - \lambda)^2 - 4 = 0 \implies 1 - \lambda = \pm 2 \implies \lambda_1 = 3, \lambda_2 = -1.$$



求解特征向量: 对于  $\lambda_1 = 3$ :  $\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 = x_2$ . 单位化得  $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ . 对于  $\lambda_2 = -1$ :  $\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 = -x_2$ . 单位化得  $\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ . 所以:

$$D = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}, \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

2.  $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  这是一个实反对称矩阵 (正规矩阵), 特征值是纯虚数.

$$\det(A - \lambda I) = \lambda^2 + 1 = 0 \implies \lambda = \pm i.$$

对于  $\lambda_1 = i$ :  $\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 = ix_2$ . 取  $\mathbf{v} = (i, 1)^T$ .  $\|\mathbf{v}\| = \sqrt{1+1} = \sqrt{2}$ . 单位化得  $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}$ . 对于  $\lambda_2 = -i$ : 由于矩阵是实矩阵, 特征向量是  $\mathbf{u}_1$  的共轭, 即  $\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix}$ . 所以:

$$D = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}.$$

3.  $A = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}$  实对称矩阵. 特征多项式:

$$\det(A - \lambda I) = -\lambda^3 + 3(4)\lambda + 2(2)(2)(2) = -\lambda^3 + 12\lambda + 16 = 0.$$

试根发现  $\lambda = 4$  是根 ( $-64 + 48 + 16 = 0$ ). 多项式分解:  $-(\lambda - 4)(\lambda + 2)^2 = 0$ . 特征值:  $\lambda_1 = 4$

(单重),  $\lambda_2 = -2$  (二重). 对于  $\lambda_1 = 4$ :  $\begin{pmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{pmatrix} \mathbf{x} = \mathbf{0}$ . 容易看出  $\mathbf{x} = (1, 1, 1)^T$ .

单位化  $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ . 对于  $\lambda_2 = -2$ :  $\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 + x_2 + x_3 = 0$ .

这是垂直于  $\mathbf{u}_1$  的平面. 我们需要在其中找一组标准正交基. 取  $\mathbf{v}_2 = (-1, 1, 0)^T$ . 单位化  $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(-1, 1, 0)^T$ . 取  $\mathbf{v}_3 = \mathbf{u}_1 \times \mathbf{u}_2$  或直接找垂直于  $\mathbf{v}_2$  和  $\mathbf{u}_1$  的向量.  $(1, 1, 1) \times (-1, 1, 0) = (-1, -1, 2)^T$ . 单位化  $\mathbf{u}_3 = \frac{1}{\sqrt{6}}(-1, -1, 2)^T$ . 所以:

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad U = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}.$$

6.2. 正确. 这是谱定理 (Spectral Theorem) 对于正规矩阵 (Normal Matrices) 的表述. 如果  $A = UDU^*$ , 其中  $U$  是酉矩阵, 那么  $U$  的列向量构成了  $A$  的特征向量的标准正交基 (当然也是正交基). 反之亦然.

6.3. 证明: 实数情况 (假设  $A = A^*$ , 即自伴随): 计算右边第一项:  $(A(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) = (A\mathbf{x} + A\mathbf{y}, \mathbf{x} + \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) + (A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{x}) + (A\mathbf{y}, \mathbf{y})$ . 由于是实内积且  $A$  自伴随,  $(A\mathbf{y}, \mathbf{x}) = (\mathbf{y}, A^*\mathbf{x}) = (\mathbf{y}, A\mathbf{x}) = (A\mathbf{x}, \mathbf{y})$ . 所以  $(A(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) + 2(A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{y})$ . 同理,  $(A(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) - 2(A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{y})$ . 两式相减得  $4(A\mathbf{x}, \mathbf{y})$ . 除以 4 即得证.

复数情况: 记  $q(\mathbf{v}) = (A\mathbf{v}, \mathbf{v})$ . 右边求和为  $\sum_{\alpha \in \{1, -1, i, -i\}} \alpha q(\mathbf{x} + \alpha \mathbf{y})$ . 展开  $q(\mathbf{x} + \alpha \mathbf{y}) = (A\mathbf{x} + \alpha A\mathbf{y}, \mathbf{x} + \alpha \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) + \bar{\alpha}(A\mathbf{x}, \mathbf{y}) + \alpha(A\mathbf{y}, \mathbf{x}) + \alpha\bar{\alpha}(A\mathbf{y}, \mathbf{y})$ . 注意到  $\alpha\bar{\alpha} = |\alpha|^2 = 1$ .  $q(\mathbf{x} + \alpha \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) + (A\mathbf{y}, \mathbf{y}) + \bar{\alpha}(A\mathbf{x}, \mathbf{y}) + \alpha(A\mathbf{y}, \mathbf{x})$ . 现在对  $\alpha$  求和  $\alpha q(\dots)$ : 项  $\alpha[(A\mathbf{x}, \mathbf{x}) + (A\mathbf{y}, \mathbf{y})]$ : 和是  $(1 - 1 + i - i)[\dots] = 0$ . 项  $\alpha\bar{\alpha}(A\mathbf{x}, \mathbf{y}) = 1 \cdot (A\mathbf{x}, \mathbf{y})$ : 和是

$(1+1+1+1)(\mathbf{Ax}, \mathbf{y}) = 4(\mathbf{Ax}, \mathbf{y})$ . 项  $\alpha^2(\mathbf{Ay}, \mathbf{x})$ : 和是  $(1+1+(-1)+(-1))(\mathbf{Ay}, \mathbf{x}) = 0$ . 所以总和为  $4(\mathbf{Ax}, \mathbf{y})$ . 除以 4 得证.

6.4. 证明: 设  $U$  和  $V$  是酉矩阵, 即  $U^*U = I$  且  $V^*V = I$ . 考虑乘积  $UV$ .

$$(UV)^*(UV) = (V^*U^*)(UV) = V^*(U^*U)V = V^*IV = V^*V = I.$$

因此  $UV$  也是酉矩阵. 对于实正交矩阵, 证明完全相同 (将  $*$  替换为  $T$ ).

6.5. 解: **a) 正确**. 由极化恒等式 (习题 6.3), 内积可以用范数完全表示. 如果范数保持不变, 则内积保持不变, 即  $(U\mathbf{x}, U\mathbf{y}) = (\mathbf{x}, \mathbf{y})$ . 这蕴含  $U^*U = I$ . 由于是有限维空间,  $U^*U = I$  意味着  $U$  是可逆的且  $U^{-1} = U^*$ , 所以  $U$  是酉算子.

**b) 错误**. 反例: 在  $\mathbb{R}^2$  中, 设标准基  $\mathbf{e}_1, \mathbf{e}_2$ . 定义  $U\mathbf{e}_1 = \mathbf{e}_1$  且  $U\mathbf{e}_2 = \mathbf{e}_1$ . 显然  $\|U\mathbf{e}_1\| = 1 = \|\mathbf{e}_1\|$  且  $\|U\mathbf{e}_2\| = 1 = \|\mathbf{e}_2\|$ . 但是  $U$  甚至不可逆 (秩为 1), 更不用说酉算子了. 而且,  $\|U(\mathbf{e}_1 + \mathbf{e}_2)\| = \|2\mathbf{e}_1\| = 2$ , 而  $\|\mathbf{e}_1 + \mathbf{e}_2\| = \sqrt{2}$ . 范数未保持.

6.6. 证明: **a)** 设  $B = UAU^*$ , 其中  $U$  是酉矩阵 ( $U^*U = UU^* = I$ ).

$$B^*B = (UAU^*)^*(UAU^*) = (U^*A^*U)(UAU^*) = U^*A^*(U^*U)AU^* = U^*(A^*A)U^*.$$

这意味着  $B^*B$  与  $A^*A$  相似. 因为相似矩阵具有相同的迹 ( $\text{trace}(XY) = \text{trace}(YX)$ ):

$$\text{trace}(B^*B) = \text{trace}(U(A^*A)U^*) = \text{trace}((A^*A)U^*U) = \text{trace}(A^*A).$$

**b)** 计算  $M^*M$  的迹.  $(M^*M)_{jj} = \sum_{k=1}^n (M^*)_{jk} M_{kj} = \sum_{k=1}^n \overline{M_{kj}} M_{kj} = \sum_{k=1}^n |M_{kj}|^2$ . 因此,

$$\text{trace}(M^*M) = \sum_{j=1}^n (M^*M)_{jj} = \sum_{j=1}^n \sum_{k=1}^n |M_{kj}|^2.$$

结合 **a)** 的结论, 得证  $\sum_{j,k} |A_{j,k}|^2 = \sum_{j,k} |B_{j,k}|^2$ . (注: 这个量被称为 Frobenius 范数的平方, 记为  $\|A\|_F^2$ ).

**c)** 对于矩阵  $A = \begin{pmatrix} 1 & 2 \\ 2 & i \end{pmatrix}$ :

$$\|A\|_F^2 = |1|^2 + |2|^2 + |2|^2 + |i|^2 = 1 + 4 + 4 + 1 = 10.$$

对于矩阵  $B = \begin{pmatrix} i & 4 \\ 1 & 1 \end{pmatrix}$ :

$$\|B\|_F^2 = |i|^2 + |4|^2 + |1|^2 + |1|^2 = 1 + 16 + 1 + 1 = 19.$$

因为  $10 \neq 19$ , 所以这两个矩阵不是酉等价的.

6.7. 解: **a) 不酉等价**. 迹不同 ( $2 \neq 0$ ).

**b) 不酉等价**. 行列式不同 ( $-1 \neq -1/4$ ).

**c) 不酉等价**. 特征值不同. 第一个矩阵特征值是  $1, i, -i$ ; 第二个是  $2, -1, 0$ .

**d) 是酉等价的**. 第一个矩阵  $A$  是正规矩阵 (反对称块 + 对角), 特征值为  $1, \pm i$ . 第二个矩阵  $B$  是对角矩阵, 特征值为  $1, \pm i$ . 因为两个都是正规矩阵且具有相同的特征值, 由谱定理, 它们都酉等价于同一个对角矩阵  $D = \text{diag}(1, i, -i)$ .

**e) 不酉等价**. 虽然它们具有相同的特征值 ( $1, 2, 3$ ) 和相同的迹、行列式, 但 Frobenius 范数不同. 第一个矩阵 (上三角):  $\|A\|_F^2 = 1^2 + 1^2 + 2^2 + 2^2 + 3^2 = 1 + 1 + 4 + 4 + 9 = 19$ . 第二个矩阵 (对角):  $\|B\|_F^2 = 1^2 + 2^2 + 3^2 = 14$ .  $19 \neq 14$ , 由 6.6 题结论, 不酉等价.

6.8. 证明: 设  $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . 因为  $U$  是正交的 ( $U^T U = I$ ), 列向量是标准正交的. 1.  $a^2 + c^2 = 1$ .

存在  $\alpha$  使得  $a = \cos \alpha, c = \sin \alpha$ . 2.  $b^2 + d^2 = 1$ . 3.  $ab + cd = 0$ . 由  $ab + cd = 0$  和  $a, c$  的形式, 向量  $(b, d)^T$  必须垂直于  $(\cos \alpha, \sin \alpha)^T$ . 所以  $(b, d)^T$  只能是  $(-\sin \alpha, \cos \alpha)^T$  或者  $(\sin \alpha, -\cos \alpha)^T$ . 计算行列式:  $\det U = ad - bc$ . 如果是第一种情况:  $\cos^2 \alpha - (-\sin^2 \alpha) = 1$ . 如果是第二种情况:  $-\cos^2 \alpha - \sin^2 \alpha = -1$ . 题目已知  $\det U = 1$ , 所以必须是第一种情况. 即  $U = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$ , 这就是旋转矩阵.

6.9. 证明: **a)** 实矩阵的特征多项式是实系数的 3 次多项式, 根据介值定理, 至少有一个实根. 设特征值为  $\lambda_1, \lambda_2, \lambda_3$ . 由于  $U$  是正交的, 所有特征值的模长为 1. 可能的实特征值只能是 1 或 -1. 复特征值成对共轭出现, 设为  $e^{i\theta}, e^{-i\theta}$ , 它们的积为 1. 因为  $\det U = \lambda_1 \lambda_2 \lambda_3 = 1$ . 情况 1: 三个实根. 可能是 1, 1, 1 或 1, -1, -1. 都有 1. 情况 2: 一个实根, 两个复根.  $\lambda_1 \cdot (e^{i\theta} e^{-i\theta}) = 1 \implies \lambda_1 = 1$ . 所以 1 必然是特征值.

**b)** 设  $\mathbf{v}_1$  是对应  $\lambda = 1$  的单位特征向量. 扩充为标准正交基  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ . 在这些基下, 算子  $U$  的矩阵表示的第一列由  $U\mathbf{v}_1 = 1\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3$  决定, 即  $(1, 0, 0)^T$ . 因为  $U$  是酉算子 (在实数域为正交算子), 其矩阵表示也是正交矩阵. 第一列的范数为 1, 且与其他列正交. 这意味着第一行必须是  $(1, 0, 0)$  (因为第一行第一列是 1, 为了保持第一列范数为 1, 其余元素为 0; 且为了保持第一行范数为 1, 第一行其余元素也必须为 0). 更严格地说: 因为是正交矩阵, 行也是标准正交的. 第一行必须有范数 1, 第一元素已经是 1, 所以其他必须是 0). 矩阵形式为:

$$[U] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & u_{22} & u_{23} \\ 0 & u_{32} & u_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & U' \end{pmatrix}.$$

由于  $\det[U] = 1 \cdot \det U' = 1$ , 所以  $\det U' = 1$ . 同时  $U'$  是  $2 \times 2$  的正交矩阵 (作为正交矩阵的子块). 根据习题 6.8, 行列式为 1 的  $2 \times 2$  正交矩阵必然是旋转矩阵. 得证.

8.1. 证明: 我们计算总和中每一项  $z_k \bar{w}_k$  的实部.

$$z_k \bar{w}_k = (x_k + iy_k)(u_k - iv_k) = (x_k u_k + y_k v_k) + i(y_k u_k - x_k v_k).$$

取实部:

$$\operatorname{Re}(z_k \bar{w}_k) = x_k u_k + y_k v_k.$$

因此,

$$\operatorname{Re} \left( \sum_{k=1}^n z_k \bar{w}_k \right) = \sum_{k=1}^n \operatorname{Re}(z_k \bar{w}_k) = \sum_{k=1}^n (x_k u_k + y_k v_k) = \sum_{k=1}^n x_k u_k + \sum_{k=1}^n y_k v_k.$$

这正是  $\mathbb{R}^{2n}$  中对应向量的标准内积.

8.2. 证明: 我们需要验证实内积的三个公理: 对称性、线性和正定性. 设  $(\mathbf{x}, \mathbf{y})_{\mathbb{R}} := \operatorname{Re}(\mathbf{x}, \mathbf{y})_{\mathbb{C}}$ .

1. **对称性:** 由于复内积满足共轭对称性  $(\mathbf{y}, \mathbf{x})_{\mathbb{C}} = \overline{(\mathbf{x}, \mathbf{y})_{\mathbb{C}}}$ . 一个复数的共轭的实部等于该复数本身的实部 ( $\operatorname{Re}(\bar{z}) = \operatorname{Re}(z)$ ). 所以,  $(\mathbf{y}, \mathbf{x})_{\mathbb{R}} = \operatorname{Re}(\mathbf{y}, \mathbf{x})_{\mathbb{C}} = \operatorname{Re}(\overline{(\mathbf{x}, \mathbf{y})_{\mathbb{C}}}) = \operatorname{Re}(\mathbf{x}, \mathbf{y})_{\mathbb{C}} = (\mathbf{x}, \mathbf{y})_{\mathbb{R}}$ .

2. **线性** (对实标量): 对于  $\alpha, \beta \in \mathbb{R}$  和  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ :

$$(\alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z})_{\mathbb{R}} = \operatorname{Re}(\alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z})_{\mathbb{C}}$$

利用复内积的线性:

$$= \operatorname{Re}(\alpha(\mathbf{x}, \mathbf{z})_{\mathbb{C}} + \beta(\mathbf{y}, \mathbf{z})_{\mathbb{C}})$$

由于  $\alpha, \beta$  是实数, 可以提出来:

$$= \alpha \operatorname{Re}(\mathbf{x}, \mathbf{z})_{\mathbb{C}} + \beta \operatorname{Re}(\mathbf{y}, \mathbf{z})_{\mathbb{C}} = \alpha(\mathbf{x}, \mathbf{z})_{\mathbb{R}} + \beta(\mathbf{y}, \mathbf{z})_{\mathbb{R}}.$$

3. **正定性:**

$$(\mathbf{x}, \mathbf{x})_{\mathbb{R}} = \operatorname{Re}(\mathbf{x}, \mathbf{x})_{\mathbb{C}} = \operatorname{Re}(\|\mathbf{x}\|_{\mathbb{C}}^2) = \|\mathbf{x}\|_{\mathbb{C}}^2.$$

因为复范数平方是非负实数. 所以  $(\mathbf{x}, \mathbf{x})_{\mathbb{R}} \geq 0$ , 且等号成立当且仅当  $\mathbf{x} = \mathbf{0}$ .

8.3. 证明: 我们需要证明  $(U\mathbf{x}, \mathbf{x}) = 0$ . 由于  $U$  是正交变换, 它保内积, 即  $(U\mathbf{u}, U\mathbf{v}) = (\mathbf{u}, \mathbf{v})$ . 令  $\mathbf{u} = U\mathbf{x}, \mathbf{v} = \mathbf{x}$ , 则:

$$(U\mathbf{x}, \mathbf{x}) = (U(U\mathbf{x}), U\mathbf{x}) = (U^2\mathbf{x}, U\mathbf{x}).$$

利用性质  $U^2 = -I$ , 我们有  $U^2\mathbf{x} = -\mathbf{x}$ . 所以:

$$(U\mathbf{x}, \mathbf{x}) = (-\mathbf{x}, U\mathbf{x}) = -(\mathbf{x}, U\mathbf{x}).$$

在实内积空间中, 内积是对称的, 即  $(\mathbf{x}, U\mathbf{x}) = (U\mathbf{x}, \mathbf{x})$ . 因此我们得到:

$$(U\mathbf{x}, \mathbf{x}) = -(U\mathbf{x}, \mathbf{x}) \implies 2(U\mathbf{x}, \mathbf{x}) = 0 \implies (U\mathbf{x}, \mathbf{x}) = 0.$$

得证。

8.4. 证明: 由于  $U$  是正交的, 根据定义有  $U^*U = I$ , 这意味着  $U^* = U^{-1}$ . 由假设  $U^2 = -I$ , 两边同时右乘  $U^{-1}$  (或左乘), 得到  $U = -U^{-1}$ . 即  $U^{-1} = -U$ . 结合上面两式, 得到  $U^* = -U$ . (这说明  $U$  是反对称的/斜伴随的)。

8.5. 证明: **1. 证明维数为偶数** 设  $X$  的维数为  $d$ .  $U$  的行列式满足  $\det(U^2) = \det(-I) = (-1)^d$ . 另一方面,  $\det(U^2) = (\det U)^2$ . 由于  $U$  是实矩阵, 其行列式是实数, 所以  $(\det U)^2 \geq 0$ . 因此  $(-1)^d \geq 0$ , 这意味  $d$  必须是偶数. 设  $\dim X = 2n$ .

**2. 构造子空间  $E$  和分解** 我们可以使用归纳法或直接构造. 任取非零向量  $\mathbf{v}_1 \in X$ . 令  $\mathbf{w}_1 = U\mathbf{v}_1$ . 由习题 8.3 知  $\mathbf{v}_1 \perp \mathbf{w}_1$ . 且  $\|\mathbf{w}_1\| = \|\mathbf{v}_1\|$  (正交变换保范数). 考虑子空间  $V_1 = \text{span}\{\mathbf{v}_1, \mathbf{w}_1\}$ .  $V_1$  是二维的且在  $U$  下是不变的 (因为  $U\mathbf{v}_1 = \mathbf{w}_1, U\mathbf{w}_1 = U^2\mathbf{v}_1 = -\mathbf{v}_1 \in V_1$ ). 由于  $U$  是正交变换, 它将  $V_1^\perp$  映射到  $V_1^\perp$ . 我们可以继续在  $V_1^\perp$  中取向量  $\mathbf{v}_2$ , 构造  $V_2 = \text{span}\{\mathbf{v}_2, \mathbf{w}_2\}$ , 以此类推. 最终我们得到一组正交基  $\{\mathbf{v}_1, \mathbf{w}_1, \mathbf{v}_2, \mathbf{w}_2, \dots, \mathbf{v}_n, \mathbf{w}_n\}$ , 其中  $\mathbf{w}_k = U\mathbf{v}_k$ .

定义  $E = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ . 显然  $\dim E = n$ . 定义  $E^\perp$ . 由于  $\mathbf{w}_k \perp \mathbf{v}_j$  (对所有  $j, k$ ),  $E^\perp$  实际上由  $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  生成.

**3. 确定块矩阵形式** 我们将  $X$  分解为  $X = E \oplus E^\perp$ . 对于任意  $\mathbf{x} \in E$ ,  $\mathbf{x} = \sum \alpha_k \mathbf{v}_k$ .  $U\mathbf{x} = \sum \alpha_k U\mathbf{v}_k = \sum \alpha_k \mathbf{w}_k \in E^\perp$ . 定义  $U_0: E \rightarrow E^\perp$  为  $U$  在  $E$  上的限制, 即  $U_0\mathbf{x} = U\mathbf{x}$ .

这是一个正交变换 (因为  $U$  保范数, 且  $E, E^\perp$  维数相同). 因此, 在块矩阵  $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$  中,

$A$  表示  $E \rightarrow E$  的映射 (为  $\mathbf{0}$ ),  $C$  表示  $E \rightarrow E^\perp$  的映射 (为  $U_0$ ).

接下来考虑  $\mathbf{y} \in E^\perp$ .  $\mathbf{y} = \sum \beta_k \mathbf{w}_k = \sum \beta_k U\mathbf{v}_k$ .  $U\mathbf{y} = \sum \beta_k U^2\mathbf{v}_k = \sum \beta_k (-\mathbf{v}_k) = -\sum \beta_k \mathbf{v}_k \in E$ . 这说明  $D (E^\perp \rightarrow E^\perp)$  为  $\mathbf{0}$ . 我们只需确定  $B (E^\perp \rightarrow E)$ . 由习题 8.4, 我们知道  $U^* = -U$ . 块矩阵的伴随是转置并共轭 (实数为转置):

$$U^* = \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix} = \begin{pmatrix} \mathbf{0} & U_0^T \\ B^T & \mathbf{0} \end{pmatrix}.$$

我们要  $U^* = -U$ , 即:

$$\begin{pmatrix} \mathbf{0} & U_0^T \\ B^T & \mathbf{0} \end{pmatrix} = -\begin{pmatrix} \mathbf{0} & B \\ U_0 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & -B \\ -U_0 & \mathbf{0} \end{pmatrix}.$$

这给出了  $B^T = -U_0 \implies B = -U_0^T = -U_0^*$  (实矩阵下). 所以  $U$  的形式为:

$$U = \begin{pmatrix} \mathbf{0} & -U_0^* \\ U_0 & \mathbf{0} \end{pmatrix}.$$

得证。

## 第六章习题解答

1.1. 证明: 根据算子的上三角表示定理 (定理 1.1), 对于任何  $n \times n$  矩阵  $A$  (或算子), 存在一个酉矩阵  $U$  和一个上三角矩阵  $T$ , 使得  $A = UTU^*$ . 这意味着  $A$  与  $T$  是相似的 (因为  $U^* = U^{-1}$ ).

**1. 特征值的一致性** 首先, 我们需要确认  $T$  的对角线元素就是  $A$  的特征值. 计算  $T$  的特征多项式:

$$\det(T - \lambda I) = \prod_{k=1}^n (T_{kk} - \lambda).$$

因为  $T - \lambda I$  仍然是上三角矩阵, 其行列式是对角线元素的乘积。这意味着  $T$  的特征值正是其对角线元素  $T_{11}, T_{22}, \dots, T_{nn}$ . 由于相似矩阵具有相同的特征多项式, 因此  $A$  的特征值 (计入代数重数) 正是  $T$  的对角线元素。我们记  $\lambda_k = T_{kk}$ .

**2. 行列式** 利用行列式的乘法性质:

$$\det(A) = \det(UTU^*) = \det(U) \det(T) \det(U^*) = \det(T) \det(UU^*) = \det(T) \det(I) = \det(T).$$

对于上三角矩阵  $T$ , 其行列式等于对角线元素的乘积:

$$\det(T) = \prod_{k=1}^n T_{kk} = \prod_{k=1}^n \lambda_k.$$

因此,  $\det(A) = \prod_{k=1}^n \lambda_k$ .

**3. 迹** 利用迹的循环性质  $\text{trace}(XY) = \text{trace}(YX)$ :

$$\text{trace}(A) = \text{trace}(U(TU^*)) = \text{trace}((TU^*)U) = \text{trace}(T(U^*U)) = \text{trace}(TI) = \text{trace}(T).$$

对于任何矩阵, 迹定义为对角线元素之和:

$$\text{trace}(T) = \sum_{k=1}^n T_{kk} = \sum_{k=1}^n \lambda_k.$$

因此,  $\text{trace}(A) = \sum_{k=1}^n \lambda_k$ . 得证。

2.1. 解: **a) 正确**. 酉算子满足  $U^*U = UU^* = I$ , 这符合正规算子  $U^*U = UU^*$  的定义。

**b) 错误**. 酉矩阵必然可逆 (逆为  $U^*$ ), 但可逆矩阵不一定是酉矩阵 (例如  $2I$  是可逆的但不是酉的)。

**c) 正确**. 如果  $B = UAU^*$  且  $U$  是酉的, 由于  $U^* = U^{-1}$ , 则  $B = UAU^{-1}$ , 即它们是相似的。

**d) 正确**.  $(A+B)^* = A^* + B^* = A + B$ .

**e) 正确**. 若  $U$  是酉的, 则  $U^* = U^{-1}$ .  $(U^*)^*(U^*) = UU^{-1} = I$ , 所以  $U^*$  也是酉的。

**f) 正确**. 设  $N$  是正规的 ( $N^*N = NN^*$ ). 我们需要检查  $N^*$ .  $(N^*)^*(N^*) = NN^*$  且  $(N^*)(N^*)^* = N^*N$ . 由于  $N$  正规, 这两者相等, 所以  $N^*$  正规。

**g) 错误**. 反例:  $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ . 特征值全是 1, 但  $A^*A \neq I$ , 不是酉的。

**h) 正确**. 正规算子可以被酉对角化。即  $A = UDU^*$ . 如果特征值全是 1, 则  $D = I$ . 于是  $A = UIU^* = UU^* = I$ .

**i) 错误**. 由极化恒等式 (复数域或实数域), 内积可以完全由范数确定。如果保持范数, 必然保持内积。

2.2. 解: **错误**. 反例: 设  $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$  和  $B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ .  $A$  是反对称的 (实数下即斜伴随),

$B$  是对称的 (自伴随)。两者都是正规矩阵。它们的和  $S = A + B = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$ . 计算  $SS^* = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$ . 计算  $S^*S = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$ .  $SS^* \neq S^*S$ , 所以和不是正规的。

2.3. 证明: 设  $A = UDU^*$ , 其中  $U$  是酉矩阵,  $D$  是对角矩阵。我们需要证明  $A^*A = AA^*$ .

$$A^*A = (UDU^*)^*(UDU^*) = (UD^*U^*)(UDU^*) = UD^*(U^*U)DU^* = U(D^*D)U^*.$$

同理,

$$AA^* = (UDU^*)(UDU^*)^* = UDU^*(UD^*U^*) = UD(U^*U)D^*U^* = U(DD^*)U^*.$$

对于对角矩阵  $D$ , 其元素为  $d_i$ , 则  $D^*$  的元素为  $\bar{d}_i$ .  $D^*D$  和  $DD^*$  都是对角矩阵, 且第  $i$  个对角元均为  $|d_i|^2$ . 因此  $D^*D = DD^*$ . 代回上式, 得  $A^*A = AA^*$ .

- 2.4. 解: 特征多项式:  $(3 - \lambda)^2 - 4 = 0 \implies \lambda^2 - 6\lambda + 5 = 0 \implies \lambda_1 = 5, \lambda_2 = 1$ . 特征向量:  $\lambda_1 = 5$ :  $\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 = x_2$ . 单位化  $\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, 1)^T$ .  $\lambda_2 = 1$ :  $\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0} \implies x_1 = -x_2$ . 单位化  $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1, -1)^T$ . 所以  $A = UDU^*$ , 其中  $D = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}, U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ .  
求平方根  $B$ . 若  $B^2 = A$ , 且  $A$  可对角化, 则  $B$  与  $A$  拥有相同的特征向量 (或  $B$  在  $A$  的基下是对角的).  $B = U\sqrt{D}U^*$ .  $\sqrt{D}$  可以是  $\begin{pmatrix} \pm\sqrt{5} & 0 \\ 0 & \pm 1 \end{pmatrix}$ . 有 4 种组合. 最常用的 (正定) 平方根取正号:  $B_1 = U \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 1 \end{pmatrix} U^* = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{5}+1 & \sqrt{5}-1 \\ \sqrt{5}-1 & \sqrt{5}+1 \end{pmatrix}$ . 其他三个平方根通过改变  $\sqrt{5}$  和 1 的符号得到.

- 2.5. 解: **错误**. 自伴随矩阵的特征值是实数, 但不一定是非负的. 例如  $A = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ . 它是自伴随的. 如果  $B^2 = -I$ , 则  $B$  的特征值必须是  $\pm i$ . 如果  $B$  是自伴随的, 它的特征值必须是实数. 矛盾. 所以  $-I$  没有自伴随的平方根. (注: 如果限制  $A$  为半正定自伴随矩阵, 则命题成立.)

- 2.6. 解: 特征方程:  $(7 - \lambda)(4 - \lambda) - 4 = \lambda^2 - 11\lambda + 24 = (\lambda - 8)(\lambda - 3) = 0$ . 特征值:  $\lambda_1 = 8, \lambda_2 = 3$ . 特征向量:  $\lambda_1 = 8$ :  $\begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix} \implies x_1 = 2x_2$ .  $\mathbf{u}_1 = \frac{1}{\sqrt{5}}(2, 1)^T$ .  $\lambda_2 = 3$ :  $\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} \implies x_2 = -2x_1$ .  $\mathbf{u}_2 = \frac{1}{\sqrt{5}}(1, -2)^T$ . 对角化:

$$A = \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}}_U \underbrace{\begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix}}_D \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}}_{U^*}$$

具有正特征值的平方根  $B$  为:

$$B = U\sqrt{D}U^* = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{3} \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}.$$

- 2.7. 解: **a) 错误**.  $(AB)^* = B^*A^* = BA$ . 只有当  $A, B$  对易 ( $AB = BA$ ) 时, 乘积才自伴随.

反例:  $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ .  $AB = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$  不对称.

**b) 正确**.  $(A^k)^* = (A \dots A)^* = A^* \dots A^* = A \dots A = A^k$ .

- 2.8. 证明: **a)**  $(A^*A)^* = A^*(A^*)^* = A^*A$ . 所以是自伴随的.

**b)** 设  $A^*A\mathbf{v} = \lambda\mathbf{v}$  ( $\mathbf{v} \neq \mathbf{0}$ ).  $(\mathbf{v}, A^*A\mathbf{v}) = (\mathbf{v}, \lambda\mathbf{v}) = \lambda\|\mathbf{v}\|^2$ . 另一方面,  $(\mathbf{v}, A^*A\mathbf{v}) = (A\mathbf{v}, A\mathbf{v}) = \|A\mathbf{v}\|^2 \geq 0$ . 所以  $\lambda = \frac{\|A\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \geq 0$ .

**c)**  $A^*A + I$  的特征值是  $\lambda_i + 1$ . 由 **b)** 知  $\lambda_i \geq 0$ , 所以  $\lambda_i + 1 \geq 1 > 0$ . 所有特征值非零, 所以行列式非零, 矩阵可逆.

- 2.9. 解: **a) 正确**.  $A$  的特征值  $\lambda$  均为实数.  $A + iI$  的特征值为  $\lambda + i$ . 模长  $\sqrt{\lambda^2 + 1} \geq 1 \neq 0$ . 所以可逆.

**b) 正确**.  $U$  的特征值  $\mu$  满足  $|\mu| = 1$ .  $U + \frac{3}{4}I$  的特征值为  $\mu + 0.75$ . 由三角不等式  $|\mu - (-0.75)| \geq \|\mu\| - 0.75 = |1 - 0.75| = 0.25 > 0$ . 特征值不为 0, 可逆.

**c) 错误**. 反例:  $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ .  $A$  的特征值为  $\pm i$ .  $i$  是  $A$  的特征值, 意味着  $\det(A - iI) = 0$ . 不可逆.

- 2.10. 解: 特征方程:  $(\cos \alpha - \lambda)^2 + \sin^2 \alpha = 0 \implies (\lambda - \cos \alpha)^2 = -\sin^2 \alpha \implies \lambda =$

$\cos \alpha \pm i \sin \alpha = e^{\pm i\alpha}$ . 特征向量: 对于  $\lambda_1 = e^{-i\alpha} = \cos \alpha - i \sin \alpha$ :

$$\begin{pmatrix} i \sin \alpha & -\sin \alpha \\ \sin \alpha & i \sin \alpha \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0} \implies ix_1 = x_2.$$

取  $\mathbf{v}_1 = (1, i)^T$ . 单位化  $\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, i)^T$ . 对于  $\lambda_2 = e^{i\alpha}$ , 特征向量为  $\bar{\mathbf{u}}_1 = \frac{1}{\sqrt{2}}(1, -i)^T$ . 所以

$$R_\alpha = \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}}_U \begin{pmatrix} e^{-i\alpha} & 0 \\ 0 & e^{i\alpha} \end{pmatrix} \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}}_{U^*}.$$

2.11. 解: 这是一个实对称矩阵 (也是正交的, 因为  $A^2 = I$ ), 所以特征值是实数. 特征多项式:  $\det(A - \lambda I) = \lambda^2 - (-\cos^2 \alpha - \sin^2 \alpha) = \lambda^2 - 1$ . 特征值:  $\lambda_1 = 1, \lambda_2 = -1$ . 特征向量:

1.  $\lambda = 1$ :  $\begin{pmatrix} \cos \alpha - 1 & \sin \alpha \\ \sin \alpha & -\cos \alpha - 1 \end{pmatrix} \mathbf{x} = \mathbf{0}$ . 利用提示  $\cos \alpha - 1 = -2\sin^2(\alpha/2)$ ,  $\sin \alpha = 2\sin(\alpha/2)\cos(\alpha/2)$ .  $-2\sin^2(\alpha/2)x_1 + 2\sin(\alpha/2)\cos(\alpha/2)x_2 = 0$ . 消去  $2\sin(\alpha/2)$  (假设  $\alpha$  不是  $2\pi$  倍数):  $-\sin(\alpha/2)x_1 + \cos(\alpha/2)x_2 = 0$ . 取  $\mathbf{u}_1 = \begin{pmatrix} \cos(\alpha/2) \\ \sin(\alpha/2) \end{pmatrix}$ .

2.  $\lambda = -1$ : 实对称矩阵不同特征值的向量必正交. 取  $\mathbf{u}_2 = \begin{pmatrix} -\sin(\alpha/2) \\ \cos(\alpha/2) \end{pmatrix}$ . 所以  $A = U \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} U^T$ , 其中  $U = \begin{pmatrix} \cos \frac{\alpha}{2} & -\sin \frac{\alpha}{2} \\ \sin \frac{\alpha}{2} & \cos \frac{\alpha}{2} \end{pmatrix}$ .

2.12. 解:  $A$  是一个特征值为 1 和  $-1$  的正交矩阵.  $\lambda = 1$  的特征向量  $\mathbf{u}_1 = (\cos \frac{\alpha}{2}, \sin \frac{\alpha}{2})^T$  指向角度为  $\alpha/2$  的方向.  $\lambda = -1$  的特征向量  $\mathbf{u}_2$  垂直于  $\mathbf{u}_1$ . 变换保持  $\mathbf{u}_1$  方向不变, 反转  $\mathbf{u}_2$  方向. 这是一个关于经过原点且倾角为  $\alpha/2$  的直线的反射 (镜像).

2.13. 证明: 因为  $A$  是正规的, 所以存在酉矩阵  $U$  使得  $A = UDU^*$ , 其中  $D$  包含特征值  $\lambda_k$ . 验证  $A$  是否为酉算子, 即验证  $A^*A = I$ .

$$A^*A = (UDU^*)^*(UDU^*) = UD^*U^*UDU^* = UD^*DU^*.$$

$D^*D$  是对角矩阵, 其对角元为  $\bar{\lambda}_k \lambda_k = |\lambda_k|^2$ . 由题设  $|\lambda_k| = 1$ , 所以  $|\lambda_k|^2 = 1$ . 因此  $D^*D = I$ . 所以  $A^*A = UIU^* = I$ . 得证.

2.14. 证明: 正规算子  $A$  可以被酉对角化:  $A = UDU^*$ . 由于特征值是实数, 对角矩阵  $D$  是实矩阵, 即  $D = D^*$  ( $D$  自伴随). 取伴随:

$$A^* = (UDU^*)^* = UD^*U^* = UDU^* = A.$$

因此  $A$  是自伴随的.

2.15. 解: **a) 可对角化但不容许正交特征基:** 考虑复对称矩阵  $A = \begin{pmatrix} 2i & 1 \\ 1 & 0 \end{pmatrix}$ . ( $A^T = A$ ). 特征值:  $\lambda^2 - 2i\lambda - 1 = 0 \implies (\lambda - i)^2 = 0$ . 这是一个亏损矩阵, 不可对角化. 让我们换一个例子.  $A = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ . 特征值  $\lambda^2 - (-1) = 0 \implies \lambda = \pm i$ . 不同的特征值, 所以可对角化.

特征向量:  $\lambda = i: (i, i)^T \rightarrow (1, 1)^T$ .  $\lambda = -i: (i, -i)^T \rightarrow (1, -1)^T$ . 在标准复内积下,  $(1, 1)^T$  和  $(1, -1)^T$  是正交的 ( $1 \cdot 1 + 1 \cdot (-1) = 0$ ). 这个例子不行 (因为它是正规的). 我们需要一个非正规的复对称矩阵. 取  $A = \begin{pmatrix} 1 & i \\ i & 2 \end{pmatrix}$ .  $A$  是对称的.  $A^* = \begin{pmatrix} 1 & -i \\ -i & 2 \end{pmatrix} \neq A$ . 非自伴随.

伴随.  $AA^* = \begin{pmatrix} 2 & i \\ -i & 5 \end{pmatrix}$ ,  $A^*A = \begin{pmatrix} 2 & -i \\ i & 5 \end{pmatrix}$ . 不相等, 非正规. 因为它非正规, 根据谱定理, 它没有正交特征向量基. 但既然是对称矩阵, 只要特征值不同, 它就是可对角化的. 特征方程:  $(1 - \lambda)(2 - \lambda) + 1 = \lambda^2 - 3\lambda + 3 = 0$ . 判别式  $9 - 12 = -3 \neq 0$ . 有两个不同特征值, 所以可对角化, 但没有正交基.

**b) 不能被对角化:** 考虑  $A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$ . ( $A^T = A$ ). 特征多项式:  $(1 - \lambda)(-1 - \lambda) - (i)^2 = -(1 - \lambda^2) + 1 = \lambda^2 - 1 + 1 = \lambda^2$ . 特征值  $\lambda = 0$  (代数重数 2). 零空间:  $\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} =$

$0 \Rightarrow x + iy = 0$ . 只有一个线性无关的特征向量  $(-i, 1)^T$ . 几何重数  $1 <$  代数重数  $2$ . 所以不可对角化.

- 3.1. 证明: 设  $A$  的奇异值分解 (SVD) 为  $A = W\Sigma V^*$ , 其中  $W$  和  $V$  是酉矩阵 (可逆),  $\Sigma$  是对角矩阵, 对角线上是奇异值  $\sigma_1, \sigma_2, \dots, \sigma_n$ . 矩阵的秩在乘以可逆矩阵后保持不变. 因此:

$$\text{rank}(A) = \text{rank}(W\Sigma V^*) = \text{rank}(\Sigma).$$

对于对角矩阵  $\Sigma$ , 其秩显然等于非零对角元素的数量. 由于  $\Sigma$  的对角元素即为  $A$  的奇异值, 因此  $A$  的秩等于非零奇异值的数量.

- 3.2. 解: 1) 对于矩阵  $A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$ : 计算  $A^*A = \begin{pmatrix} 2 & 0 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 13 \end{pmatrix}$ . 特征方程:  $(4 - \lambda)(13 - \lambda) - 36 = \lambda^2 - 17\lambda + 52 - 36 = \lambda^2 - 17\lambda + 16 = (\lambda - 16)(\lambda - 1) = 0$ . 特征值:  $\lambda_1 = 16, \lambda_2 = 1$ . 奇异值:  $\sigma_1 = 4, \sigma_2 = 1$ .

计算  $A^*A$  的特征向量 (即  $\mathbf{v}_k$ ): 对于  $\lambda_1 = 16$ :  $\begin{pmatrix} -12 & 6 \\ 6 & -3 \end{pmatrix} \mathbf{x} = \mathbf{0} \Rightarrow 2x_1 = x_2$ . 单位化得  $\mathbf{v}_1 = \frac{1}{\sqrt{5}}(1, 2)^T$ . 对于  $\lambda_2 = 1$ : 与  $\mathbf{v}_1$  正交,  $\mathbf{v}_2 = \frac{1}{\sqrt{5}}(2, -1)^T$ .

计算  $\mathbf{w}_k = \frac{1}{\sigma_k} A \mathbf{v}_k$ :  $\mathbf{w}_1 = \frac{1}{4} \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{4\sqrt{5}} \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ .

$\mathbf{w}_2 = \frac{1}{1} \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ .

施密特分解为:

$$A = 4\mathbf{w}_1\mathbf{v}_1^* + 1\mathbf{w}_2\mathbf{v}_2^* = 4 \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} + 1 \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix}.$$

2) 对于第二个矩阵: (注: 此矩阵计算较为繁琐, 通常课堂练习会给出整洁的数字, 这里给出方法) 计算  $A^*A$  的特征值得到  $\sigma_k^2$ , 求出单位特征向量  $\mathbf{v}_k$ , 再由  $\mathbf{w}_k = \sigma_k^{-1} A \mathbf{v}_k$  得到  $\mathbf{w}_k$ .

3) 对于第三个矩阵: 同上方法.

- 3.3. 解: 1) 对于  $A^*$ : 取伴随:

$$A^* = (W\Sigma V^*)^* = (V^*)\Sigma^*W^* = V\Sigma W^*.$$

因为  $\Sigma$  是实对角矩阵,  $\Sigma^* = \Sigma$ . 且  $V, W$  是酉矩阵, 所以  $V$  和  $W$  分别也是酉矩阵. 这已经是 SVD 的形式 (左酉矩阵  $V$ , 中间对角矩阵  $\Sigma$ , 右酉矩阵的伴随  $W^*$ ). 奇异值不变, 左/右奇异向量互换.

2) 对于  $A^{-1}$ : 取逆:

$$A^{-1} = (W\Sigma V^*)^{-1} = (V^*)^{-1}\Sigma^{-1}W^{-1} = V\Sigma^{-1}W^*.$$

$\Sigma^{-1}$  是对角矩阵, 对角元素为  $1/\sigma_k$ . 注意: SVD 通常要求奇异值按降序排列. 如果  $\sigma_1 \geq \dots \geq \sigma_n > 0$ , 那么  $1/\sigma_n \geq \dots \geq 1/\sigma_1$ . 我们需要重新排列  $V$  和  $W$  的列以满足标准定义, 但就分解形式而言,  $V\Sigma^{-1}W^*$  是正确的.

- 3.4. 解: a) 观察到  $A$  的列是线性相关的:  $\text{Col}_2 = -\frac{1}{3}\text{Col}_1$ . 所以秩为 1. 计算  $A^*A$ :

$$A^*A = \begin{pmatrix} -3 & 6 & 6 \\ 1 & -2 & -2 \end{pmatrix} \begin{pmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{pmatrix} = \begin{pmatrix} 81 & -27 \\ -27 & 9 \end{pmatrix}.$$

迹 = 90, 行列式 = 0. 特征值:  $\lambda_1 = 90, \lambda_2 = 0$ . 奇异值:  $\sigma_1 = \sqrt{90} = 3\sqrt{10}, \sigma_2 = 0$ .

求  $V$  ( $A^*A$  的特征向量):  $\lambda_1 = 90$ :  $\begin{pmatrix} -9 & -27 \\ -27 & -81 \end{pmatrix} \Rightarrow x_1 = -3x_2$ .  $\mathbf{v}_1 = \frac{1}{\sqrt{10}}(3, -1)^T$ .

$\lambda_2 = 0$ :  $\mathbf{v}_2 = \frac{1}{\sqrt{10}}(1, 3)^T$ . 所以  $V = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}$ .



求  $W$ :  $\mathbf{w}_1 = \frac{1}{\sigma_1} A \mathbf{v}_1 = \frac{1}{3\sqrt{10}} \begin{pmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{pmatrix} \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix} = \frac{1}{30} \begin{pmatrix} -10 \\ 20 \\ 20 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$ . 我们需要将  $\mathbf{w}_1$  扩充为  $\mathbb{R}^3$  的标准正交基来得到酉矩阵  $W$ . 可以选择  $\mathbf{w}_2 = \frac{1}{\sqrt{5}}(2, 1, 0)^T$  (正交于  $\mathbf{w}_1$ ), 然后  $\mathbf{w}_3 = \mathbf{w}_1 \times \mathbf{w}_2$ . 或者简单观察:  $(2, 1, 0)^T$  和  $(2, -2, 3)^T$  等. 构造  $W = (\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ .  

$$\Sigma = \begin{pmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

b) 计算  $AA^*$  (因为是  $2 \times 3$ , 计算  $2 \times 2$  更简单):

$$AA^* = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}.$$

特征值:  $(17 - \lambda)^2 - 64 = 0 \Rightarrow \lambda = 17 \pm 8$ .  $\lambda_1 = 25, \lambda_2 = 9$ . 奇异值:  $\sigma_1 = 5, \sigma_2 = 3$ .

$AA^*$  的特征向量 (构成  $W$ ):  $\lambda_1 = 25$ :  $\begin{pmatrix} -8 & 8 \\ 8 & -8 \end{pmatrix} \Rightarrow \mathbf{w}_1 = \frac{1}{\sqrt{2}}(1, 1)^T$ .  $\lambda_2 = 9$ :

$\mathbf{w}_2 = \frac{1}{\sqrt{2}}(1, -1)^T$ .  $W = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ .

求  $V$ :  $\mathbf{v}_1 = \frac{1}{\sigma_1} A^* \mathbf{w}_1 = \frac{1}{5} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{5\sqrt{2}} \begin{pmatrix} 5 \\ 5 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ .  $\mathbf{v}_2 = \frac{1}{\sigma_2} A^* \mathbf{w}_2 =$

$\frac{1}{3} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$ . 需要找  $\mathbf{v}_3$  正交于  $\mathbf{v}_1, \mathbf{v}_2$ .  $\mathbf{v}_1 \times \mathbf{v}_2 \propto (1, 1, 0)^T \times$

$(1, -1, 4)^T = (4, -4, -2)^T \propto (-2, 2, 1)^T$ . 单位化  $\mathbf{v}_3 = \frac{1}{3}(-2, 2, 1)^T$ .  $V = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ .

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}.$$

3.5. 解: 由习题 3.2, 我们已知 SVD 元素:  $\sigma_1 = 4, \sigma_2 = 1$ .  $V = (\mathbf{v}_1, \mathbf{v}_2) = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$ .

$$W = (\mathbf{w}_1, \mathbf{w}_2) = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}.$$

a) 最大值为  $\sigma_1 = 4$ . 达到的向量为右奇异向量  $\mathbf{v}_1 = \frac{1}{\sqrt{5}}(1, 2)^T$  (及其反方向).

b) 最小值为  $\sigma_2 = 1$ . 达到的向量为右奇异向量  $\mathbf{v}_2 = \frac{1}{\sqrt{5}}(2, -1)^T$ .

c) 像  $A(B)$  是一个实心椭圆.

其半长轴长度为  $\sigma_1 = 4$ , 方向沿  $\mathbf{w}_1 = \frac{1}{\sqrt{5}}(2, 1)^T$ . 其半短轴长度为  $\sigma_2 = 1$ , 方向沿  $\mathbf{w}_2 = \frac{1}{\sqrt{5}}(1, -2)^T$ .

3.6. 证明: 利用极分解  $A = U|A|$ , 其中  $U$  是酉矩阵,  $|A| = \sqrt{A^*A}$  是半正定矩阵. 两边取行列式:

$$\det A = \det(U|A|) = \det U \cdot \det |A|.$$

两边取绝对值 (模):

$$|\det A| = |\det U| \cdot |\det |A||.$$

因为  $U$  是酉矩阵,  $\det U$  的模为 1 (即  $|\det U| = 1$ ). 因为  $|A|$  是半正定矩阵, 其特征值非负, 故  $\det |A| \geq 0$ , 所以  $|\det |A|| = \det |A|$ . 因此  $|\det A| = \det |A|$ .

3.7. 解: a) 错误. 例如  $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ , 特征值为 0, 但奇异值为 1, 0.

b) 错误. 奇异值是  $A^*A$  特征值的算术平方根.

c) 正确.  $(cA)^*(cA) = \bar{c}cA^*A = |c|^2A^*A$ . 特征值变为  $|c|^2\sigma^2$ , 开方得  $|c|\sigma$ .

d) 正确. 根据定义, 奇异值是半正定矩阵  $|A|$  的特征值, 非负.

e) 错误. 它们等于特征值的绝对值. 例如  $A = -I$ , 特征值 -1, 奇异值 1. (若  $A$  半正定, 则命题成立).

- 3.8. 证明: 设  $\lambda \neq 0$  是  $A^*A$  的特征值, 对应特征向量  $\mathbf{v}$ . 即  $A^*A\mathbf{v} = \lambda\mathbf{v}$ . 左乘  $A$ :  $A(A^*A\mathbf{v}) = A(\lambda\mathbf{v}) \implies (AA^*)(A\mathbf{v}) = \lambda(A\mathbf{v})$ . 由于  $\lambda \neq 0, \mathbf{v} \neq \mathbf{0}$ , 且  $A^*A\mathbf{v} = \lambda\mathbf{v} \neq \mathbf{0}$ , 这意味着  $A\mathbf{v} \neq \mathbf{0}$ . 所以  $A\mathbf{v}$  是  $AA^*$  的特征向量, 特征值为  $\lambda$ . 同理可证  $AA^*$  的非零特征值也是  $A^*A$  的. 零特征值的重数相同当且仅当  $A$  是方阵 ( $m = n$ ). 因为特征值总数分别为  $n$  和  $m$ , 非零个数均为  $r = \text{rank}(A)$ , 所以零特征值个数分别为  $n - r$  和  $m - r$ .
- 3.9. 证明: 最大奇异值  $s = \|A\|$  (算子范数). 设  $\lambda$  为任意特征值,  $\mathbf{v}$  为对应单位特征向量 ( $\|\mathbf{v}\| = 1$ ).  $A\mathbf{v} = \lambda\mathbf{v}$ . 两边取范数:  $\|A\mathbf{v}\| = \|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\| = |\lambda|$ . 根据算子范数定义  $\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| \geq \|A\mathbf{v}\|$ . 所以  $s \geq |\lambda|$ .
- 3.10. 解: 见习题 3.1 的解答.
- 3.11. 证明: 设奇异值为  $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ . 算子范数  $\|A\| = \sigma_1$ . Frobenius 范数  $\|A\|_2 = \sqrt{\sum_{k=1}^r \sigma_k^2}$ . 显然  $\sqrt{\sum_{k=1}^r \sigma_k^2} = \sigma_1 \iff \sum_{k=1}^r \sigma_k^2 = \sigma_1^2 \iff \sigma_2 = \sigma_3 = \cdots = 0$ . 即只有一个非零奇异值, 也就是  $\text{rank}(A) = 1$ .

- 3.12. 解: 首先找出  $A$  的 SVD. 注意这个矩阵是习题 3.2 中矩阵  $B = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$  的广义变体 (符号差异).  $A^*A = \begin{pmatrix} 2 & 0 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & -6 \\ -6 & 13 \end{pmatrix}$ . 特征值与之前相同:  $\lambda = 16, 1$ . 奇异值  $\sigma_1 = 4, \sigma_2 = 1$ . 特征向量 ( $V$  的列):  $\lambda = 16$ :  $\begin{pmatrix} -12 & -6 \\ -6 & -3 \end{pmatrix} \rightarrow \mathbf{v}_1 = \frac{1}{\sqrt{5}}(1, -2)^T$ .  $\lambda = 1$ :  $\mathbf{v}_2 = \frac{1}{\sqrt{5}}(2, 1)^T$ . 条件  $\|A\mathbf{x}\| \leq 1$  等价于  $\|\Sigma V^* \mathbf{x}\| \leq 1$ . 令  $\mathbf{y} = V^* \mathbf{x}$  (即  $\mathbf{x}$  在基  $V$  下的坐标). 条件变为  $\|\Sigma \mathbf{y}\| \leq 1$ , 即:

$$16y_1^2 + y_2^2 \leq 1 \quad (\text{注意特征值是奇异值的平方, 这里用的是 } \sigma y \text{ 的模方}).$$

修正:  $\|\Sigma \mathbf{y}\|^2 = \sigma_1^2 y_1^2 + \sigma_2^2 y_2^2 = 16y_1^2 + y_2^2 \leq 1$ . 这是一个在  $\mathbf{y}$  坐标系 (由  $\mathbf{v}_1, \mathbf{v}_2$  定义) 中的椭圆. 半短轴长度为  $1/4$  (沿  $\mathbf{v}_1$  方向), 半长轴长度为  $1$  (沿  $\mathbf{v}_2$  方向). 几何描述: 逆像是  $\mathbb{R}^2$  中的一个实心椭圆, 长轴沿向量  $(2, 1)^T$ , 短轴沿向量  $(1, -2)^T$ .

- 4.1. 解: **a)** 计算  $A^*A$ :

$$A^*A = \begin{pmatrix} 4 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 17 & 3 \\ 3 & 9 \end{pmatrix}.$$

特征方程:

$$\lambda^2 - \text{trace}(A^*A)\lambda + \det(A^*A) = \lambda^2 - 26\lambda + (153 - 9) = \lambda^2 - 26\lambda + 144 = 0.$$

因式分解:  $(\lambda - 18)(\lambda - 8) = 0$ . 特征值为  $\lambda_1 = 18, \lambda_2 = 8$ . 奇异值为  $s_1 = \sqrt{18} = 3\sqrt{2}, s_2 = \sqrt{8} = 2\sqrt{2}$ .

范数:  $\|A\| = s_1 = 3\sqrt{2}$ .

条件数:  $\kappa(A) = s_1/s_2 = 3\sqrt{2}/2\sqrt{2} = 1.5$ .

构造例子: 要使不等式变为等式, 我们需要“最坏情况”. 根据误差界理论, 当  $\mathbf{b}$  沿着对应于最大奇异值  $s_1$  的左奇异向量  $\mathbf{w}_1$  方向, 且误差  $\Delta \mathbf{b}$  沿着对应于最小奇异值  $s_2$  的左奇异向量  $\mathbf{w}_2$  方向时, 相对误差被最大程度放大 (或者反过来, 取决于公式的导数, 但标准结论是: 输入  $\mathbf{b}$  在“最大放大”方向, 而输入误差  $\Delta \mathbf{b}$  在“最小放大”方向时, 解的相对误差并不大; 反之, 如果  $\mathbf{b}$  在“最小放大”方向 (解向量  $\mathbf{x}$  很大), 而  $\Delta \mathbf{b}$  在“最大放大”方向 ( $\Delta \mathbf{x}$  很大), 则比值最大). 让我们仔细检查:  $\mathbf{x} = A^{-1}\mathbf{b}, \Delta \mathbf{x} = A^{-1}\Delta \mathbf{b}$ .  $\|\mathbf{x}\| \geq (1/s_1)\|\mathbf{b}\|$ ,  $\|\Delta \mathbf{x}\| \leq (1/s_n)\|\Delta \mathbf{b}\|$ . 要最大化  $\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|}$ , 我们需要  $\|\Delta \mathbf{x}\|$  尽可能大 (即  $\Delta \mathbf{b}$  沿  $A^{-1}$  的最大放大方向, 即对应  $A$  的最小奇异值  $s_n$  的方向  $\mathbf{w}_n$ ), 同时  $\|\mathbf{x}\|$  尽可能小 (即  $\mathbf{b}$  沿  $A^{-1}$  的最小放大方向, 即对应  $A$  的最大奇异值  $s_1$  的方向  $\mathbf{w}_1$ ).

计算  $A^*A$  的特征向量 (右奇异向量  $\mathbf{v}$ ):  $\lambda_1 = 18$ :  $\begin{pmatrix} -1 & 3 \\ 3 & -9 \end{pmatrix} \implies \mathbf{v}_1 \propto (3, 1)^T$ .  $\lambda_2 = 8$ :

$\mathbf{v}_2 \propto (1, -3)^T$  (正交于  $\mathbf{v}_1$ ). 计算左奇异向量  $\mathbf{w}_k$  (即  $A\mathbf{v}_k$  的方向):  $A\mathbf{v}_1 = \begin{pmatrix} 4 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 12 \\ 6 \end{pmatrix} = 6 \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ . 归一化  $\mathbf{w}_1 = \frac{1}{\sqrt{5}}(2, 1)^T$ .  $A\mathbf{v}_2 = \begin{pmatrix} 4 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} 4 \\ -8 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ . 归一

化  $\mathbf{w}_2 = \frac{1}{\sqrt{5}}(1, -2)^T$ .

选取  $\mathbf{b} = \mathbf{w}_1 = (2, 1)^T$  (对应最大奇异值), 选取  $\Delta \mathbf{b} = \epsilon \mathbf{w}_2 = \epsilon(1, -2)^T$  (对应最小奇异值). 此时  $\mathbf{x} = s_1^{-1} \mathbf{v}_1$ ,  $\|\mathbf{x}\| = 1/s_1$ .  $\Delta \mathbf{x} = s_2^{-1} \epsilon \mathbf{v}_2$ ,  $\|\Delta \mathbf{x}\| = \epsilon/s_2$ .  $\text{LHS} = \frac{\epsilon/s_2}{1/s_1} = \epsilon \frac{s_1}{s_2}$ .  $\text{RHS} = \frac{s_1}{s_2} \epsilon = \epsilon \frac{s_1}{s_2}$ . 等式成立.

b)  $A = \begin{pmatrix} 5 & 3 \\ -3 & 3 \end{pmatrix}$ . 计算  $A^*A$ :

$$A^*A = \begin{pmatrix} 5 & -3 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ -3 & 3 \end{pmatrix} = \begin{pmatrix} 34 & 6 \\ 6 & 18 \end{pmatrix}.$$

特征方程:  $\lambda^2 - 52\lambda + (34 \times 18 - 36) = \lambda^2 - 52\lambda + 576 = 0$ . 求根:  $\lambda = \frac{52 \pm \sqrt{2704 - 2304}}{2} = \frac{52 \pm 20}{2}$ .  $\lambda_1 = 36, \lambda_2 = 16$ . 奇异值为  $s_1 = 6, s_2 = 4$ .

范数:  $\|A\| = 6$ .

条件数:  $\kappa(A) = 6/4 = 1.5$ .

4.2. 证明: 因为  $A$  是正常算子 ( $A^*A = AA^*$ ), 根据谱定理, 存在酉矩阵  $U$  使得  $A$  可以被酉对角化:

$$A = UDU^*,$$

其中  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ . 计算  $A^*A$ :

$$A^*A = (UDU^*)^*(UDU^*) = UD^*U^*UDU^* = UD^*DU^*.$$

因为  $D$  是对角矩阵,  $D^*D$  也是对角矩阵, 其对角元素为  $\bar{\lambda}_i \lambda_i = |\lambda_i|^2$ . 这表明  $A^*A$  的特征值是  $|\lambda_1|^2, \dots, |\lambda_n|^2$ . 根据奇异值的定义 ( $A^*A$  特征值的算术平方根),  $A$  的奇异值为  $\sqrt{|\lambda_i|^2} = |\lambda_i|$ . 得证.

4.3. 解: 我们利用提示的方法. 观察矩阵  $A$  的结构:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = J + I.$$

$J$  是全 1 矩阵.  $J$  可以写成  $3P$ , 其中  $P$  是向向量  $\mathbf{v} = (1, 1, 1)^T$  张成的子空间的正交投影矩阵 ( $P = \frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2} = \frac{1}{3}J$ ).

a) 正交投影  $P$  的特征值是 1 (对应  $E$  中的向量) 和 0 (对应  $E^\perp$  中的向量).

b)  $J = 3P$  的特征值是 3 (重数 1) 和 0 (重数 2).

c)  $A = J + I$ . 如果  $\lambda$  是  $J$  的特征值, 那么  $\lambda + 1$  是  $A$  的特征值 (因为  $J$  和  $I$  交换, 特征向量相同). 所以  $A$  的特征值为  $3 + 1 = 4$  (重数 1) 和  $0 + 1 = 1$  (重数 2). 因为  $A$  是实对称矩阵 (自伴随), 其特征值均为正, 所以奇异值等于特征值.

奇异值:  $s_1 = 4, s_2 = 1, s_3 = 1$ .

范数:  $\|A\| = s_1 = 4$ .

条件数:  $\kappa(A) = 4/1 = 4$ .

4.4. 证明: 在约简奇异值分解中,  $\tilde{W}$  是  $m \times r$  矩阵, 列向量  $\mathbf{w}_1, \dots, \mathbf{w}_r$  是  $A$  的对应于非零奇异值的左奇异向量 (两两正交且归一化).  $\tilde{\Sigma}$  是  $r \times r$  对角正定矩阵.  $\tilde{V}^*$  是  $r \times n$  矩阵, 行向量是右奇异向量. 对任意  $\mathbf{x} \in \mathbb{R}^n$ :

$$A\mathbf{x} = \tilde{W}(\tilde{\Sigma}\tilde{V}^*\mathbf{x}).$$

令  $\mathbf{y} = \tilde{\Sigma}\tilde{V}^*\mathbf{x} \in \mathbb{R}^r$ . 则  $A\mathbf{x} = \tilde{W}\mathbf{y} = \sum_{k=1}^r y_k \mathbf{w}_k$ . 这说明  $\text{Ran } A \subseteq \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_r\} = \text{Ran } \tilde{W}$ . 由于  $\text{rank}(A) = r$  且  $\mathbf{w}_1, \dots, \mathbf{w}_r$  线性无关,  $\text{Ran } \tilde{W}$  的维数也是  $r$ . 由维数相等和包含关系可知  $\text{Ran } A = \text{Ran } \tilde{W}$ .

对  $A^*$  取伴随:

$$A^* = (\tilde{W}\tilde{\Sigma}\tilde{V}^*)^* = \tilde{V}\tilde{\Sigma}^*\tilde{W}^* = \tilde{V}\tilde{\Sigma}\tilde{W}^*.$$

这正是  $A^*$  的约简奇异值分解 (只是  $U, V$  角色互换). 应用同样的逻辑,  $\text{Ran } A^*$  是由左侧矩阵  $\tilde{V}$  的列张成的. 因此  $\text{Ran } A^* = \text{Ran } \tilde{V}$ .

- 4.5. 解: 如果  $A = W\Sigma V^*$ , 其中  $\Sigma = \text{diag}(s_1, \dots, s_p)$  ( $p = \min(m, n)$ , 可能包含零)。则摩尔-彭罗斯逆由下式给出:

$$A^+ = V\Sigma^+W^*,$$

其中  $\Sigma^+$  是  $n \times m$  对角矩阵 (与  $\Sigma$  转置同型), 其对角元素定义为:

$$(\Sigma^+)_{kk} = \begin{cases} 1/s_k & \text{如果 } s_k \neq 0 \\ 0 & \text{如果 } s_k = 0 \end{cases}.$$

- 4.6. 证明: 我们使用 SVD 代入证明第一个等式 (第二个类似)。设  $A = W\Sigma V^*$  (这里使用全 SVD 或约简 SVD 均可, 使用约简 SVD 更直观, 设  $r$  为秩)。 $A^*A = V\Sigma^2V^*$ . (注意  $W^*W = I$ ). 则  $A^*A + \varepsilon I = V(\Sigma^2 + \varepsilon I)V^*$  (利用  $V$  的正交性质  $VIV^* = I$ ). 求逆:

$$(A^*A + \varepsilon I)^{-1} = V(\Sigma^2 + \varepsilon I)^{-1}V^*.$$

右乘  $A^* = V\Sigma W^*$ :

$$(A^*A + \varepsilon I)^{-1}A^* = V(\Sigma^2 + \varepsilon I)^{-1}V^*V\Sigma W^* = V[(\Sigma^2 + \varepsilon I)^{-1}\Sigma]W^*.$$

中间的对角矩阵  $(\Sigma^2 + \varepsilon I)^{-1}\Sigma$  的第  $k$  个对角元素为:

$$\frac{s_k}{s_k^2 + \varepsilon}.$$

当  $\varepsilon \rightarrow 0^+$  时:

如果  $s_k \neq 0$ , 该项趋于  $\frac{s_k}{s_k^2} = \frac{1}{s_k}$ .

如果  $s_k = 0$ , 该项始终为  $\frac{0}{0+\varepsilon} = 0$ , 极限为 0.

这正是  $\Sigma^+$  的定义。所以,

$$\lim_{\varepsilon \rightarrow 0^+} V[(\Sigma^2 + \varepsilon I)^{-1}\Sigma]W^* = V\Sigma^+W^* = A^+.$$

得证。

- 6.1. 解: 设标准基为  $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$ , 新基为  $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2\} = \{\mathbf{e}_2, \mathbf{e}_1\}$ . 从新基  $\mathcal{V}$  到标准基  $\mathcal{E}$  的坐标变换矩阵 (过渡矩阵) 为  $P = (\mathbf{v}_1, \mathbf{v}_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . 设  $R_\alpha$  在标准基下的矩阵为  $A$ , 在新基下的矩阵为  $B$ . 根据相似矩阵的性质:

$$B = P^{-1}AP.$$

由于  $P$  是置换矩阵 (交换两行/列), 它是正交的且是对称的, 所以  $P^{-1} = P^T = P$ . 计算  $B$ :

$$\begin{aligned} B &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \sin \alpha & \cos \alpha \\ \cos \alpha & -\sin \alpha \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \end{aligned}$$

注: 这个结果对应于  $-\alpha$  角的旋转矩阵。这是符合直觉的, 因为我们交换了  $x$  轴和  $y$  轴, 改变了空间的方向 (翻转), 使得逆时针旋转看起来变成了顺时针旋转。

- 6.2. 证明: 我们需要构造一个连续的矩阵函数  $M(t)$ , 其中  $t \in [0, 1]$ , 使得  $M(t)$  对所有  $t$  都是可逆的, 且  $M(0) = I_2$ ,  $M(1) = R_\alpha$ . 定义  $M(t)$  为角度  $t\alpha$  的旋转矩阵:

$$M(t) = R_{t\alpha} = \begin{pmatrix} \cos(t\alpha) & -\sin(t\alpha) \\ \sin(t\alpha) & \cos(t\alpha) \end{pmatrix}, \quad t \in [0, 1].$$

显然,  $M(t)$  的元素是  $t$  的连续函数。  $M(0) = R_0 = I_2$ .  $M(1) = R_\alpha$ . 对于任何  $t$ ,  $\det(M(t)) = \cos^2(t\alpha) + \sin^2(t\alpha) = 1 \neq 0$ . 因此,  $M(t)$  始终是可逆的。得证。

- 6.3. 证明: 根据第 5 节的定理 (正交矩阵的正则形式), 因为  $U$  是正交的且  $\det U > 0$  (即  $\det U = 1$ ), 存在一个正交矩阵  $Q$  使得  $U$  可以分块对角化为:

$$U = QDQ^T = Q \operatorname{diag}(R_{\phi_1}, R_{\phi_2}, \dots, R_{\phi_k}, 1, \dots, 1)Q^T,$$

其中  $R_{\phi_j}$  是  $2 \times 2$  的旋转块. 注意: 因为  $\det U = 1$ , 所以特征值  $-1$  的数量 (如果有) 必须是偶数, 它们可以两两组合成  $180^\circ$  的旋转块  $R_\pi$ , 所以我们可以假设对角线上全是旋转块和 1. 根据习题 6.2, 每个旋转块  $R_{\phi_j}$  都可以通过路径  $R_{t\phi_j}$  连续变换为  $I_2$  (当  $t$  从 1 变到 0). 定义对角矩阵函数  $D(t)$ , 将其中的每个块  $R_{\phi_j}$  替换为  $R_{t\phi_j}$ :

$$D(t) = \operatorname{diag}(R_{t\phi_1}, \dots, R_{t\phi_k}, 1, \dots, 1), \quad t \in [0, 1].$$

显然  $D(1) = D$  且  $D(0) = I_n$ . 且  $\det D(t) = 1 \neq 0$ . 现在定义  $U(t) = QD(t)Q^T$ . 由于  $Q$  是常数矩阵,  $U(t)$  是连续的.  $\det U(t) = \det Q \cdot \det D(t) \cdot \det Q^T = 1 \cdot 1 \cdot 1 = 1 \neq 0$ .  $U(1) = QDQ^T = U$ .  $U(0) = QI_nQ^T = I_n$ . 因此, 我们构造了一条从  $I_n$  到  $U$  的可逆矩阵连续路径.

- 6.4. 证明: 由于  $A$  是正定埃尔米特矩阵, 它可以被酉对角化, 且所有特征值均为正实数. 即  $A = V\Lambda V^*$ , 其中  $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ , 且  $\lambda_i > 0$ . 考虑连接 1 和  $\lambda_i$  的直线路径  $\lambda_i(t) = (1-t) \cdot 1 + t \cdot \lambda_i$ ,  $t \in [0, 1]$ . 由于  $1 > 0$  且  $\lambda_i > 0$ , 其凸组合  $\lambda_i(t)$  对所有  $t \in [0, 1]$  恒大于 0. 定义  $\Lambda(t) = \operatorname{diag}(\lambda_1(t), \dots, \lambda_n(t))$ . 定义  $A(t) = V\Lambda(t)V^*$ . 则  $A(t)$  是连续的.  $A(0) = VIV^* = I$ .  $A(1) = V\Lambda V^* = A$ .  $\det A(t) = \prod \lambda_i(t) > 0$ , 所以  $A(t)$  始终可逆. 得证.

(另一种简单的方法是利用正定矩阵集合的凸性: 令  $A(t) = (1-t)I + tA$ . 因为  $I$  和  $A$  都是正定的, 正定矩阵构成的锥是凸的, 所以  $A(t)$  对所有  $t \in [0, 1]$  都是正定的, 从而可逆.)

- 6.5. 证明: 定理的“仅当”部分陈述为: 如果两个基  $\mathcal{A}$  和  $\mathcal{B}$  具有相同的方向 (即变换矩阵  $M = [I]_{\mathcal{B}\mathcal{A}}$  满足  $\det M > 0$ ), 则  $\mathcal{A}$  可以连续变换为  $\mathcal{B}$ . 这也等价于证明: 如果  $\det M > 0$ , 则矩阵  $M$  可以通过可逆矩阵连续变换为单位矩阵  $I$ . 对  $M$  进行极分解:

$$M = UP,$$

其中  $U$  是正交矩阵,  $P$  是正定对称矩阵 (即自伴随矩阵). 取行列式:  $\det M = \det U \det P$ . 因为  $P$  是正定的, 所以  $\det P > 0$ . 已知  $\det M > 0$ , 所以必须有  $\det U > 0$ . 现在我们需要将  $M$  连续变换到  $I$ . 我们可以分两步或同时进行: 1. 根据习题 6.4, 正定矩阵  $P$  可以连续变换为  $I$ . 设此路径为  $P(t)$ , 其中  $P(0) = P, P(1) = I$ . 2. 根据习题 6.3, 正行列式的正交矩阵  $U$  可以连续变换为  $I$ . 设此路径为  $U(t)$ , 其中  $U(0) = U, U(1) = I$ . 定义总路径  $M(t) = U(t)P(t)$ .  $M(t)$  是两个可逆矩阵的乘积, 因此是可逆的.  $M(0) = UP = M$ .  $M(1) = I \cdot I = I$ . 因此, 坐标变换矩阵  $M$  可以连续形变为  $I$ . 这对应于基  $\mathcal{A}$  连续变换为基  $\mathcal{B}$ . 得证.

## 第七章习题解答

- 1.1. 解: 设矩阵为  $A$ . 根据定义  $L(\mathbf{x}, \mathbf{y}) = (A\mathbf{x}, \mathbf{y}) = \sum_{j,k=1}^3 A_{jk}x_jy_k$ . 矩阵  $A$  的第  $j$  行第  $k$  列的元素  $A_{jk}$  对应于多项式中项  $x_jy_k$  的系数.  
 $j = 1$  ( $x_1$  的系数):  $y_1$  系数为 1,  $y_2$  系数为 2,  $y_3$  系数为 14.  $\implies$  第一行为 (1, 2, 14).  
 $j = 2$  ( $x_2$  的系数):  $y_1$  系数为 -5,  $y_2$  系数为 2,  $y_3$  系数为 -3.  $\implies$  第二行为 (-5, 2, -3).  
 $j = 3$  ( $x_3$  的系数):  $y_1$  系数为 8,  $y_2$  系数为 19,  $y_3$  系数为 -2.  $\implies$  第三行为 (8, 19, -2).  
 因此, 矩阵为:

$$A = \begin{pmatrix} 1 & 2 & 14 \\ -5 & 2 & -3 \\ 8 & 19 & -2 \end{pmatrix}.$$

1.2. 解: 设  $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T$ . 计算行列式:

$$L(\mathbf{x}, \mathbf{y}) = \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} = x_1 y_2 - x_2 y_1.$$

我们需要找到矩阵  $A$  使得  $L(\mathbf{x}, \mathbf{y}) = \sum A_{jk} x_j y_k$ .

对照系数:  $x_1 y_1$  系数为 0  $\Rightarrow A_{11} = 0$ .  $x_1 y_2$  系数为 1  $\Rightarrow A_{12} = 1$ .  $x_2 y_1$  系数为 -1  $\Rightarrow A_{21} = -1$ .  $x_2 y_2$  系数为 0  $\Rightarrow A_{22} = 0$ .

因此, 矩阵为:

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

1.3. 解: 对于二次型  $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ , 我们通常寻找对称矩阵  $A$ . 对称矩阵  $A$  的元素由下式确定:

对角元素  $A_{kk}$  等于  $x_k^2$  的系数. 非对角元素  $A_{jk} = A_{kj}$  等于  $x_j x_k$  系数的一半.

具体计算如下:  $A_{11} = 1$  ( $x_1^2$  的系数).  $A_{22} = -9$  ( $x_2^2$  的系数).  $A_{33} = 13$  ( $x_3^2$  的系数).  $A_{12} = A_{21} = 2/2 = 1$  ( $x_1 x_2$  的系数).  $A_{13} = A_{31} = -3/2 = -1.5$  ( $x_1 x_3$  的系数).  $A_{23} = A_{32} = 6/2 = 3$  ( $x_2 x_3$  的系数).

因此, 对称矩阵  $A$  为:

$$A = \begin{pmatrix} 1 & 1 & -1.5 \\ 1 & -9 & 3 \\ -1.5 & 3 & 13 \end{pmatrix}.$$

1.4. 证明: 引理 1.1 断言: 设  $(A\mathbf{x}, \mathbf{x})$  对所有  $\mathbf{x} \in \mathbb{C}^n$  都是实数. 那么  $A = A^*$ .

我们利用引理 1.2 (极化恒等式) 来证明, 或者直接推导. 已知对于任意向量  $\mathbf{v} \in \mathbb{C}^n$ ,  $(A\mathbf{v}, \mathbf{v}) \in \mathbb{R}$ , 即  $(A\mathbf{v}, \mathbf{v}) = \overline{(A\mathbf{v}, \mathbf{v})}$ . 由内积性质  $(A\mathbf{v}, \mathbf{v}) = (\mathbf{v}, A\mathbf{v}) = (A^*\mathbf{v}, \mathbf{v})$ . 因此, 我们有  $(A\mathbf{v}, \mathbf{v}) = (A^*\mathbf{v}, \mathbf{v})$ , 即:

$$((A - A^*)\mathbf{v}, \mathbf{v}) = 0, \quad \forall \mathbf{v} \in \mathbb{C}^n.$$

令  $B = A - A^*$ . 我们需要证明如果  $(B\mathbf{v}, \mathbf{v}) = 0$  对所有  $\mathbf{v}$  成立, 则  $B = \mathbf{0}$  (这将意味着  $A = A^*$ ). 利用引理 1.2 中的极化恒等式 (复数情况):

$$(B\mathbf{x}, \mathbf{y}) = \frac{1}{4} \sum_{\alpha \in \{1, -1, i, -i\}} \alpha (B(\mathbf{x} + \alpha\mathbf{y}), \mathbf{x} + \alpha\mathbf{y}).$$

由于假设对任意向量  $\mathbf{v}$ ,  $(B\mathbf{v}, \mathbf{v}) = 0$ , 上述公式右边的每一项  $(B(\mathbf{x} + \alpha\mathbf{y}), \mathbf{x} + \alpha\mathbf{y})$  均为 0. 因此,  $(B\mathbf{x}, \mathbf{y}) = 0$  对任意  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$  成立. 这意味着  $B = \mathbf{0}$ , 即  $A = A^*$ .

(如果不使用引理 1.2, 仅使用提示): 展开  $(A(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) = (A\mathbf{x}, \mathbf{x}) + (A\mathbf{y}, \mathbf{y}) + (A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{x})$ . 因为左边和前两项由假设都是实数, 所以  $(A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{x})$  必须是实数. 即  $(A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{x}) = \overline{(A\mathbf{x}, \mathbf{y}) + (A\mathbf{y}, \mathbf{x})} = \overline{(A\mathbf{x}, \mathbf{y})} + \overline{(A\mathbf{y}, \mathbf{x})} = (\mathbf{y}, A\mathbf{x}) + (\mathbf{x}, A\mathbf{y})$ . 同理, 展开  $(A(\mathbf{x} + i\mathbf{y}), \mathbf{x} + i\mathbf{y})$  可知  $-i(A\mathbf{x}, \mathbf{y}) + i(A\mathbf{y}, \mathbf{x})$  是实数. 这导致  $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y}) = \overline{(A\mathbf{y}, \mathbf{x})}$ , 结合  $A$  的伴随定义即得  $A = A^*$ .

2.1. 解: 对应的二次型为:

$$Q[\mathbf{x}] = x_1^2 + 3x_2^2 + x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3.$$

**方法一: 配方法** 我们首先处理包含  $x_1$  的项:  $x_1^2 + 4x_1x_2 + 2x_1x_3$ . 这看起来像是  $(x_1 + 2x_2 + x_3)^2$  的展开式的一部分.

$$(x_1 + 2x_2 + x_3)^2 = x_1^2 + 4x_2^2 + x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3.$$

对比原式  $Q[\mathbf{x}]$ , 我们发现:

$$Q[\mathbf{x}] = (x_1 + 2x_2 + x_3)^2 - x_2^2.$$

令  $y_1 = x_1 + 2x_2 + x_3$ ,  $y_2 = x_2$ ,  $y_3 = x_3$ . 则在新坐标下  $Q = y_1^2 - y_2^2 + 0y_3^2$ . 对角形式的矩阵为  $D = \text{diag}(1, -1, 0)$ .

**方法二：行/列运算**我们对增广矩阵  $(A|I)$  进行操作，旨在将  $A$  变为对角矩阵。注意：对  $A$  每做一次行运算，必须紧接着做相应的列运算。对右侧的  $I$  只做行运算（记录变换矩阵  $S^*$ ）。

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

消去第一列/第一行的非对角元素：1.  $R_2 \leftarrow R_2 - 2R_1, C_2 \leftarrow C_2 - 2C_1$ . 2.  $R_3 \leftarrow R_3 - R_1, C_3 \leftarrow C_3 - C_1$ .

$$\xrightarrow{R_2-2R_1} \left( \begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array} \right) \xrightarrow{C_2-2C_1} \left( \begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array} \right)$$

注意：行运算  $R_3 - R_1$  在  $R_2 - 2R_1$  之后进行（或者观察到第3行减第1行直接变0）。继续对第3行/列操作：

$$\xrightarrow{R_3-R_1} \left( \begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -2 & 1 & 0 \\ 0 & 0 & -1 & -1 & 0 & 1 \end{array} \right) \xrightarrow{C_3-C_1} \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{array} \right)$$

此时矩阵已对角化为  $D = \text{diag}(1, -1, 0)$ . （注：配方法中我们直接发现  $Q = (\dots)^2 - x_2^2$ , 这里  $D_{33}$  也是0，结果一致）。

**偏好：**对于低维且系数简单的矩阵，**配方法**通常更直观且计算量小。对于高维矩阵，行运算更系统化，不易出错。

**正定性：**对角化后的对角元素为  $1, -1, 0$ . 因为存在负数（且存在0），所以该矩阵**不是**正定的（它是不定矩阵）。正定矩阵要求所有对角元素均为正。

2.2. 解：我们需要求  $A$  的特征值和特征向量。

**步骤1：求特征值**计算特征多项式  $\det(A - \lambda I)$ :

$$\begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix}$$

观察到每一行的和都是  $4 - \lambda$ . 将第2、3列加到第1列：

$$= (4 - \lambda) \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix}$$

$R_2 - R_1, R_3 - R_1$ :

$$= (4 - \lambda) \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (4 - \lambda)(1 - \lambda)^2.$$

特征值为  $\lambda_1 = 4$  (代数重数1),  $\lambda_2 = 1$  (代数重数2)。所以对角矩阵为  $D = \text{diag}(4, 1, 1)$ .

**步骤2：求特征向量并标准化**对于  $\lambda_1 = 4$ : 解  $(A - 4I)\mathbf{x} = \mathbf{0}$ :

$$\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

得到特征向量  $\mathbf{v}_1 = (1, 1, 1)^T$ . 归一化:  $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ .

对于  $\lambda_2 = 1$ : 解  $(A - I)\mathbf{x} = \mathbf{0}$ :

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow x_1 + x_2 + x_3 = 0.$$

我们需要在平面  $x_1 + x_2 + x_3 = 0$  上找到两个正交的单位向量。取第一个向量  $\mathbf{v}_2 = (1, -1, 0)^T$  (显然满足方程)。归一化:  $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1, -1, 0)^T$ . 取第二个向量  $\mathbf{v}_3$ , 它必须

正交于  $\mathbf{v}_1$  (自动满足, 因为特征值不同) 和  $\mathbf{v}_2$ . 我们可以利用外积  $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2 = (1, 1, 1)^T \times (1, -1, 0)^T = (1, 1, -2)^T$ . 或者在平面方程中取一个与  $\mathbf{v}_2$  正交的解. 归一化:  $\mathbf{u}_3 = \frac{1}{\sqrt{6}}(1, 1, -2)^T$ .

**结论:** 对角矩阵  $D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ . 酉(正交)矩阵  $U = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{pmatrix}$ .

4.1. 解: **检查 A 的正定性:** 计算顺序主子式 (Leading Principal Minors): 1.  $\det A_1 = 4 > 0$ . 2.

$\det A_2 = \begin{vmatrix} 4 & 2 \\ 2 & 3 \end{vmatrix} = 12 - 4 = 8 > 0$ . 3.  $\det A_3 = \det A = 4(6 - 1) - 2(4 + 1) + 1(-2 - 3) = 20 - 10 - 5 = 5 > 0$ . 因为所有顺序主子式均为正, 所以 **A 是正定的**.

**检查 B 的正定性:** 1.  $\det B_1 = 3 > 0$ . 2.  $\det B_2 = \begin{vmatrix} 3 & -1 \\ -1 & 4 \end{vmatrix} = 12 - 1 = 11 > 0$ . 3.  $\det B_3 = \det B = 3(4 - 4) - (-1)(-1 + 4) + 2(2 - 8) = 0 + 3 - 12 = -9 < 0$ . 因为  $\det B < 0$ , 所以 **B 不是正定的** (它是不定的).

**其他矩阵的性质:** 由于 A 是正定的, 其特征值  $\lambda_i > 0$ .

$-A$ : 特征值为  $-\lambda_i < 0$ . 所以 **不是正定的** (它是负定的).

$A^3$ : 特征值为  $\lambda_i^3 > 0$ . 所以 **是正定的**.

$A^{-1}$ : 特征值为  $1/\lambda_i > 0$ . 所以 **是正定的**.

接下来涉及 B. 计算 B 的逆 (如果存在):  $\det B = -9 \neq 0$ , 所以 B 可逆. 由于 B 的主子式符号为  $(+, +, -)$ , 根据惯性定理或特征值符号规则, 它有负特征值. 具体来说,  $\det B < 0$  且  $\text{trace } B = 8 > 0$ , 这意味着它至少有一个负特征值和至少一个正特征值. 因此, B 是不定的,  $B^{-1}$  也是不定的 (特征值是 B 特征值的倒数, 符号不变).

$A + B^{-1}$ : 我们需要具体计算.  $B^{-1} = \frac{1}{-9} \text{adj}(B) = \frac{-1}{9} \begin{pmatrix} 0 & -3 & -6 \\ -3 & -1 & 4 \\ -6 & 4 & 11 \end{pmatrix}$ .  $A + B^{-1} =$

$\begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 1/9 & -4/9 \\ 2/3 & -4/9 & -11/9 \end{pmatrix}$ . 这计算比较繁琐, 我们检查  $9(A + B^{-1})$  的主

子式:  $M = 9A + 9B^{-1} = \begin{pmatrix} 36 & 18 & 9 \\ 18 & 27 & -9 \\ 9 & -9 & 18 \end{pmatrix} + \begin{pmatrix} 0 & 3 & 6 \\ 3 & 1 & -4 \\ 6 & -4 & -11 \end{pmatrix} = \begin{pmatrix} 36 & 21 & 15 \\ 21 & 28 & -13 \\ 15 & -13 & 7 \end{pmatrix}$ .

$M_1 = 36 > 0$ .  $M_2 = 36(28) - 21^2 = 1008 - 441 = 567 > 0$ .  $M_3 = \det M$ . 显然  $\det M < 0$ . 所以  $A + B^{-1}$  **不是正定的**.

$A + B$ :  $S = A + B = \begin{pmatrix} 7 & 1 & 3 \\ 1 & 7 & -3 \\ 3 & -3 & 3 \end{pmatrix}$ .  $S_1 = 7 > 0$ .  $S_2 = 49 - 1 = 48 > 0$ .  $S_3 =$

$7(21 - 9) - 1(3 + 9) + 3(-3 - 21) = 7(12) - 12 + 3(-24) = 84 - 12 - 72 = 0$ . 行列式为 0, 说明有零特征值. 所以它 **不是正定的** (它是半正定的).

$A - B$ :  $D = A - B = \begin{pmatrix} 1 & 3 & -1 \\ 3 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}$ .  $D_1 = 1 > 0$ .  $D_2 = -1 - 9 = -10 < 0$ . 所以 **不是正定的**.

4.2. a) **正确**. A 的特征值  $\lambda_i > 0 \implies \lambda_i^5 > 0$ .

b) **错误**. A 负定  $\implies \lambda_i < 0$ . 但  $A^8$  的特征值  $\lambda_i^8$  总是正的, 所以  $A^8$  是正定的.

c) **正确**. 偶数次幂将负特征值变为正特征值.

d) **正确**.  $A > 0$  且  $B \leq 0 \implies -B \geq 0$ . 两个正定 (或正定 + 半正定) 矩阵之和是正定的.

e) **错误**. 反例: 设  $A = \text{diag}(10, -1)$  (不定),  $B = \text{diag}(1, 100)$  (正定).  $A + B = \text{diag}(11, 99)$  是正定的.

4.3. 证明: 设  $A = \begin{pmatrix} a_{11} & a_{12} \\ \bar{a}_{12} & a_{22} \end{pmatrix}$ . 由于  $\det A \geq 0$ , 即  $a_{11}a_{22} - |a_{12}|^2 \geq 0$ , 这意味着  $a_{11}a_{22} \geq$



$|a_{12}|^2 \geq 0$ . 已知  $a_{11} \geq 0$ .

**情况 1:**  $a_{11} > 0$ . 由于  $a_{11}a_{22} \geq 0$  且  $a_{11} > 0$ , 必有  $a_{22} \geq 0$ . 迹  $\text{trace}(A) = a_{11} + a_{22} > 0$ . 特征值之积  $\lambda_1\lambda_2 = \det A \geq 0$ , 特征值之和  $\lambda_1 + \lambda_2 > 0$ . 这暗示  $\lambda_1, \lambda_2 \geq 0$ . 所以  $A$  是半正定的。

**情况 2:**  $a_{11} = 0$ . 不等式变为  $0 \cdot a_{22} - |a_{12}|^2 \geq 0 \implies -|a_{12}|^2 \geq 0$ . 这只有在  $a_{12} = 0$  时成立。此时  $A = \begin{pmatrix} 0 & 0 \\ 0 & a_{22} \end{pmatrix}$ .  $\det A = 0$ .

**注:** 题目陈述在这种情况下有瑕疵。如果  $A = \text{diag}(0, -1)$ , 则满足  $a_{11} = 0 \geq 0$  和  $\det A = 0 \geq 0$ , 但它是半负定的, 不是半正定的。如果题目隐含  $a_{11} > 0$  或者  $n \geq 3$  的反例暗示  $n = 2$  时通常成立 (忽略退化情况), 则主要逻辑见情况 1。如果必须严格证明, 则题目条件对于  $a_{11} = 0$  的情况是不充分的, 除非补充  $a_{22} \geq 0$ 。

4.4. 解: 取一个对角矩阵, 其中前  $n-1$  个元素为正 (或非负), 最后一个元素为负, 且行列式为 0. 设  $A = \text{diag}(1, 0, -1)$  (这里  $n = 3$ )。

检查主子式:  $k = 1: \det A_1 = 1 \geq 0$ .

$k = 2: \det A_2 = 1 \cdot 0 = 0 \geq 0$ .

$k = 3: \det A_3 = 1 \cdot 0 \cdot (-1) = 0 \geq 0$ .

所有  $\det A_k \geq 0$  均成立。

然而, 特征值为  $1, 0, -1$ . 因为存在负特征值  $-1$ , 所以  $A$  不是半正定的。

4.5. 证明: 1. 根据塞尔维斯特正定性判据, 因为  $\det A_k > 0$  对所有  $k = 1, \dots, n-1$  成立, 所以  $A$  的左上  $(n-1) \times (n-1)$  子矩阵  $A_{n-1}$  是**正定**的。

2. 根据特征值交错定理 (推论 4.4), 设  $A_{n-1}$  的特征值为  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{n-1}$ , 它们全是正数 ( $\mu_i > 0$ )。

3. 设  $A$  的特征值为  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . 交错定理告诉我们:

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \mu_{n-1} \geq \lambda_n.$$

4. 因为  $\mu_{n-1} > 0$ , 所以  $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$  都必须大于 0。

5. 现在只剩  $\lambda_n$  的符号未知。已知  $\det A = \lambda_1\lambda_2 \dots \lambda_n \geq 0$ . 因为前  $n-1$  个特征值都是正的, 它们的乘积也是正的。为了使总乘积  $\geq 0$ , 必须有  $\lambda_n \geq 0$ 。

6. 既然所有特征值  $\lambda_i \geq 0$ , 则  $A$  是半正定的。

4.6. 解: 我们需要利用习题 4.5 的反面情况。为了使  $A$  不是半正定, 我们需要破坏 “ $A_{n-1}$  是正定” 这一条件。即便  $\det A_2 \geq 0$ , 如果  $A_2$  不是正定的 (例如它有一个 0 特征值), 那么交错定理允许  $A$  有一个负特征值。

**构造:** 让  $A_2 = \text{diag}(1, 0)$ . 满足  $a_{11} > 0$  且  $\det A_2 = 0 \geq 0$ . 让  $A_3$  有一个负元素。取

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

**验证:** 1.  $a_{1,1} = 1 > 0$ . (满足) 2.  $\det A_2 = 1 \cdot 0 - 0 = 0 \geq 0$ . (满足) 3.  $\det A_3 = 1 \cdot 0 \cdot (-1) = 0 \geq 0$ . (满足)

**结论:** 特征值为  $1, 0, -1$ . 存在负特征值, 所以  $A$  不是半正定的。

## 第八章习题解答

1.1. 解: **a)** 设标量  $c_1, c_2, \dots, c_r$  使得

$$\sum_{j=1}^r c_j \mathbf{v}_j = \mathbf{0}.$$

我们需要证明所有  $c_j$  均为 0。对上述等式两边应用线性泛函  $\mathbf{v}'_k$  (对于任意  $k \in \{1, \dots, r\}$ ):

$$\mathbf{v}'_k \left( \sum_{j=1}^r c_j \mathbf{v}_j \right) = \mathbf{v}'_k(\mathbf{0}).$$

利用线性泛函的线性性质:

$$\sum_{j=1}^r c_j \mathbf{v}'_k(\mathbf{v}_j) = 0.$$

根据双正交性质  $\mathbf{v}'_k(\mathbf{v}_j) = \delta_{kj}$ , 求和符号中只有当  $j = k$  时的项非零, 因此:

$$c_k \cdot 1 = 0 \implies c_k = 0.$$

由于这对所有  $k$  都成立, 因此向量系统  $\mathbf{v}_1, \dots, \mathbf{v}_r$  是线性无关的。

**b)** 设  $n = \dim X$ . 如果  $\mathbf{v}_1, \dots, \mathbf{v}_r$  不是生成集, 则  $r < n$ . 根据第二章命题 5.4 (基的扩充定理), 我们可以将  $\mathbf{v}_1, \dots, \mathbf{v}_r$  扩充为  $X$  的一组基:

$$\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}.$$

设  $\mathbf{v}'_1, \dots, \mathbf{v}'_n$  是基  $\mathcal{B}$  的对偶基。根据定义, 它们满足  $\mathbf{v}'_k(\mathbf{v}_j) = \delta_{kj}$  (对于所有  $1 \leq k, j \leq n$ )。特别是, 对于  $k = 1, \dots, r$ , 这些  $\mathbf{v}'_k$  满足题目要求的条件。

现在我们构造另一组泛函。定义  $\mathbf{w}'_r$  如下:

$$\mathbf{w}'_r = \mathbf{v}'_r + \mathbf{v}'_{r+1}.$$

对于  $j = 1, \dots, r$ , 我们要验证  $\mathbf{w}'_r(\mathbf{v}_j) = \delta_{rj}$ : 如果  $j < r$ :  $\mathbf{w}'_r(\mathbf{v}_j) = \mathbf{v}'_r(\mathbf{v}_j) + \mathbf{v}'_{r+1}(\mathbf{v}_j) = 0 + 0 = 0$ . 如果  $j = r$ :  $\mathbf{w}'_r(\mathbf{v}_r) = \mathbf{v}'_r(\mathbf{v}_r) + \mathbf{v}'_{r+1}(\mathbf{v}_r) = 1 + 0 = 1$ . 因此, 泛函系统  $\mathbf{v}'_1, \dots, \mathbf{v}'_{r-1}, \mathbf{w}'_r$  也满足题目中的双正交条件。但是,  $\mathbf{w}'_r \neq \mathbf{v}'_r$ , 因为在基向量  $\mathbf{v}_{r+1}$  上:

$$\mathbf{w}'_r(\mathbf{v}_{r+1}) = \mathbf{v}'_r(\mathbf{v}_{r+1}) + \mathbf{v}'_{r+1}(\mathbf{v}_{r+1}) = 0 + 1 = 1,$$

而  $\mathbf{v}'_r(\mathbf{v}_{r+1}) = 0$ . 因此, 满足条件的双正交系统不是唯一的。

- 1.2. 证明: (1.5) 式即要求  $p(a_k) = y_k$  对所有  $k = 1, \dots, n+1$  成立。令  $V = \mathbb{P}_n$  为次数不超过  $n$  的多项式空间, 已知  $\dim V = n+1$ . 定义线性泛函  $\mathbf{f}_k \in V'$  为求值泛函:  $\mathbf{f}_k(p) = p(a_k)$ . 我们在正文中 (通过拉格朗日插值公式) 已经构造了多项式系统  $p_1, p_2, \dots, p_{n+1} \in V$ , 使得:

$$\mathbf{f}_k(p_j) = p_j(a_k) = \delta_{kj}.$$

这正是习题 1.1 中的条件。根据习题 1.1 (a) 的结论, 多项式系统  $p_1, \dots, p_{n+1}$  是线性无关的。由于  $V$  的维数为  $n+1$ , 且我们有  $n+1$  个线性无关的向量, 因此  $\{p_1, \dots, p_{n+1}\}$  构成了  $V$  的一组基。

根据线性代数的基本性质, 向量空间中的任何向量都可以唯一地表示为基向量的线性组合。设任意多项式  $p \in V$ . 我们可以将其写为:

$$p = \sum_{j=1}^{n+1} c_j p_j.$$

为了确定系数  $c_j$ , 我们应用线性泛函  $\mathbf{f}_k$  (即代入  $t = a_k$ ):

$$p(a_k) = \mathbf{f}_k \left( \sum_{j=1}^{n+1} c_j p_j \right) = \sum_{j=1}^{n+1} c_j \mathbf{f}_k(p_j) = \sum_{j=1}^{n+1} c_j \delta_{kj} = c_k.$$

因此, 为了满足条件  $p(a_k) = y_k$ , 我们必须且只能取  $c_k = y_k$ . 这就证明了满足条件的多项式  $p$  存在且形式唯一, 即

$$p(t) = \sum_{k=1}^{n+1} y_k p_k(t).$$

- 3.1. 证明: 将等式移项, 定义线性变换  $S = T - T_1$ . 我们需要证明  $S$  是零变换。条件变为:

$$\langle (T - T_1)\mathbf{x}, \mathbf{y}' \rangle = 0 \implies \langle S\mathbf{x}, \mathbf{y}' \rangle = 0, \quad \forall \mathbf{x} \in X, \forall \mathbf{y}' \in Y'.$$

固定任意一个向量  $\mathbf{x} \in X$ , 令  $\mathbf{v} = S\mathbf{x} \in Y$ . 上述条件意味着对于这个特定的向量  $\mathbf{v}$ , 我们有:

$$\langle \mathbf{v}, \mathbf{y}' \rangle = \mathbf{y}'(\mathbf{v}) = 0, \quad \forall \mathbf{y}' \in Y'.$$

根据本章引理 1.3 (或者推论 1.4, 关于对偶配对的非退化性): 如果一个向量  $\mathbf{v}$  使得所有线性泛函在它身上的作用都为 0, 那么这个向量必须是零向量。因此,  $\mathbf{v} = \mathbf{0}$ . 即  $S\mathbf{x} = \mathbf{0}$ . 因为  $\mathbf{x}$  是  $X$  中任意选取的, 所以  $S$  将每一个向量都映射为零向量。即  $S = 0$ , 从而  $T = T_1$ .

- 3.2. 解: 设  $H$  是一个内积空间 (实或复),  $T: H \rightarrow H$  是一个线性算子。我们要定义  $T$  的伴随  $T^*$ . 对于任意固定的  $\mathbf{y} \in H$ , 考虑映射  $L_{\mathbf{y}}: H \rightarrow \mathbb{F}$ , 定义为:

$$L_{\mathbf{y}}(\mathbf{x}) = (\mathbf{x}, T^*\mathbf{y})$$

这看起来是倒果为因, 我们应该模仿 3.1.3 节的构造: 固定  $\mathbf{y} \in H$ . 考虑映射  $\varphi: H \rightarrow \mathbb{F}$ , 定义为  $\varphi(\mathbf{x}) = (T\mathbf{x}, \mathbf{y})$ . 由于  $T$  是线性的, 且内积对第一个参数是线性的, 所以  $\varphi$  是  $H$  上的一个线性泛函。根据里斯表示定理 (定理 2.1), 存在唯一的向量  $\mathbf{z} \in H$  使得

$$\varphi(\mathbf{x}) = (\mathbf{x}, \mathbf{z}) \quad \forall \mathbf{x} \in H.$$

也就是  $(T\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{z})$ . 我们将这个依赖于  $\mathbf{y}$  的向量  $\mathbf{z}$  定义为  $T^*\mathbf{y}$ . 即定义映射  $T^*: H \rightarrow H$  使得  $T^*\mathbf{y} = \mathbf{z}$ . 因此, 埃尔米特伴随  $T^*$  被定义为满足以下等式的唯一变换:

$$(T\mathbf{x}, \mathbf{y}) = (\mathbf{x}, T^*\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in H.$$

**关于线性的说明:** 类似于 3.1.3 节, 我们需要验证  $T^*$  是线性的 (在复空间中这很重要, 因为内积的第二项是共轭线性的)。对于  $\mathbf{y}_1, \mathbf{y}_2 \in H$  和标量  $\alpha$ :

$$(\mathbf{x}, T^*(\alpha\mathbf{y}_1 + \mathbf{y}_2)) = (T\mathbf{x}, \alpha\mathbf{y}_1 + \mathbf{y}_2) = \bar{\alpha}(T\mathbf{x}, \mathbf{y}_1) + (T\mathbf{x}, \mathbf{y}_2).$$

另一方面:

$$(\mathbf{x}, \alpha T^*\mathbf{y}_1 + T^*\mathbf{y}_2) = \bar{\alpha}(\mathbf{x}, T^*\mathbf{y}_1) + (\mathbf{x}, T^*\mathbf{y}_2) = \bar{\alpha}(T\mathbf{x}, \mathbf{y}_1) + (T\mathbf{x}, \mathbf{y}_2).$$

(注意: 内积的共轭线性性质在两边抵消了, 或者更严谨地说, 我们利用里斯定理的唯一性)。比较两端, 由向量的唯一性 (里斯定理), 可得  $T^*(\alpha\mathbf{y}_1 + \mathbf{y}_2) = \alpha T^*\mathbf{y}_1 + T^*\mathbf{y}_2$ . 所以  $T^*$  是线性的。

- 3.3. 证明: 回顾对偶基的定义:  $\mathbf{v}'_i(\mathbf{v}_j) = \delta_{ij}$ . 零化子  $E^\perp$  的定义为:  $E^\perp = \{f \in X' : f(\mathbf{x}) = 0, \forall \mathbf{x} \in E\}$ .

**第一步: 证明**  $\text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\} \subset E^\perp$ . 设  $f \in \text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\}$ . 则  $f$  可以表示为:

$$f = \sum_{k=r+1}^n c_k \mathbf{v}'_k.$$

对于任意  $\mathbf{x} \in E$ , 由于  $\mathbf{v}_1, \dots, \mathbf{v}_r$  生成  $E$ ,  $\mathbf{x}$  可以表示为  $\mathbf{x} = \sum_{j=1}^r a_j \mathbf{v}_j$ . 计算  $f(\mathbf{x})$ :

$$f(\mathbf{x}) = \left( \sum_{k=r+1}^n c_k \mathbf{v}'_k \right) \left( \sum_{j=1}^r a_j \mathbf{v}_j \right) = \sum_{k=r+1}^n \sum_{j=1}^r c_k a_j \mathbf{v}'_k(\mathbf{v}_j).$$

注意求和下标的范围:  $k$  从  $r+1$  到  $n$ , 而  $j$  从 1 到  $r$ . 因此  $k$  永远不等于  $j$ , 所以  $\mathbf{v}'_k(\mathbf{v}_j) = \delta_{kj} = 0$ . 所以  $f(\mathbf{x}) = 0$ . 这意味着  $f \in E^\perp$ .

**第二步: 证明**  $E^\perp \subset \text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\}$ . 设  $g \in E^\perp$ . 由于  $\mathbf{v}'_1, \dots, \mathbf{v}'_n$  是  $X'$  的基, 我们可以将  $g$  展开为:

$$g = \sum_{k=1}^n c_k \mathbf{v}'_k.$$

因为  $g \in E^\perp$ , 所以对于  $E$  中的基向量  $\mathbf{v}_j$  ( $1 \leq j \leq r$ ), 必须有  $g(\mathbf{v}_j) = 0$ . 代入计算:

$$0 = g(\mathbf{v}_j) = \left( \sum_{k=1}^n c_k \mathbf{v}'_k \right) (\mathbf{v}_j) = \sum_{k=1}^n c_k \delta_{kj} = c_j.$$

这意味着对于所有  $j = 1, \dots, r$ , 系数  $c_j$  必须为 0。因此,  $g$  的展开式中只剩下  $k > r$  的项:

$$g = \sum_{k=r+1}^n c_k \mathbf{v}'_k.$$

这说明  $g \in \text{span}\{\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n\}$ .

综上所述, 集合相等得证。

- 3.4. 证明: 设  $\dim X = n$ . 根据习题 3.3, 我们可以选取  $E$  的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_r$ , 其中  $r = \dim E$ . 根据基扩充定理, 将其扩充为  $X$  的基  $\mathbf{v}_1, \dots, \mathbf{v}_n$ . 设  $\mathbf{v}'_1, \dots, \mathbf{v}'_n$  为其对偶基。由习题 3.3 的结论可知,  $E^\perp$  的基正是  $\mathbf{v}'_{r+1}, \dots, \mathbf{v}'_n$ . 我们需要计算这个基中向量的个数。指标  $k$  从  $r+1$  到  $n$ , 共有  $n - (r+1) + 1 = n - r$  个向量。因此:

$$\dim E^\perp = n - r = \dim X - \dim E.$$

移项即得:

$$\dim E + \dim E^\perp = \dim X.$$

得证。

- 4.1. 证明: 设旧坐标为  $x_1, \dots, x_n$ , 新坐标为  $\tilde{x}_1, \dots, \tilde{x}_n$ . 我们假设坐标变换是线性的 (或者我们在切空间中考虑局部线性化), 坐标变换关系记为:

$$x_k = x_k(\tilde{x}_1, \dots, \tilde{x}_n).$$

根据多元微积分的链式法则, 对新坐标的偏导数可以表示为对旧坐标偏导数的线性组合:

$$\frac{\partial}{\partial \tilde{x}_j} = \sum_{k=1}^n \frac{\partial x_k}{\partial \tilde{x}_j} \frac{\partial}{\partial x_k}.$$

现在, 假设微分算子  $D$  在新坐标系下的表示为:

$$D = \sum_{j=1}^n \tilde{v}_j \frac{\partial}{\partial \tilde{x}_j}.$$

我们将链式法则代入上式:

$$D = \sum_{j=1}^n \tilde{v}_j \left( \sum_{k=1}^n \frac{\partial x_k}{\partial \tilde{x}_j} \frac{\partial}{\partial x_k} \right).$$

交换求和顺序:

$$D = \sum_{k=1}^n \left( \sum_{j=1}^n \frac{\partial x_k}{\partial \tilde{x}_j} \tilde{v}_j \right) \frac{\partial}{\partial x_k}.$$

将此式与  $D$  在旧坐标系下的定义  $D = \sum_{k=1}^n v_k \frac{\partial}{\partial x_k}$  进行比较。由于偏导算子  $\frac{\partial}{\partial x_k}$  构成了切空间的一组基, 对应系数必须相等:

$$v_k = \sum_{j=1}^n \frac{\partial x_k}{\partial \tilde{x}_j} \tilde{v}_j.$$

如果坐标变换是线性的, 即  $x = S\tilde{x}$  (其中  $S$  是从新基到旧基的坐标变换矩阵, 或者旧坐标对新坐标的雅可比矩阵), 那么  $\frac{\partial x_k}{\partial \tilde{x}_j}$  正是矩阵  $S$  的元素  $S_{kj}$ . 于是上式变为:

$$v_k = \sum_{j=1}^n S_{kj} \tilde{v}_j,$$

或者用矩阵形式表示:  $\mathbf{v} = S\tilde{\mathbf{v}}$ . 这等价于  $\tilde{\mathbf{v}} = S^{-1}\mathbf{v}$ . 这正是向量坐标的变换规则 (坐标随基变换的逆矩阵进行变换)。因此, 微分算子  $D$  的系数  $v_k$  确实像向量的坐标一样进行变换。

- 5.1. 证明: 根据注记 5.3, 向量的张量积  $T = \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_p$  被定义为一个作用在对偶空间  $V'_1 \times \cdots \times V'_p$  上的多线性泛函. 具体地, 对于任意  $f_1 \in V'_1, \dots, f_p \in V'_p$ , 其定义为:

$$(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p)(f_1, \dots, f_p) = f_1(\mathbf{v}_1)f_2(\mathbf{v}_2)\cdots f_p(\mathbf{v}_p).$$

我们需要证明映射  $(\mathbf{v}_1, \dots, \mathbf{v}_p) \mapsto \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p$  在每个变量  $\mathbf{v}_k$  上是线性的. 不失一般性, 我们考虑第  $k$  个变量. 设  $\mathbf{v}_k = \alpha\mathbf{u} + \beta\mathbf{w}$ , 其中  $\mathbf{u}, \mathbf{w} \in V_k$ ,  $\alpha, \beta \in \mathbb{F}$ . 其他变量  $\mathbf{v}_j$  ( $j \neq k$ ) 固定不变. 考虑张量  $\mathbf{v}_1 \otimes \cdots \otimes (\alpha\mathbf{u} + \beta\mathbf{w}) \otimes \cdots \otimes \mathbf{v}_p$  作用在任意泛函组  $(f_1, \dots, f_p)$  上:

$$\begin{aligned} & [\mathbf{v}_1 \otimes \cdots \otimes (\alpha\mathbf{u} + \beta\mathbf{w}) \otimes \cdots \otimes \mathbf{v}_p](f_1, \dots, f_p) \\ &= f_1(\mathbf{v}_1) \cdots f_k(\alpha\mathbf{u} + \beta\mathbf{w}) \cdots f_p(\mathbf{v}_p) \\ &= f_1(\mathbf{v}_1) \cdots [\alpha f_k(\mathbf{u}) + \beta f_k(\mathbf{w})] \cdots f_p(\mathbf{v}_p) \quad (\text{因为 } f_k \text{ 是线性的}) \\ &= \alpha[f_1(\mathbf{v}_1) \cdots f_k(\mathbf{u}) \cdots f_p(\mathbf{v}_p)] + \beta[f_1(\mathbf{v}_1) \cdots f_k(\mathbf{w}) \cdots f_p(\mathbf{v}_p)] \\ &= \alpha(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{u} \otimes \cdots \otimes \mathbf{v}_p)(f_1, \dots, f_p) + \beta(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{w} \otimes \cdots \otimes \mathbf{v}_p)(f_1, \dots, f_p). \end{aligned}$$

由于这对所有  $(f_1, \dots, f_p)$  都成立, 根据函数相等的定义, 我们有:

$$\mathbf{v}_1 \otimes \cdots \otimes (\alpha\mathbf{u} + \beta\mathbf{w}) \otimes \cdots \otimes \mathbf{v}_p = \alpha(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{u} \otimes \cdots \otimes \mathbf{v}_p) + \beta(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{w} \otimes \cdots \otimes \mathbf{v}_p).$$

得证.

- 5.2. 证明: 我们通过一个简单的反例来证明 (假设  $\dim V_k \geq 2$ ). 考虑  $p = 2$ , 且  $V_1 = V_2 = \mathbb{R}^2$ . 设  $\mathbf{e}_1, \mathbf{e}_2$  是  $\mathbb{R}^2$  的标准基. 张量积空间  $V_1 \otimes V_2$  的一组基为  $\{\mathbf{e}_1 \otimes \mathbf{e}_1, \mathbf{e}_1 \otimes \mathbf{e}_2, \mathbf{e}_2 \otimes \mathbf{e}_1, \mathbf{e}_2 \otimes \mathbf{e}_2\}$ . 考虑张量  $T = \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2$ . 显然  $T \in V_1 \otimes V_2$ . 我们证明  $T$  不能写成单一的张量积  $\mathbf{v} \otimes \mathbf{w}$  的形式. 假设存在  $\mathbf{v} = (x_1, x_2)^T$  和  $\mathbf{w} = (y_1, y_2)^T$  使得  $T = \mathbf{v} \otimes \mathbf{w}$ . 将  $\mathbf{v}$  和  $\mathbf{w}$  在基下展开:

$$\begin{aligned} \mathbf{v} \otimes \mathbf{w} &= (x_1\mathbf{e}_1 + x_2\mathbf{e}_2) \otimes (y_1\mathbf{e}_1 + y_2\mathbf{e}_2) \\ &= x_1y_1(\mathbf{e}_1 \otimes \mathbf{e}_1) + x_1y_2(\mathbf{e}_1 \otimes \mathbf{e}_2) + x_2y_1(\mathbf{e}_2 \otimes \mathbf{e}_1) + x_2y_2(\mathbf{e}_2 \otimes \mathbf{e}_2). \end{aligned}$$

这就要求该展开式的系数与  $T$  的系数相同.  $T$  的系数矩阵 (对应于基向量的系数) 为:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

而  $\mathbf{v} \otimes \mathbf{w}$  的系数矩阵为:

$$\begin{pmatrix} x_1y_1 & x_1y_2 \\ x_2y_1 & x_2y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} y_1 & y_2 \end{pmatrix}.$$

这就意味着我们需要找到列向量  $\mathbf{v}$  和行向量  $\mathbf{w}^T$  使得它们的乘积等于  $2 \times 2$  单位矩阵  $I$ . 然而, 矩阵  $\mathbf{v}\mathbf{w}^T$  的秩最多为 1 (它的列是  $\mathbf{v}$  的倍数), 而单位矩阵  $I$  的秩为 2. 这导致矛盾. 因此,  $T$  不能表示为  $\mathbf{v} \otimes \mathbf{w}$ . 这意味着形如  $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p$  的元素 (通常称为 “纯张量” 或 “秩-1 张量”) 的集合只是整个张量积空间的一个真子集.

- 5.3. 证明: 命题 5.6 声明: 给定一个张量  $\tilde{F} \in L(V_1, \dots, V_p, V'; \mathbb{F})$ , 存在一个唯一的变换  $F \in L(V_1, \dots, V_p; V)$  使得

$$\tilde{F}(\mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{v}') = \langle F(\mathbf{v}_1, \dots, \mathbf{v}_p), \mathbf{v}' \rangle$$

对所有  $\mathbf{v}_k \in V_k$  和  $\mathbf{v}' \in V'$  成立. 我们假设存在两个这样的变换  $F_1$  和  $F_2$  都满足上述条件. 那么对于任意固定的输入  $(\mathbf{v}_1, \dots, \mathbf{v}_p)$ , 以及任意的  $\mathbf{v}' \in V'$ , 我们有:

$$\langle F_1(\mathbf{v}_1, \dots, \mathbf{v}_p), \mathbf{v}' \rangle = \tilde{F}(\mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{v}') = \langle F_2(\mathbf{v}_1, \dots, \mathbf{v}_p), \mathbf{v}' \rangle.$$

这意味着:

$$\langle F_1(\mathbf{v}_1, \dots, \mathbf{v}_p) - F_2(\mathbf{v}_1, \dots, \mathbf{v}_p), \mathbf{v}' \rangle = 0, \quad \forall \mathbf{v}' \in V'.$$

令向量  $\mathbf{u} = F_1(\mathbf{v}_1, \dots, \mathbf{v}_p) - F_2(\mathbf{v}_1, \dots, \mathbf{v}_p) \in V$ . 上式表明  $\langle \mathbf{u}, \mathbf{v}' \rangle = 0$  对所有  $\mathbf{v}' \in V'$  成立. 根据引理 1.3 (或本章前面的习题 3.1), 如果一个向量被对偶空间中的所有泛函零化, 则该向量必须是零向量. 因此  $\mathbf{u} = \mathbf{0}$ , 即:

$$F_1(\mathbf{v}_1, \dots, \mathbf{v}_p) = F_2(\mathbf{v}_1, \dots, \mathbf{v}_p).$$

由于这对定义域中的所有向量组  $(\mathbf{v}_1, \dots, \mathbf{v}_p)$  都成立, 所以变换  $F_1$  和  $F_2$  是相等的。唯一性得证。

## 第九章习题解答

- 1.1. 证明: 设  $A = SDS^{-1}$ , 其中  $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . 首先回顾相似矩阵的一个基本性质: 若  $A = SDS^{-1}$ , 则对于任何正整数  $k$ , 有

$$A^k = (SDS^{-1})(SDS^{-1}) \dots (SDS^{-1}) = SD^k S^{-1}.$$

对于多项式  $p(t) = \sum_{k=0}^n c_k t^k$ , 我们可以计算  $p(A)$ :

$$\begin{aligned} p(A) &= \sum_{k=0}^n c_k A^k = \sum_{k=0}^n c_k (SD^k S^{-1}) \\ &= S \left( \sum_{k=0}^n c_k D^k \right) S^{-1} \quad (\text{利用矩阵乘法的线性}) \\ &= Sp(D)S^{-1}. \end{aligned}$$

现在我们需要计算  $p(D)$ . 由于  $D$  是对角矩阵, 其幂  $D^k$  也是对角矩阵, 且对角线元素为  $\lambda_i^k$ . 因此,  $p(D)$  也是对角矩阵, 其对角线元素为  $p(\lambda_i)$ :

$$p(D) = \text{diag}\{p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)\}.$$

这里的  $\lambda_i$  是对角矩阵  $D$  的对角元。因为  $A$  与  $D$  相似, 它们具有相同的特征多项式, 且  $D$  的对角元正是其特征值 (也就是特征多项式  $p(\lambda)$  的根)。根据特征值和特征多项式的定义, 对于所有的  $i = 1, \dots, n$ , 都有:

$$p(\lambda_i) = 0.$$

因此,

$$p(D) = \text{diag}\{0, 0, \dots, 0\} = \mathbf{0}.$$

最后, 将其代回  $p(A)$  的表达式中:

$$p(A) = S \cdot \mathbf{0} \cdot S^{-1} = \mathbf{0}.$$

得证。

- 2.1. 证明: **方法一: 使用谱映射定理** 设  $A$  是幂零的, 即存在整数  $k \geq 1$  使得  $A^k = \mathbf{0}$ . 定义多项式  $p(z) = z^k$ . 根据谱映射定理 (定理 2.1), 我们有:

$$\sigma(p(A)) = p(\sigma(A)).$$

因为  $p(A) = A^k = \mathbf{0}$ , 而零算子的谱只包含  $0$  (即  $\sigma(\mathbf{0}) = \{0\}$ ), 所以等式左边为  $\{0\}$ . 等式右边为  $\{\lambda^k : \lambda \in \sigma(A)\}$ . 因此, 我们有:

$$\{0\} = \{\lambda^k : \lambda \in \sigma(A)\}.$$

这意味着对于  $A$  的任何特征值  $\lambda$ , 必须满足  $\lambda^k = 0$ . 在复数域 (或实数域) 中,  $\lambda^k = 0$  意味着  $\lambda = 0$ . 因为在复向量空间中算子的谱非空, 所以  $\sigma(A) = \{0\}$ .

**方法二: 不使用谱映射定理 (直接证明)** 设  $\lambda \in \sigma(A)$  是  $A$  的一个特征值。根据定义, 存在一个非零向量  $\mathbf{v} \neq \mathbf{0}$ , 使得

$$A\mathbf{v} = \lambda\mathbf{v}.$$

我们在等式两边反复应用算子  $A$ :

$$A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda^2\mathbf{v}$$

$$A^3\mathbf{v} = \lambda^3\mathbf{v}$$

$$\vdots$$

$$A^k\mathbf{v} = \lambda^k\mathbf{v}.$$

由假设  $A^k = \mathbf{0}$ , 所以等式左边  $A^k\mathbf{v} = \mathbf{0}\mathbf{v} = \mathbf{0}$ . 因此我们得到:

$$\mathbf{0} = \lambda^k\mathbf{v}.$$

因为  $\mathbf{v} \neq \mathbf{0}$ , 根据向量空间的公理, 标量系数必须为 0, 即:

$$\lambda^k = 0.$$

这推导出  $\lambda = 0$ . 所以  $A$  唯一的特征值是 0, 即  $\sigma(A) = \{0\}$ .

(THE END)





# 索引

$\delta_{k,j}$ , *see* 克罗内克符号

$M_{m,n}$ , 3

$M_{m \times n}$ , 3

爱因斯坦求和约定, 206

伴随

算子的, 102

不变子空间, 223

重数, 88

代数, 88

几何, 88

等距同构, 127

对偶基, 190

对偶空间, 189

对偶性, 213

多项式矩阵, 81

多线性泛函, *see* 张量

多线性函数, 209

二次型, 171

泛函

线性, 189

范数, 101

弗罗贝尼乌斯, 158

欧几里得 (在  $\mathbb{R}^n$  或  $\mathbb{C}^n$  中), 101

算子 (线性变换的), 158

希尔伯特-施密特, 158

赋范空间, 106

傅里叶分解

抽象, 非正交, 193

抽象, 正交, 111

弗罗贝尼乌斯范数, 158

格拉姆-施密特正交化, 114

广义特征向量, 227

广义特征子空间, 227

基, 5

标准正交, 110

对偶, 190

双正交, *see* 基, 对偶

正交, 110

子空间的, 93

迹, 20

价

张量的, 209

计入重数, 88

矩阵, 3

埃尔米特, 143

对称, 4

反对称, 10

三角, 70

上三角, 28, 70

下三角, 70

克罗内克符号, 192

空间

行空间, 52

核空间, 27

基本子空间, 52

- 列空间, 51
- 零空间, 27
- 像空间, 27
- 左零空间, 52
- 摩尔-彭罗斯逆, 160
- 内积
  - 抽象, 104
  - 在  $\mathbb{C}^n$  中, 103
- 内积空间, 103
- 谱, 86
- 谱理论, 85
- 奇异值, 150
- 奇异值分解, 153
- 若尔当标准
  - 基, 235
  - 型, 235
  - 型 (幂零算子的), 234
- 若尔当标准
  - 基 (幂零算子的), 234
- 三角矩阵, 70
  - 特征值, 88
- 生成系统, 6
- 施密特分解, 151
- 双线性型, 171
- 算子范数, 157
- 特征向量, 85
- 特征值, 85
- 特征子空间, 86
- 投影
  - 正交, 112
- 项, 3
  - 矩阵的, 3
  - 张量的, 210
- 相似, 63
- 线性泛函, 189
- 线性无关, 7
- 线性相关, 7
- 线性组合, 5
  - 平凡, 7
- 酉算子, 128
- 张量, 209
  - 逆变, 213
  - $r$ -协变  $s$ -逆变, 213
  - 协变, 213
- 正规算子, 144
- 正交补, 116
- 正交分解, 116
- 正交 (相似) 对角化, 174
- 转置, 4
- 自伴算子, 143
- 子矩阵, 82
- 子空间
  - 不变, 223
- 子式, 82
- 最小二乘解, 117
  - 最小范数, 160
- 坐标, 5
  - 向量在基下的, 5

# 译后记

从大一开学后的一个月（十月中上旬）到今天，经过两个多月，我终于完成了这本书的翻译工作。

事情缘起于李耀文老师的线性代数课及我们使用的主要参考书《Linear Algebra Done Wrong》。这本教材极为优秀，但全英文内容对许多同学而言确实是一个不小的挑战。

为了帮助大家，并提升自我，我设立了几个目标：力求使译文精准且易于理解、采用专业的  $\text{\LaTeX}$  排版以匹配原书的阅读体验、成果永久免费共享，并通过 GitHub 开源以汇集大家的智慧。

如今回顾，我可以欣慰地说，早先设定的目标和承诺已然实现。对标英文原稿的排版，亲自输入书中每一个复杂的  $\text{\LaTeX}$  公式——这一工作量确实远超出了最初的预期，但正是在这一过程中，我获得了极大的成长与满足。

在这里感谢李耀文老师的关注与指导，感谢匡亚明学院各位同学的支持与鼓励，也感谢英语张子源老师分享的电子词典。正是在你们的帮助下，我才能顺利完成这本书的翻译工作。

希望这份译稿能够切实帮助同级同学及未来的学弟学妹们。若它能为你的学习带来些许便利，那么我所有的努力便具有了意义。

董耀择

2025 年 12 月